Concise report

Measurement properties of the Health Assessment Questionnaire Disability Index for generalized osteoarthritis

Nienke Cuperus¹, Elien A. M. Mahler¹, Theodora P. M. Vliet Vlieland², Thomas J. Hoogeboom³ and Cornelia H. van den Ende¹

Abstract

Objective. Generalized OA (GOA) is highly prevalent in OA. Individuals with GOA typically suffer from limitations of both upper and lower extremity function, yet we lack a validated instrument to assess their activity limitations. An appropriate instrument might be the HAQ Disability Index (HAQ-DI). Therefore the aim of this study was to evaluate the measurement properties of the HAQ-DI in GOA.

Methods. Data were used from a randomized controlled trial comparing the effectiveness of two multidisciplinary treatment programmes for patients with GOA. One hundred and thirty-seven of 147 included patients completed a standardized set of questionnaires before and after treatment. Interpretability, validity, reliability and responsiveness of the HAQ-DI were assessed using the Consensus-Based Standards for the Selection of Health Status Measurement Instruments checklist (COSMIN).

Results. Floor and ceiling effects were present. The content validity was questionable since the HAQ-DI encompasses activities that are either not relevant or too easy to perform as judged by patients and experts. Construct validity was good since 90% of the hypotheses were confirmed. Factor analysis confirmed the unidimensionality of the HAQ-DI (root mean square error of approximation = 0.057, $\chi^2/df$ ratio = 1.48). Cronbach’s $\alpha$ was 0.90, confirming internal consistency and the ICC was 0.81, reflecting good reliability. The minimal important change was 0.25 and the smallest detectable change was 0.60. We could not establish the responsiveness of the HAQ-DI.

Conclusion. The HAQ-DI showed good construct validity, internal consistency and reliability, whereas its content validity and responsiveness were limited. We recommend updating the items of the HAQ-DI in future research focusing on functional limitations in GOA.


Key words: generalized osteoarthritis, functional limitations, HAQ-DI, activity limitations, measurement properties.

Introduction

Patient-reported assessments of functional ability have emerged as an integral part of clinical practice and research in OA. Numerous self-administered instruments (such as the WOMAC) are available to assess activity limitations in OA [1]. However, many of these questionnaires focus on a specific localization of OA (e.g. hands, hips or knees), limiting their use in the assessment of patients with generalized OA (GOA). Individuals with GOA typically suffer from OA in multiple joints [2], resulting in limitations of both upper and lower extremity function. To comprehensively
assess activity limitations in people with GOA, instruments should contain items of both upper and lower extremity function. A growing body of evidence suggests that GOA is highly prevalent in patients with OA [3, 4], yet we lack a validated instrument to assess their activity limitations.

An appropriate instrument to measure activity limitations in patients with GOA might be the Stanford Health Assessment Questionnaire Disability Index (HAQ-DI), as it comprises activities of fine movement of the upper extremities, locomotor activities of the lower extremities and activities that involve both upper and lower extremities [5]. Since the HAQ-DI has been designed for individuals with RA, it comprises activities frequently impaired in RA patients. The HAQ-DI could be a useful instrument to assess and monitor functional limitations in patients with GOA as well, as the two conditions have many similarities. In both RA and GOA, pain and stiffness are major symptoms and multiple joints in both the upper and lower extremities are affected. However, there are also differences, such as the pattern of joint involvement. Functional disabilities might therefore differ between the two conditions. Therefore an assessment of the measurement properties of the HAQ-DI in patients with GOA seems warranted. The aim of this study was to evaluate the measurement properties (i.e. interpretability, validity, reliability and responsiveness) of the HAQ-DI in patients with GOA.

Methods

Study design and participants

The current study was part of a randomized controlled trial comparing the effectiveness of two non-pharmacological multidisciplinary treatment programmes in patients with GOA. The Institutional Review Board of the Radboud University Nijmegen Medical Centre approved the study and patients signed informed consent. The complete study design has been described previously (Dutch Trial Register NTR2137) [2].

