This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Uimonen, Mikko M.; Hulkkonen, Sina M.; Ryhänen, Jorma; Ponkilainen, Ville T.; Häkkinen, Arja H.; Karppinen, Jaro; Repo, Jussi P.

Title: Assessment of construct validity of the Finnish versions of the Disabilities of Arm, Shoulder and Hand Instrument and the Michigan Hand Outcomes Questionnaire

Year: 2020

Version: Accepted version (Final draft)

Copyright: © 2019 Hanley & Belfus, an imprint of Elsevier Inc. All rights reserved.

Rights: CC BY-NC-ND 4.0

Rights url: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Please cite the original version:

Uimonen, M. M., Hulkkonen, S. M., Ryhänen, J., Ponkilainen, V. T., Häkkinen, A. H., Karppinen, J., & Repo, J. P. (2020). Assessment of construct validity of the Finnish versions of the Disabilities of Arm, Shoulder and Hand Instrument and the Michigan Hand Outcomes Questionnaire. *Journal of Hand Therapy*, 33(4), 571-579. <https://doi.org/10.1016/j.jht.2019.03.008>

1 **Assessment of construct validity of the Finnish versions of the Disabilities of Arm, Shoulder**
2 **and Hand Instrument and the Michigan Hand Outcomes Questionnaire**

3 Mikko M. Uimonen, BM¹; Sina M. Hulkkonen, MD², Jorma Ryhänen, MD, PhD²; Ville T.
4 Ponkilainen, BM¹; Arja H. Häkkinen, PhD^{3,4}; Jaro Karppinen, MD, PhD^{5,6}; Jussi P. Repo, MD,
5 PhD¹

6

7 *1. Department of Surgery, Central Finland Central Hospital, Jyväskylä, Finland.*

8 *2. Department of Hand Surgery, Helsinki University Hospital and University of Helsinki, Finland.*

9 *3. Health Sciences, Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä,*
10 *Finland.*

11 *4. Department of Physical Medicine, Central Finland Health Care District, Jyväskylä, Finland.*

12 *5. Medical Research Center Oulu, University of Oulu and Oulu University Hospital, Oulu, Finland*

13 *6. Finnish Institute of Occupational Health, Aapistie 1, 90220 Oulu, Oulu, Finland*

14

15 **Corresponding author:**

16 Mikko M. Uimonen

17 *Department of Surgery, Central Finland Central Hospital, Jyväskylä, Finland*

18 *Email: uimonen.mikko.m@student.uta.fi*

19

20 **Abstract**

21 **Background:** There is a lack of information on the measurement properties of patient-reported
22 upper extremity instruments and their association to health-related quality of life. The existing
23 upper extremity specific measures need further validation. The aim of this study was to measure and
24 compare the measurement properties and construct validity of the Disabilities of Arm, Shoulder,
25 and Hand (DASH) Instrument and the Michigan Hand Questionnaire (MHQ) using a heterogeneous
26 sample of patients with hand and wrist problems.

27 **Methods:** Two hundred-fifty consecutive patients visiting a general orthopedic outpatient clinic due
28 to hand/wrist problems were invited to participate in the study. Altogether 230 patients agreed to
29 participate and finally, a total of 193 (77%) participants provided sufficient patient-reported
30 outcome data and were included in the analysis. Participants with various hand or wrist complaints
31 completed the DASH, the MHQ, the EQ-5D-3L health-related quality of life (HRQoL) and Pain on
32 a visual analogue scale instruments. Grip and key pinch forces were measured. Scale targeting,
33 relatedness of demographics and construct validity of the DASH and the MHQ were assessed.

34 **Results:** Both the DASH and the MHQ had good targeting, but the DASH had wider coverage. The
35 convergence between the DASH and the MHQ was high. The DASH was more closely related to
36 HRQoL than the MHQ in terms of EQ-5D scores.

37 **Conclusions:** The DASH instrument appeared to measure hand function and disability from a
38 perspective of HRQoL superior to the MHQ among patients with heterogeneous hand and wrist
39 complaints.

40 **Introduction**

41 Increased interest in the outcomes of medical treatment has accelerated the development and use of
42 patient-reported outcome (PRO) instruments as a part of clinical outcome assessment¹⁻³. Extensive
43 research and validation of such instruments have shown them to be useful in evaluating and
44 comparing treatment outcomes ⁴. However, as the applicability of different PRO instruments may
45 vary in diverse study samples ². In order to more accurately select a proper instrument for the
46 population under examination ⁵, it is beneficial to understand the measurement properties of
47 different PRO instruments in head-to-head comparison. A systematic review of Van de Ven-
48 Stevens et al. ⁶ reviewed the clinimetric properties of 23 instruments for assessing hand function
49 after hand injury, including five PRO instruments. There were vital shortages in the reported
50 properties of all the surveyed PRO instruments. This finding refers to insufficient understanding of
51 the key features of these hand-specific measures.