Patients clinically diagnosed with GOA were included when they had at least two objective signs indicating OA in two or more joint areas (malalignment, crepitation, limited range of motion, palpable osteophytes or nodules or radiographic signs including the presence of joint space narrowing and/or osteophytes); complaints in three or more out of eight joint areas; and were limited in daily activities (HAQ-DI score \( \geq 0.5 \)). Patients diagnosed with another rheumatic disease were excluded [2].

Data collection and procedure

Participants completed a standardized set of questionnaires prior to the start of the treatment (baseline) and at 6, 26 and 52 weeks. Data from baseline and the 6-week follow-up were used. Patients were included if they completed the baseline measurement. The recommendations of the Consensus-based Standards for the Selection of Health Status Measurement Instruments (COSMIN) [6] and the quality criteria for measurement properties proposed by Terwee et al. [7] were followed to assess the measurement properties of the HAQ-DI.

Outcome measures

Participants completed the Dutch consensus HAQ-DI [8], determining difficulties with the performance of 20 daily activities scored from 0 (without difficulty) to 3 (unable to do) and classified into eight categories: dressing and grooming, arising, eating, walking, personal hygiene, reaching, gripping and usual activities [5]. The total HAQ-DI score represents a disability index ranging from 0 (no disability) to 3 (very severe disability) [5]. We did not correct HAQ-DI scores for the use of assistive devices, preventing overestimation of the patients’ functional disability.

In addition, we measured quality of life [36-item Short Form Health Survey (SF-36) and the EuroQol five-dimension questionnaire (EQ-5D)], fatigue [Checklist of Individual Strength (CIS)], self-efficacy [General Self-Efficacy Scale (GSES)] and illness cognition [Illness Cognition Questionnaire (ICQ)]. Patient-specific complaints [Patient Specific Complaints questionnaire (PSK)] was measured by asking patients to select the three most important GOA-related activity limitations and to score their severity on a 0–10 scale. In the follow-up questionnaire, a seven-point Likert transition question was used to measure changes in daily function ranging from completely recovered to vastly worsened [9].

Interpretability

The scale and category scores of the HAQ-DI were assessed for normality and missing data. We assessed floor and ceiling effects at baseline for each HAQ-DI category, which were considered present if \( >15\% \) of patients scored the worst (3) or best (0) possible score. We determined the minimal important change (MIC) score using an anchor-based method [10].

Content validity

We assessed whether the HAQ-DI items adequately represent daily function from both the patient’s and health professional’s perspective. We compared the International Classification of Functioning, Disability and Health (ICF) codes of all HAQ-DI items presented in a study by Stamm et al. [11] with the ICF codes of an aggregated list (unpublished observations) representing the 10 most important activity limitations reported by patients with GOA. Content validity was arbitrarily considered good when all 10 activity limitations were covered by the HAQ-DI. Secondly, we asked 17 health professionals experienced with GOA to estimate the relevance of each HAQ-DI item, expressed as the percentage of GOA patients having actual difficulty performing that item. Items were considered not relevant when health professionals expected items were difficult to perform for \( \leq 50\% \) of patients.

Construct validity and responsiveness

We assessed construct validity by testing predefined hypotheses about expected correlations between HAQ-DI scores and scores on related measures (convergent), unrelated measures (discriminant) or differences between
groups (discriminative) [7]. Since we consider responsiveness as a measure of longitudinal validity, we assessed responsiveness in analogy with construct validity by testing predefined hypotheses using change scores [7]. In this way, responsiveness is independent of the magnitude of a change but measures the quality of the measurement instrument. In addition, for our responsiveness analysis we formulated two hypotheses on the size of the area under the receiver operating characteristic curve (AUC) [7]. An expert group of researchers, health professionals and a rheumatologist independently formulated hypotheses that were discussed during a consensus meeting. Construct validity and responsiveness were considered positive if >75% of the hypotheses were confirmed [7].

Internal consistency and reliability

Confirmatory factor analysis (CFA) was performed using the following thresholds for a good model fit: root mean square error of approximation (RMSEA) ≤ 0.06, Comparative Fit Index (CFI) ≥ 0.95 and χ²/df ratio ≤ 3 [12]. Internal consistency was considered adequate if Cronbach’s α was ≥ 0.7. Test-retest reliability was acceptable when the intraclass correlation coefficient (ICC2,1) was ≥ 0.70 in stable patients (i.e. patients reported no change on the transition question). We determined the limits of agreement as a parameter of agreement [13] and we quantified the smallest detectable change (SDC = S.E.M. × 1.96 × √2) [7].

Statistical analysis

Pearson or Spearman rank correlation coefficients (when appropriate) were computed to test the hypotheses. The CFA was conducted using LISREL software (Scientific Software International, Skokie, IL, USA). All other analyses were performed using STATA 10.1 (StataCorp, College Station, TX, USA).

Results

Patient characteristics

A total of 147 patients completed the baseline measurement (93%), of whom 137 completed the follow-up measurement (93%). The majority of patients remained stable in their daily function after 6 weeks (Table 1).

Interpretability

The item getting in or out of bath was returned with >10% missing values. HAQ-DI scores were normally distributed with a mean score of 1.27 (s.d. 0.50) at baseline and 1.20 (s.d. 0.47) at follow-up. Ceiling effects were present in the categories eating, dressing and gripping, and a floor effect was present in the category hygiene. The MIC for improvement was 0.25 points.

Content validity

Seven of 10 ICF codes representing the most relevant activity limitations for patients with GOA were covered by the HAQ-DI items. Activities that were not covered were maintaining a body position (d415), recreation and leisure (d920) and driving (d475). Health professionals considered nine items of the HAQ-DI relevant for 50% of GOA patients whereas 11 items were considered relevant for >50% of patients. The least relevant items were lifting a cup to your mouth, shampoo your hair, wash your body and opening car doors. The most relevant items were getting down a 5 pound object from above your head, opening a new milk carton and doing chores.

Construct validity and responsiveness

The HAQ-DI showed good construct validity since 9 out of 10 hypotheses (90%) were confirmed (Table 2). We were unable to confirm the hypothesized strong correlation between HAQ-DI scores and the EQ-5D item usual activities since we found a correlation of 0.23. Four out of 10 hypotheses (40%) to assess the responsiveness of the HAQ-DI were confirmed. The correlation between change in HAQ-DI score and change in SF-36 physical function score was weak (−0.20), whereas we hypothesized a moderate correlation because of the associated constructs. We could not confirm the hypotheses on the AUC.

Internal consistency and reliability

The initial CFA (with uncorrelated measurement errors between categories) did not meet adequate fit criteria. Modification indices provided by LISREL suggested a correlated error term between eating and gripping. The modified model resulted in a good fit to the data: RMSEA = 0.057, CFI = 0.98 and χ²/df ratio = 1.48. Cronbach’s α was 0.90 (lower limit of the 95% CI 0.88),
Table 2

Results of predefined hypotheses to assess the construct validity and longitudinal validity (responsiveness) of the HAQ-DI in generalized OA