52 The Disabilities of Arm, Shoulder and Hand (DASH) instrument ⁷ and the Michigan
53 Hand Outcomes Questionnaire (MHQ) ⁸ are widely adopted PRO instruments for evaluating the
54 performance and disability of upper limbs or hands ⁵. The clinimetric properties of the DASH have
55 been investigated using classical test theory ⁹ and the Rasch measurement theory ¹⁰⁻¹², and several
56 reports have assessed the validity of the MHQ among hand patients ^{8, 13-16}. A study by Dias et al.
57 compared three upper extremity-specific PRO instruments, the Patient evaluation measure (PEM),
58 the DASH, and the MHQ ¹⁷. It found the DASH and the MHQ to be valid and reliable for a sample
59 of patients with various wrist or finger complaints although there were shortages in construct
60 validity of all PRO instruments measured by correlation testing between the instruments scores and
61 hand symptom severity ¹⁷. Nonetheless, thus far there has been a lack of high-quality comparison of
62 these two PRO instruments with a perspective of health-related quality of life (HRQoL) outcomes.
63 Comparison of the association of the DASH and the MHQ to HRQoL provides valuable
64 information for researchers and clinicians dealing with hand and wrist problems. The results could

65 potentially guide us to choose the right instrument for assessing the function or disability of
66 patients.

67 The aim of this study was to measure and compare the scale targeting and construct
68 validity of the Finnish versions of the DASH and the MHQ and their association to HRQoL using a
69 heterogeneous sample of patients with hand and wrist problems in order to better understand the
70 clinimetrics of these two widely used PRO instruments.

71 **Materials and Methods**

72 In 2017, two hundred-fifty consecutive patients with hand and wrist problems treated at the general
73 orthopedic outpatient clinic in Länsi-Pohja Central Hospital in Kemi, Finland, were invited face-to-
74 face to participate in the study. The inclusion criteria were age of 18 years or above, full
75 understanding of spoken and written Finnish, a lack of cognitive disabilities, and the ability to give
76 signed informed consent to participating in the study. Overall, 230 invited patients were willing to
77 participate in the study. Of these 230 patients, 193 had completed all the PRO instruments
78 sufficiently and were included in the analysis, giving us an effective response rate of 77%.

79 The participants were clinically examined and cognitively debriefed by a surgeon
80 during their appointment at the hospital. Both hands were examined. The participants were asked to
81 complete the Finnish versions of the DASH instrument and the MHQ. Both PRO instruments are
82 available in Finnish ^{12, 18}. If both hands were affected, we selected the MHQ scores of the hand that
83 the participant considered worse for further analysis whereas the DASH does not distinguish
84 between hands and thus, we analyzed the DASH scores regardless of which hand was affected. The
85 participants also filled in the EQ-5D-3L generic health-related quality of life (HRQoL)
86 questionnaire. In addition, the participants were asked to complete a visual analogue scale (VAS)
87 pain instrument (“Place a line perpendicular to the line from 0 to 100 at the point that represents
88 their pain intensity with 0 representing ‘No pain at all’ and 100 representing ‘The worst imaginable

pain”). They returned the completed questionnaires during the final appointment at the hospital. We collected the clinical and demographic details. Grip and key pinch forces were measured using a Baseline hydraulic hand dynamometer together with a pinch gauge (White Plains, NY, USA). The Ethics Committee of the Northern Ostrobothnia Hospital District approved the study protocol. All the participants gave their written informed consent.

Disabilities of the arm, shoulder and hand (DASH)

The DASH is an upper limb-specific PRO instrument. It has been validated among patients with various hand and upper limb complaints^{9, 19-22}. The instrument was developed to evaluate patients’ disability and performance^{7, 9, 23}. The DASH has been widely adopted in use and is available in several languages¹⁸. It consists of two sections, one covers physical activities (23 items) and the other covers symptoms (7 items)²⁴. All items are on a Likert scale from 1 to 5, with higher score indicating higher disability over the preceding week²⁴. At least 27 items must be completed to enable calculation of the total score. The total score is scaled from 0 to 100 by dividing it by the number of items responded to and then subtracting one. Finally, the result is multiplied by 25²⁴. Cronbach’s alpha indicates high internal consistency at 0.97^{9, 25}. Test-retest reliability is estimated as high, as intraclass correlation coefficient (ICC) has been reported as 0.96 (95% CI: 0.93-0.98) with a three- to five-day interval between the assessments⁹. A difference of 10 points is considered the minimum significant change²⁶.

Michigan Hand Outcomes Questionnaire (MHQ)

The MHQ is a 37-item questionnaire, which was developed to evaluate the health state of patients with hand disorders⁸. The questionnaire includes six domains: Overall hand function, Activities of daily living (ADL), Pain, Work performance, Aesthetics, and patient’s Satisfaction with hand function⁸. The raw scores for each domain are calculated as a sum of the items in the category⁸. The raw score is converted to 0 to 100 scale, in which a higher score indicates better hand

113 performance⁸. The Pain domain is reversed, as a higher score indicates less pain⁸. The developers
114 of the questionnaire estimated Cronbach's alphas to be from 0.86 to 0.97 for all dimensions
115 indicating high internal consistency⁸. Test-retest reliability has proven to be high, as the ICC for all
116 dimensions ranges from 0.81 to 0.97⁸. The minimum significant change between the dimensions
117 has been presented as 17²⁷.

118 *EuroQol instrument (EQ-5D-3L)*

119 The EQ-5D-3L is a generic HRQoL instrument²⁸. It consists of five dimensions: Mobility, Self-
120 Care, Usual Activities, Pain/Discomfort, and Anxiety/Depression. Each dimension has three
121 response categories: no problems, some problems or severe problems. The instrument also includes
122 an item on overall health state, on a 0–100 visual analogue scale (EQ-VAS). The result is presented
123 as a five-digit number, a health state, and contains responses to all dimensions. This can be further
124 converted into a single index, varying from -0.011 to 1 in the Finnish version of EQ-5D-3L; a lower
125 score indicating poorer health-related quality of life.

126 **Statistical Methods**

127 The scores of all the instruments were converted to scale from 0 to 100. To obtain a parallel effect
128 direction, the scores of the EQ-5D-3L index, EQ-VAS and the MHQ were inverted by subtracting
129 the score from 100. After conversion, lower scores indicated better outcomes in all scales and
130 higher scores indicated worse outcomes. Clinical, demographic and questionnaire data are presented
131 as means with standard deviations (SD) or 95% confidence intervals (95% CI) or counts with
132 percentages. Scale targeting was assessed by evaluating score distributions, and floor and ceiling
133 effects. Floor or ceiling effects were considered confirmed if 15% of the participants scored the
134 minimum or maximum points². Relatedness of the DASH and the MHQ with demographic features
135 was examined by calculating Pearson correlation coefficients between the instruments' scores and
136 the ages of the participants, and by comparing the mean scores of male and female participants with

137 independent samples t-test. To produce comparable and generalizable results, the instruments
138 should ideally be unrelated of demographic factors.

139 Construct validity refers to the extent to which the PRO instrument measures what it is
140 supposed to. We evaluated construct validity of the DASH and the MHQ by assessing the
141 convergence between these two hand specific PRO instruments scores, association with objective
142 hand outcome measures, grip and key pinch forces, and relationship with HRQoL.

143 The relationship between grip and key pinch forces and the DASH and the MHQ was
144 examined by calculating Pearson correlation coefficients. To evaluate convergence of the DASH
145 and the MHQ, we calculated the Pearson correlation coefficients between the instruments. The
146 correlations were represented according to the Rule of Thumb for interpreting the size of the
147 correlation coefficient ²⁹: 0.00–0.30 negligible, 0.30–0.50 low, 0.50–0.70 moderate, 0.70–0.90 high
148 and 0.90–1.00 very high correlation. To further examine the underlying constructs measured by the
149 DASH and the MHQ from a perspective of HRQoL, we conducted Principal Component Analysis
150 (PCA) to reduce the variables to the main factors to enable us to examine their influence on the
151 DASH and MHQ scores. All EQ-5D scores and VAS Pain were included in the PCA. The
152 continuous variables (the EQ-5D index, EQ-VAS, and VAS Pain) were transformed into a
153 logarithmic scale ³⁰. The Principal Component (PC) was required to have an eigenvalue either equal
154 to or higher than 1 to be selected for further examination, in accordance with the Kaiser criteria ³¹.
155 We evaluated the distributions of the DASH and MHQ scores against the selected PCs. Locally
156 estimated scatterplot smoothing (LOESS) curves with 95% CIs were formed to illustrate the
157 correlation of the PCs with the DASH and MHQ scores. Similar distributions of the DASH and the
158 MHQ scores around the PCs indicate high convergence of the instruments from the perspective of
159 HRQoL related PCs.

160 We used HRQoL instrument EQ-5D and its subscales, EQ-VAS and Pain VAS as a

161 criterion measures when assessing relatedness of the hand outcome instruments of HRQoL. Pearson
162 correlation coefficients between the criterion measures and the DASH and the MHQ were
163 calculated. In addition, we used age- and gender-standardized linear regression analyses to identify
164 the strength of the influence of the DASH and MHQ scores on the perceived HRQoL of the
165 participants. Regression coefficients (beta, β) indicated how strongly the DASH and MHQ scores
166 influenced the criterion variables', EQ-5D scores and VAS Pain scores. The β was measured in
167 units of SD. The Cohen reference values for regression coefficient β were <0.1, <0.3 and <0.5 for
168 small, moderate and strong influence, respectively. In addition, we examined the association
169 between HRQoL and the instruments' scores by dividing the participants into subgroups according
170 to the EQ-5D index. The 25%, 50% and 75% percentiles were used as cut-off values between
171 subgroups. Residual analysis of linear regression of the DASH and the MHQ scores against the EQ-
172 5D index was conducted to assess the strength of linear relationship within the subgroups.