<table>
<thead>
<tr>
<th>Hypothesis construct validity</th>
<th>Results</th>
<th>Confirmed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. There is a strong correlation ($r &gt; -0.5$) between HAQ-DI and SF-36 physical function</td>
<td>$-0.68$ ($-0.76$ to $-0.58$)</td>
<td>Yes</td>
</tr>
<tr>
<td>2. The correlation between HAQ-DI and SF-36 physical function is stronger than the correlation between HAQ-DI and SF-36 pain</td>
<td>$-0.68$ ($-0.76$ to $-0.58$)</td>
<td>Yes</td>
</tr>
<tr>
<td>3. There is a strong correlation ($r &gt; 0.5$) between HAQ-DI and the EQ-5D item usual activities</td>
<td>$0.23$ ($0.07$ to $0.37$)</td>
<td>No</td>
</tr>
<tr>
<td>4. There is a moderate correlation ($0.3 &lt; r &lt; 0.5$) between HAQ-DI and PSK</td>
<td>$0.34$ ($0.19$ to $0.48$)</td>
<td>Yes</td>
</tr>
<tr>
<td>5. There is a weak correlation ($r &lt; -0.3$) between HAQ-DI and SF-36 mental function</td>
<td>$-0.11$ ($-0.27$ to $0.05$)</td>
<td>Yes</td>
</tr>
<tr>
<td>6. There is a weak correlation ($r &lt; 0.3$) between HAQ-DI and fatigue</td>
<td>$0.27$ ($0.11$ to $0.41$)</td>
<td>Yes</td>
</tr>
<tr>
<td>7. There is a weak correlation ($r &lt; 0.3$) between HAQ-DI and sociodemographic characteristics</td>
<td>$-0.27$ ($-0.41$ to $-0.11$)</td>
<td>Yes</td>
</tr>
<tr>
<td>8. There is a weak correlation ($r &lt; -0.3$) between HAQ-DI and self-efficacy</td>
<td>$-0.05$ ($-0.21$ to $0.12$)</td>
<td>Yes</td>
</tr>
<tr>
<td>9. There is a weak correlation ($r &lt; -0.3$) between HAQ-DI and illness cognitions subscale acceptance</td>
<td>$-0.18$ ($-0.33$ to $-0.01$)</td>
<td>Yes</td>
</tr>
<tr>
<td>10. Patients with musculoskeletal complaints in more than five joint areas have higher HAQ-DI scores than patients with musculoskeletal complaints in three to five joint areas</td>
<td>$0.38$ ($0.22$ to $0.54$)</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Confirmed, n/n (%) | 9/10 (90)

Hypothesis longitudinal validity (responsiveness) | Results | Confirmed |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The HAQ-DI can distinguish patients who reported major improvements (much improved, completely recovered) from those who reported no important changes (slightly improved, no change, slightly worse) based on the transition question (AUC &gt; 0.7)</td>
<td>AUC = 0.69</td>
<td>No</td>
</tr>
<tr>
<td>2. The HAQ-DI can distinguish patients who were improved (slightly improved, completely recovered) from those who were unchanged (no change) based on the transition question (AUC &gt; 0.7)</td>
<td>AUC = 0.53</td>
<td>No</td>
</tr>
<tr>
<td>3. There is a moderate correlation ($0.3 &lt; r &lt; 0.5$) between ΔHAQ-DI and the transition question</td>
<td>$0.14$ ($-0.03$ to $0.30$)</td>
<td>No</td>
</tr>
<tr>
<td>4. There is a moderate correlation ($0.3 &lt; r &lt; -0.5$) between ΔHAQ-DI and ΔSF-36 physical function</td>
<td>$-0.20$ ($0.36$ to $-0.03$)</td>
<td>No</td>
</tr>
<tr>
<td>5. The correlation between ΔHAQ-DI and ΔSF-36 physical function is stronger than the correlation between ΔHAQ-DI and ΔSF-36 pain</td>
<td>$-0.20$ ($0.36$ to $-0.03$)</td>
<td>No</td>
</tr>
<tr>
<td>6. There is a weak correlation ($r &lt; 0.3$) between ΔHAQ-DI and ΔSF-36 mental function</td>
<td>$-0.05$ ($-0.22$ to $0.12$)</td>
<td>Yes</td>
</tr>
<tr>
<td>7. There is a weak correlation ($r &lt; 0.3$) between ΔHAQ-DI and Δenergy</td>
<td>$0.09$ ($-0.08$ to $0.26$)</td>
<td>Yes</td>
</tr>
<tr>
<td>8. There is a weak correlation ($r &lt; -0.3$) between ΔHAQ-DI and Δfatigue</td>
<td>$0.13$ ($-0.30$ to $0.04$)</td>
<td>Yes</td>
</tr>
<tr>
<td>9. There is a weak correlation ($r &lt; -0.3$) between ΔHAQ-DI and Δillness cognitions subscale acceptance</td>
<td>$-0.14$ ($-0.30$ to $0.03$)</td>
<td>Yes</td>
</tr>
<tr>
<td>10. The effect size of the HAQ-DI in the group of patients with major improvements is &gt;0.8</td>
<td>0.49</td>
<td>No</td>
</tr>
</tbody>
</table>