173 The statistical analyses were performed using R (version 1.1.453) and IBM SPSS 25.0
174 statistics software. The results of this study are interpreted and reported in adherence to the
175 Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement ³² and
176 the Consensus-based Standards for the selection of health status Measurement Instruments
177 (COSMIN) guidelines ³³.

178 **Results**

179 Table 1 presents the sociodemographic and clinical data of the participants. We observed no floor or
180 ceiling effects in either PRO instrument, although six (3.1%) participants scored the minimum score
181 in the DASH. Figure 1 shows the distributions of DASH and MHQ scores. The distribution of the
182 MHQ scores followed normal distribution, whereas the DASH score distribution was skewed
183 towards lower disability. However, the DASH scores covered the scale more comprehensively than
184 the MHQ scores. In addition, the EQ-5D score was distributed normally.

185 There was similar pattern of the DASH and the MHQ score distributions in diagnostic
186 subgroups. The subgroups of Distal radius fracture (mean DASH score = 41, MHQ score = 51),
187 Other fractures (DASH score = 36, MHQ score = 49) and Carpometacarpal 1 joint osteoarthritis
188 (DASH score = 39, MHQ score = 44) obtained the highest scores from both the DASH and the
189 MHQ indicating worst hand function, whereas the subgroups of Dupuytren's disease (mean DASH
190 score = 17, MHQ score = 32), Ulnar nerve entrapment (DASH score = 19, MHQ score = 32) and
191 Ganglion cyst (DASH score = 20, MHQ score = 33) obtained the lowest scores indicating, in turn,
192 best hand function.

193 The Pearson correlation coefficients between the instruments' scores and age showed
194 no significant correlations. On the other hand, female participants obtained higher scores from both
195 instruments as the mean DASH scores were 35 for female vs. 25 for male ($p = 0.003$) and the mean
196 MHQ scores were 42 vs. 38 ($p = 0.018$). Both instruments' scores correlated negatively with grip
197 and key pinch forces as the Pearson correlation coefficients of the DASH were -0.50 ($p < 0.001$) for
198 grip pinch and -0.45 ($p < 0.001$) for key pinch and of the MHQ were -0.53 ($p < 0.001$) for grip
199 pinch and -0.48 ($p < 0.001$) for key pinch forces.

200 Pearson correlation coefficient r between the DASH and MHQ scores was 0.75 (95%
201 CI = 0.68 to 0.82; $p < 0.001$), which indicates high convergent validity of the instruments (Figure
202 2). We observed moderate correlations between the DASH score and the EQ-5D index (Figure 3A),
203 Self-care, Usual activity, Pain/Discomfort, EQ-VAS and the VAS Pain scales (Table 2). We also
204 observed low but still notable correlations between MHQ score and the EQ-5D index (Figure 3B),
205 Self-care, Usual activity, Anxiety/Depression, EQ-VAS and VAS Pain, as well as between the
206 DASH score and Anxiety/Depression. All the correlations of the DASH with the reference
207 outcomes were higher than those of the MHQ.

208 Figure 4 presents the age- and sex-adjusted regression coefficient β of the DASH and
209 MHQ scores on the HRQoL measures. According to the Cohen reference values, all scores except

210 those of Mobility and Pain/Discomfort reached beta values of over 0.3 against the MHQ score,
211 indicating a moderate influence. The MHQ scores had the strongest influence on Usual activity ($\beta =$
212 0.49). The DASH scores had a strong influence on the EQ-5D index, Self-care, Usual activity, EQ-
213 VAS, and Pain VAS with betas over 0.5 and at least a moderate influence on all scores except that
214 of Mobility with betas over 0.3. The highest β was for the EQ-5D index ($\beta = 0.61$). All of the
215 associations were statistically significant (Figure 3). Overall, according to the regression
216 coefficients, the DASH had a notably stronger association with the HRQoL measures than the
217 MHQ. Thus, the HRQoL related construct validity of the DASH was considered higher than that of
218 the MHQ.

219 Figure 5 presents the MHQ and DASH scores in each HRQoL group. The DASH
220 scores show consistent improving trend when HRQoL improves from Very bad to Very good. In
221 addition, the MHQ scores show improving trend when HRQoL improves from Very bad to Good
222 but when HRQoL improves from Good to Very good, the trend seems to plateau. The residual
223 analysis of linear regression of the DASH and the MHQ scores against the EQ-5D index within the
224 HRQoL groups showed consistent strong relationship between the DASH and the EQ-5D index
225 across the HRQoL groups. However, the regression coefficient beta values showed that the strength
226 of the relationship decreases as the HRQoL improves as the beta values were 1.39 ($p < 0.001$), 1.00
227 ($p < 0.001$), 0.85 ($p < 0.001$) and 0.53 ($p < 0.001$) for Very bad, Bad, Good and Very good HRQoL,
228 respectively. On the other hand, there was significant linear relationship between the MHQ and the
229 EQ-5D index in only subgroup of Very bad HRQoL ($\beta = 0.47$, $p < 0.001$).

230 The principal component analysis generated two principal components that met the
231 Kaiser criteria³¹ of an eigenvalue over 1, and were thus included for further examination. The
232 eigenvalues of the first principal component (PC1) and the second principal component (PC2) were
233 3.16 and 1.26, respectively. PC1 explained 39.6% of the total variance of reference outcome
234 measures, whereas PC2 explained 15.7%. Overall, the selected PCs explained 55.3% of the total

235 variance. The loadings of the reference outcome measures on the PCs are presented in Figure 6.
236 According to the loading plot, the loading vectors can be divided to two groups. The first group
237 includes the EQ-5D index, EQ-VAS, Pain/Distress, and the VAS Pain scales, representing the
238 measures of general HRQoL and pain. The second group includes Self-care, Usual activities, and
239 Anxiety/Depression, which can be interpreted as representing the independency and performance of
240 participants. The mobility dimension was loaded towards the first group, but the loading was merely
241 weak. Figure 7A-D presents scatter plots and the LOESS curves with the 95% CIs of the DASH and
242 the MHQ against PCs. The distributions of the DASH and MHQ scores were similar for both PCs.

243 **Discussion**

244 The main finding of this study was that both the DASH and the MHQ instruments' scores strongly
245 correlated with each other as well as with the generic HRQoL instrument scores. Both instruments
246 had comparable measurement properties when the outcomes of various hand complaints were
247 evaluated (Table 3). However, in terms of its strong relationship with HRQoL, the DASH
248 instrument seemed to have more suitable measurement properties for the study sample than the
249 MHQ when evaluating HRQoL related outcomes after hand complaints.

250 The differences in the distributions of the DASH and MHQ scores pointed to wider
251 spread of the DASH scores than of the MHQ scores. Although, the distribution of the DASH scores
252 was skewed, it still covered the participants well, whereas the MHQ scores focused on more narrow
253 spectrum. On the other hand, we observed no floor or ceiling effects, indicating that both
254 instruments measured the spectrum of the hand complaints in the present sample well, which in turn
255 indicates good targeting for both instruments. Previous studies on the DASH and the MHQ have
256 also found similar distributions ^{9, 34-36}.

257 The evaluation of relatedness of the instruments with demographics showed similar
258 limitations for both instruments as female participants scored higher than male indicating worse

259 outcomes for female than male. The finding might be due to male participants having more strength
260 which in turn might help compensating the problems caused by hand complaints. On the other hand,
261 age was not associated with the instruments' scores. As there were similar pattern between the
262 DASH and the MHQ concerning the relatedness with sex, the issue should be taken into account
263 when these instruments are applied in further studies.

264 As expected, the correlation between the DASH and the MHQ scores was high, in line
265 with prior studies ^{17, 37}. Furthermore, the principal component analysis gave parallel results as the
266 DASH and the MHQ scores were distributed similarly for both recognized principal components.
267 Lastly, both instruments correlated significantly with grip and key pinch forces. The findings
268 indicated that both instruments measured the same constructs, which, in turn, proposes good
269 construct validity of each instruments.

270 Both instruments correlated strongly with the generic HRQoL scores (EQ-5D)
271 although the DASH correlations were stronger than those of the MHQ. In addition, the regression
272 analysis results indicated that the influence of the DASH score on HRQoL was stronger than that of
273 the MHQ. We also observed this in the comparison of the instruments' scores in the subgroups of
274 HRQoL, as the DASH scores shifted consistently in compliance with HRQoL while the MHQ
275 scores were unable to distinguish the differences when the HRQoL was good or very good. The
276 residual analysis of linear regression in HRQoL subgroups showed strong relationship between the
277 DASH score and the EQ-5D index across the subgroups while regarding the MHQ, the relationship
278 was found only in Very bad-subgroup. The findings propose that the DASH score was superior to
279 the MHQ in assessing the HRQoL outcomes in our sample. The highest correlation and regression
280 coefficients of each hand instrument were in the EQ-5D's Daily activity dimension and the EQ-5D
281 index, suggesting that the management of everyday tasks plays a key role in the assessment of hand
282 function. Prior studies have had parallel results, proposing that hand performance measures
283 associate with general physical function ^{8, 38}.

284 Although both instruments focus mostly on the same themes on hand outcomes, there
285 are differences in the perspectives of the instruments. While the subscales of the DASH focus on
286 the functions of daily living and symptoms of the hand, the MHQ has more comprehensive point of
287 view as the subscales also cover the management of daily chores and work as well as aesthetics of
288 the hand. However, despite the wider perspective of the MHQ on hand outcomes, the DASH score
289 was more closely related to HRQoL measured by EQ-5D than the MHQ score. According to our
290 findings, the DASH seems to capture the key aspects of hand outcomes related to HRQoL better
291 than the MHQ in a sample of patients with various hand complaints. On the other hand, there is a
292 fundamental difference in the scopes of the DASH and the MHQ. As the MHQ scores were
293 analyzed for the worse hand, the DASH does not distinguish between the hands. Thus, the DASH
294 scores might not be as sensitive to hand complaints if the unaffected hand compensate the issues
295 with affected hand. Hence, if the DASH score shows impairment, it is more likely, that the hand
296 issues affect daily living and HRQoL of the patient. This may explain the stronger association
297 between the DASH and the EQ-5D index than between the MHQ and the EQ-5D index. In addition,
298 the distributions of the DASH and the MHQ scores supported this idea as the MHQ scores showed
299 more hand issues than the DASH. Thus, the MHQ might be more sensitive if only one hand is
300 affected.