Confirmed, n/n (%) | 4/10 (40)

*Values are correlations (95% CI) unless indicated otherwise. **Significantly higher correlation, $P < 0.05$. ***Sociodemographic characteristics included BMI, age, education level, work and family status. The strongest correlation is depicted, i.e. education level. ^Mean difference in scores (95% CI), $P < 0.01$. °Difference between correlation is not significant, $P > 0.05$. °Effect size = mean change in HAQ-DI/SD. b: mean baseline HAQ-DI. AUC: area under the receiver operating characteristic curve; Δ: change in; EQ-5D: EuroQol five-dimensions questionnaire; HAQ-DI: Health Assessment Questionnaire Disability Index; PSK: patient-specific complaints; SF-36: 36-item Short Form Health Survey.

Discussion

This is the first study to comprehensively assess the measurement properties of the HAQ-DI in patients clinically diagnosed with GOA. The results showed a questionable content validity. The construct validity, internal
consistency and reliability were confirmed; however, the responsiveness was limited.

Based on the results, we need to question the content validity of the HAQ-DI. Important activities frequently impaired in patients with GOA were not assessed by the HAQ-DI. Moreover, the ceiling effects suggest that the HAQ-DI measures activities that are either not relevant or too easy to perform for patients with GOA, a finding supported in RA. In particular, these activities are lifting a cup to your mouth, shampooing your hair and opening car doors, which were also rated by the health professionals as least relevant. The activity getting in or out of the bath also seems less relevant given the high number of missing values, a result supported in RA [14]. An explanation for our findings might be that certain activities of the HAQ-DI are ‘outdated’, since the questionnaire was developed three decades ago [5], a period during which some activities have become easier to perform due to technological and social developments.

We were not able to confirm the responsiveness of the HAQ-DI. This might be explained by the reported floor and ceiling effects in four of eight HAQ-DI categories, a finding supported by other studies [15]. Several studies examining the responsiveness of the HAQ-DI in rheumatic diseases showed conflicting results and/or used different approaches [14–16]. There is no consensus on which method is best to assess responsiveness; however, it has been recommended that anchor-based methods should be used, as traditionally accepted methods (e.g. effect size) measure the magnitude of change rather than longitudinal validity [17]. An important limitation in our responsiveness analysis is the very small number of patients with major improvements, which prevents firm conclusions. In addition, there is debate regarding the use of transition questions since they are biased by patients’ current status, particularly as the time span increases [18]. To limit recall bias, we decided to use only the 6-week follow-up measurement.

The MIC of 0.25 in this study is comparable to that of patients with RA [19]. The SDC was 0.60, indicating that important changes cannot be distinguished from measurement error in individuals. The HAQ-DI is therefore not suitable for monitoring individual GOA patients in daily clinical practice.

Some limitations need to be addressed. One could question the extent to which our study population adequately represents GOA patients. We used a pragmatic GOA definition, as there is no widely accepted definition [20]. Moreover, we did not assess the WOMAC. We considered the face validity of the WOMAC inadequate, as it only contains items on lower extremity function. However, in future research it would be worthwhile comparing the measurement properties of existing physical function measures in GOA, such as the WOMAC or the SF-36, before developing new measurement instruments assessing functional limitations in patients with GOA.

In conclusion, the HAQ-DI showed good construct validity and reliability. Given the unsatisfactory content validity, we recommend an update of the items of the HAQ-DI when using the HAQ-DI in future clinical practice and research focusing on functional limitations in GOA. This update might also be worthwhile for RA and all other rheumatic diseases.

**Rheumatology key messages**

- The HAQ-DI showed good construct validity, internal consistency and reliability in patients with generalized OA.
- The content validity is questionable; items are either not relevant or too easy to perform for generalized OA patients.
- We recommend an update of the items of the HAQ-DI when focusing on functional limitations in generalized OA.

**Funding:** None.

**Disclosure statement:** The authors have declared no conflicts of interest.

**References**


7 Terwee CB, Bot SD, de Boer MR et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34–42.