301 Our study had some limitations. First of all, we did not examine the responsiveness to
302 change of the instruments under observation, although it is an essential measurement property in the
303 validity evaluation of PRO instruments ³³. Second, our study was conducted in only one hospital
304 district. However, the strength of this study was its comprehensive, large sample of various hand
305 and wrist complaints. In addition, as the statistical methods it used have not previously been used
306 for evaluating and comparing the clinimetric properties of the DASH and the MHQ, our study
307 provides further insight into the construct validity of these two outcome instruments.

308 **Conclusion**

309 The scores of the DASH and the MHQ were highly correlated. The DASH scores had a stronger
310 relationship with the HRQoL outcomes. Thus, the DASH instrument appeared to be superior to the
311 MHQ in evaluating the outcomes from a perspective of HRQoL among patients with heterogeneous
312 hand and wrist complaints. On the other hand, the MHQ might be more specific instrument when
313 measuring performance of the affected hand.

314

315

316 **Figures and tables**

317 **Table 1.** Participants’ sociodemographic and clinical details.

	N=193
Age in years, mean (SD)	54 (15)
Women, n (%)	114 (59)
BMI, mean (SD)	28 (5.6)
Right hand affected, n (%)	109 (56)
Grip force of affected hand, kg, mean (SD)	29 (16)
Pinch force of affected hand, kg, mean (SD)	7.2 (2.9)
Number of diagnoses	
Carpal tunnel syndrome	82
Trigger finger	25
Distal radius fracture	20
Other fracture of the hand/wrist	20
Ganglion cyst	17
Dupuytren’s disease	17
Carpometacarpal 1 joint osteoarthritis	16
Ulnar nerve entrapment	3
Other	10
DASH score, mean (SD)	32 (22)
MHQ score, mean (SD)	40 (13)
EQ-5D index, mean (SD)	34 (17)
EQ-VAS, mean (SD)	26 (17)
VAS Pain, mean (SD)	36 (25)

318

319

320 **Table 2.** Pearson correlation coefficients between DASH and MHQ, and EQ-5D scores and VAS
 321 Pain.

Variable	DASH, r (95% CI)		MHQ, r (95% CI)	
EQ-5D index	0.64 (0.53 to 0.73)	***	0.43 (0.28 to 0.56)	***
EQ-5D dimensions				
Mobility	0.23 (0.07 to 0.39)	**	0.21 (0.06 to 0.36)	**
Self-care	0.56 (0.40 to 0.69)	***	0.40 (0.23 to 0.53)	***
Usual activity	0.56 (0.42 to 0.67)	***	0.48 (0.34 to 0.60)	***
Pain/Discomfort	0.50 (0.38 to 0.61)	***	0.25 (0.07 to 0.41)	**
Anxiety/Depression	0.43 (0.25 to 0.60)	***	0.34 (0.17 to 0.47)	***
EQ-VAS	0.51 (0.36 to 0.64)	***	0.44 (0.28 to 0.57)	***
VAS Pain	0.55 (0.44 to 0.65)	***	0.34 (0.14 to 0.50)	***

322 * p < 0.05, ** p < 0.01, *** p < 0.001

323

324

325 **Table 3.** Clinimetric features, hypotheses and conclusions for the DASH and the MHQ.

Feature	Hypothesis	DASH	MHQ
<i>Scale targeting</i>			
No Floor effect	Min score <15%	Confirmed	Confirmed
No Ceiling effect	Max score <15%	Confirmed	Confirmed
<i>Relatedness with demographics</i>	Non-significant or negligible associations with age and sex	Confirmed / Rejected	Confirmed / Rejected
<i>Construct validity</i>			
Convergence between the PRO instruments	Significant and at least low correlation with each other	Confirmed	
	Similar distributions for all recognized PCs	Confirmed	
Relatedness with objective hand function measures	Significant and at least low correlation with grip and pinch forces	Confirmed	Confirmed
Relatedness with HRQoL			
Correlation	Significant and at least low correlation with EQ-5D index	Confirmed	Confirmed
Regression	Significant and at least low linear dependency with EQ-5D index	Confirmed	Confirmed

326

327

328

329 **Figure 1.** Distributions of DASH and MHQ scores.

330

331

332 **Figure 2.** Correlation between DASH and MHQ.

333

Figure 3A-B. Correlations of the DASH and the MHQ with the EQ-5D index.

338

339

340

341

342

Figure 4. Age- and sex-adjusted regression coefficient β of predictors of DASH and MHQ in units of SD. Boxes represent mean scores and whiskers show 95% CIs. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

343

344

345 **Figure 5.** DASH and MHQ scores in HRQoL subgroups presented as 0 to 100 scores. Division in
346 subgroups is made by 25%, 50%, and 75% percentiles of EQ-5D index.

347

348

349

350 **Figure 6.** Loading vectors of reference outcome measures on PCs. The dots represent the
351 participants.

352

353

354 **Figure 7A-D.** Relationships of DASH and MHQ scores with PC1 (A-B) and PC2 (C-D). The
 355 LOESS curve shows the deterministic part of the variation in the data. The grey area around the
 356 curve describes the 95% CIs.

357

358

359

360

References

361

- 362 1. Guyatt G, Cairns J, Churchill D, et al. Evidence-based medicine: a new approach to
363 teaching the practice of medicine. *Jama* 1992; 268: 2420-2425.
- 364 2. McHorney CA and Tarlov AR. Individual-patient monitoring in clinical practice: are
365 available health status surveys adequate? *Quality of Life Research* 1995; 4: 293-307.
- 366 3. Garratt A, Schmidt L, Mackintosh A, et al. Quality of life measurement: bibliographic
367 study of patient assessed health outcome measures. *Bmj* 2002; 324: 1417.
- 368 4. Nelson EC, Eftimovska E, Lind C, et al. Patient reported outcome measures in
369 practice. *Bmj* 2015; 350: g7818.
- 370 5. MacDermid JC. Patient-reported outcomes: state-of-the-art hand surgery and future
371 applications. *Hand clinics* 2014; 30: 293-304.
- 372 6. van de Ven-Stevens LA, Munneke M, Terwee CB, et al. Clinimetric properties of
373 instruments to assess activities in patients with hand injury: a systematic review of the literature.
374 *Arch Phys Med Rehabil* 2009; 90: 151-169. 2009/01/22. DOI: 10.1016/j.apmr.2008.06.024.
- 375 7. Hudak PL, Amadio PC, Bombardier C, et al. Development of an upper extremity
376 outcome measure: the DASH (disabilities of the arm, shoulder, and hand). *American journal of*
377 *industrial medicine* 1996; 29: 602-608.
- 378 8. Chung KC, Pillsbury MS, Walters MR, et al. Reliability and validity testing of the
379 Michigan Hand Outcomes Questionnaire. *Journal of Hand Surgery* 1998; 23: 575-587.
- 380 9. Beaton DE, Katz JN, Fossel AH, et al. Measuring the whole or the parts?: Validity,
381 reliability, and responsiveness of the disabilities of the arm, shoulder and hand outcome measure in
382 different regions of the upper extremity. *Journal of Hand Therapy* 2001; 14: 128-142.
- 383 10. Forget NJ, Jerosch-Herold C, Shepstone L, et al. Psychometric evaluation of the
384 Disabilities of the Arm, Shoulder and Hand (DASH) with Dupuytren's contracture: validity
385 evidence using Rasch modeling. *BMC musculoskeletal disorders* 2014; 15: 361.
- 386 11. Franchignoni F, Giordano A, Sartorio F, et al. Suggestions for refinement of the
387 Disabilities of the Arm, Shoulder and Hand Outcome Measure (DASH): a factor analysis and Rasch
388 validation study. *Archives of physical medicine and rehabilitation* 2010; 91: 1370-1377.
- 389 12. Ikonen J HS, Ryhänen J, Häkkinen AH, Karppinen J, Repo JP. . Validation of the
390 Finnish version of the Disabilities of Arm, Shoulder and Hand questionnaire using Rasch model. .
391 *Submitted manuscript*.
- 392 13. Wehrli M, Hensler S, Schindele S, et al. Measurement Properties of the Brief
393 Michigan Hand Outcomes Questionnaire in Patients With Dupuytren Contracture. *The Journal of*
394 *hand surgery* 2016; 41: 896-902.
- 395 14. Kotsis SV, Lau FH and Chung KC. Responsiveness of the Michigan Hand Outcomes
396 Questionnaire and physical measurements in outcome studies of distal radius fracture treatment.
397 *The Journal of hand surgery* 2007; 32: 84-90.
- 398 15. Sambandam SN, Priyanka P, Gul A, et al. Critical analysis of outcome measures used
399 in the assessment of carpal tunnel syndrome. *International orthopaedics* 2008; 32: 497-504.
- 400 16. Chatterjee JS and Price PE. Comparative responsiveness of the Michigan Hand
401 Outcomes Questionnaire and the Carpal Tunnel Questionnaire after carpal tunnel release. *The*
402 *Journal of hand surgery* 2009; 34: 273-280.
- 403 17. Dias J, Rajan R and Thompson J. Which questionnaire is best? The reliability, validity
404 and ease of use of the Patient Evaluation Measure, the Disabilities of the Arm, Shoulder and Hand

405 and the Michigan Hand Outcome Measure. *Journal of Hand Surgery (European Volume)* 2008; 33:
406 9-17.

407 18. Hacklin E TS, Madanat R, Strandberg N, Aro H. . DASH-kyselykaavakkeen
408 suomentaminen ja kulttuuriadaptaatio. *Suomen Ortopedia ja Traumatologia* 2009; 32: 252-254.

409 19. Aktekin LA, Eser F, Başkan BM, et al. Disability of Arm Shoulder and Hand
410 Questionnaire in rheumatoid arthritis patients: relationship with disease activity, HAQ, SF-36.
411 *Rheumatology international* 2011; 31: 823-826.

412 20. Huisstede BM, Feleus A, Bierma-Zeinstra SM, et al. Is the disability of arm, shoulder,
413 and hand questionnaire (DASH) also valid and responsive in patients with neck complaints. *Spine*
414 2009; 34: E130-E138.

415 21. Buntragulpoontawee M, Phutrit S, Tongprasert S, et al. Construct validity, test-retest
416 reliability and internal consistency of the Thai version of the disabilities of the arm, shoulder and
417 hand questionnaire (DASH-TH) in patients with carpal tunnel syndrome. *BMC research notes*
418 2018; 11: 208.

419 22. Wong JY, Fung BK, Chu MM, et al. The use of Disabilities of the Arm, Shoulder, and
420 Hand Questionnaire in rehabilitation after acute traumatic hand injuries. *Journal of Hand Therapy*
421 2007; 20: 49-56.

422 23. Schmitt JS and Di Fabio RP. Reliable change and minimum important difference
423 (MID) proportions facilitated group responsiveness comparisons using individual threshold criteria.
424 *J Clin Epidemiol* 2004; 57: 1008-1018. 2004/11/06. DOI: 10.1016/j.jclinepi.2004.02.007.

425 24. Kennedy CA. *The DASH and quickDASH outcome measure user's manual*. Institute
426 for Work & Health, 2011.

427 25. Marx RG, Bombardier C, Hogg-Johnson S, et al. Clinimetric and psychometric
428 strategies for development of a health measurement scale. *Journal of clinical epidemiology* 1999;
429 52: 105-111.

430 26. Gummesson C, Atroshi I and Ekdahl C. The disabilities of the arm, shoulder and hand
431 (DASH) outcome questionnaire: longitudinal construct validity and measuring self-rated health
432 change after surgery. *BMC Musculoskeletal Disorders* 2003; 4: 11. journal article. DOI:
433 10.1186/1471-2474-4-11.

434 27. Marks M, Audigé L, Herren DB, et al. Measurement properties of the German
435 Michigan Hand Outcomes Questionnaire in patients with trapeziometacarpal osteoarthritis. *Arthritis*
436 *care & research* 2014; 66: 245-252.

437 28. Brooks R and Group E. EuroQol: the current state of play. *Health policy* 1996; 37: 53-
438 72.

439 29. Mukaka MM. A guide to appropriate use of correlation coefficient in medical
440 research. *Malawi Medical Journal* 2012; 24: 69-71.

441 30. Venables WN and Ripley BD. *Modern Applied Statistics with S-PLUS*. Springer-
442 Verlag New York, 1997.

443 31. Kaiser HF. The Application of Electronic Computers to Factor Analysis. *Educational*
444 *and Psychological Measurement* 1960; 20: 141-151. DOI: 10.1177/001316446002000116.

445 32. Von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of
446 Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting
447 observational studies. *PLoS medicine* 2007; 4: e296.

448 33. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the
449 methodological quality of studies on measurement properties of health status measurement
450 instruments: an international Delphi study. *Quality of Life Research* 2010; 19: 539-549. DOI:
451 10.1007/s11136-010-9606-8.

452 34. Atroshi I, Gummesson C, Andersson B, et al. The disabilities of the arm, shoulder and
453 hand (DASH) outcome questionnaire: Reliability and validity of the Swedish version evaluated in

454 176 patients. *Acta Orthopaedica Scandinavica* 2000; 71: 613-618. DOI:
455 10.1080/000164700317362262.
456 35. Waljee JF, Chung KC, Kim HM, et al. Validity and responsiveness of the Michigan
457 Hand Questionnaire in patients with rheumatoid arthritis: a multicenter, international study.
458 *Arthritis care & research* 2010; 62: 1569-1577.
459 36. McMillan CR and Binhammer PA. Which outcome measure is the best? Evaluating
460 responsiveness of the Disabilities of the Arm, Shoulder, and Hand Questionnaire, the Michigan
461 Hand Questionnaire and the Patient-Specific Functional Scale following hand and wrist surgery.
462 *Hand* 2009; 4: 311-318.
463 37. Horng Y-S, Lin M-C, Feng C-T, et al. Responsiveness of the Michigan Hand
464 Outcomes Questionnaire and the Disabilities of the Arm, Shoulder, and Hand questionnaire in
465 patients with hand injury. *The Journal of hand surgery* 2010; 35: 430-436.
466 38. Gay RE, Amadio PC and Johnson JC. Comparative responsiveness of the disabilities
467 of the arm, shoulder, and hand, the carpal tunnel questionnaire, and the SF-36 to clinical change
468 after carpal tunnel release. *The Journal of hand surgery* 2003; 28: 250-254.
469
470