

## Concise report

## Comparison of measures of functional disability in patients with gout

Peter M. ten Klooster<sup>1</sup>, Martijn A. H. Oude Voshaar<sup>1</sup>, Erik Taal<sup>1</sup> and Mart A. F. J. van de Laar<sup>1,2</sup>

## Abstract

**Objective.** To compare the measurement properties of the HAQ disability index (HAQ-DI), HAQ-II and short form 36 physical functioning scale (PF-10) in patients with gout.

**Methods.** A cross-sectional sample of 97 patients with gout completed all three measures. Reliability was assessed by examining the internal consistency of the scales. Validity was assessed by testing for expected intercorrelations and associations with other aspects of health status and the ability to discriminate between patients with different levels of general health. Additionally, distributional properties were examined.

**Results.** All three measures demonstrated high reliability (Cronbach's  $\alpha \geq 0.93$ ), strong intercorrelations ( $r \geq 0.75$ ), and the expected pattern of external correlations. The HAQ-DI and HAQ-II performed somewhat better in discriminating between patients. However, both demonstrated ceiling effects of 34.0 and 25.8%, respectively, compared with only 7.2% of the patients scoring no disability on the PF-10.

**Conclusions.** The HAQ-DI, HAQ-II and PF-10 demonstrated similar and adequate reliability and validity for measuring functional disability in patients with gout. The large ceiling effects of both HAQ versions, however, may point to limited content validity and responsiveness to change. Further research should examine whether current instruments cover all aspects of physical functioning relevant to patients with gout.

**Key words:** Outcome measures, Gout, Functional disability, Health assessment questionnaire, SF-36, Psychometrics.

## Introduction

Functional disability is an important domain for outcome measurement in gout [1, 2]. However, there is a paucity of validated outcome measures for patients with gout in general [1, 3], and only few studies have specifically examined the psychometric properties of physical disability measures.

Most studies that have measured disability in gout have used the generic HAQ disability index (HAQ-DI) or short form 36 (SF-36) physical functioning scale (PF-10). However, only one study briefly examined the

measurement properties of the PF-10 in patients with gout and did not report any detailed results [4]. The HAQ-DI was more thoroughly validated in two recent studies, which both concluded that it was a valid and reliable measure of gout-related functional disability [5, 6]. However, high ceiling effects were observed with, respectively, 42.2 and 20.5% of the patients scoring no disability. Moreover, similar high ceiling effects were found in observational studies of both chronic stable gout [7] and treatment-failure gout patients [8]. High ceiling effects indicate that items may be missing at the upper end of the scale, suggesting limited content validity and responsiveness [9]. Consequently, Taylor *et al.* [6] suggested that the more robust HAQ-II should be evaluated as a possible alternative scale for measuring functional status in gout.

To date, no studies have simultaneously examined the measurement properties of the HAQ-DI, HAQ-II and PF-10 for measuring functional disability in gout. Therefore, the aim of this study was to compare the reliability and validity of all three measures in a cross-sectional sample of patients with gout.

<sup>1</sup>Department of Psychology and Communication of Health and Risk, Institute for Behavioural Research, University of Twente and  
<sup>2</sup>Department of Rheumatology, Medisch Spectrum Twente, Enschede, The Netherlands.

Submitted 14 September 2010; revised version accepted 20 October 2010.

Correspondence to: Peter M. ten Klooster, Institute for Behavioural Research, Faculty of Behavioural Sciences, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands.  
E-mail: p.m.tenklooster@utwente.nl

## Methods

### Patients

Respondents for the study were recruited during several waves of data collection in the period between 2005 and 2008 at our outpatient clinic for rheumatology in Enschede, the Netherlands. Details about the study design have been reported elsewhere [10]. According to the Dutch law for medical research with humans [Wet Medisch-wetenschappelijk onderzoek met Mensen (WMO)], approval by an ethics committee was not necessary for this survey study. In total, 102 patients with gout were included, of whom 97 completed all three measures. The diagnosis of gout was confirmed by identification of urate crystals in SF or material aspirated from tophi in 80% of the patients and by elevated serum urate (normal range 0.20–0.40 mmol/l) in 20% of the patients.

### Measures

The HAQ-DI contains 20 items measuring physical disabilities over the past week in eight categories of daily living [11]. Items are scored on a 4-point rating scale from 0 (without any difficulty) to 3 (unable to do). A total score is calculated by averaging the highest score in each category (corrected for the use of aids and devices) if at least six categories are completed.

The HAQ-II was developed to address some of the problems of the HAQ-DI in RA, including its marked ceiling effect in patients with lower levels of disability [12]. Like the HAQ-DI, the HAQ-II is a generic questionnaire in the sense that it assesses physical disability in general and does not focus on specific disease-associated impairments. It consists of 10 items, 5 of which stem directly from the HAQ-DI. The HAQ-II is scored by taking the mean of the items if at least eight items are completed, also resulting in a score from 0 to 3, with higher scores indicating more disability.

The generic SF-36 version 2.0 is a multidimensional questionnaire assessing different aspects of health represented in eight scales [13]. Its 10-item physical functioning scale measures current limitations in a variety of physical activities [14]. Items are scored from 1 (yes, limited a lot) to 3 (no, not limited at all). Scores on the PF-10 items are summed and linearly transformed to range between 0 and 100, with higher scores indicating better functioning.

Additionally, patient-reported pain and general health were assessed on 11-point numerical rating scales (NRSs) ranging from 0 ('no pain' or 'very good') to 10 ('unbearable pain' or 'very bad').

### Analyses

Missing values for individual items were low, ranging from 0 to 5.2% for the HAQ-DI, from 0 to 4.1% for the HAQ-II, and from 0 to 6.2% for the PF-10. Use of the standard scoring methods [12, 15, 16], resulted in no missing values for total scores.

Internal consistency of the scales was assessed by Cronbach's  $\alpha$  coefficients, where values  $\geq 0.70$  were considered adequate for group comparisons and values

between 0.90 and 0.95 for individual comparisons [17]. For concurrent validity, it was hypothesized that the scales should be at least strongly intercorrelated. As a first test of construct validity, the convergence and divergence of the scales with other aspects of health as measured by the SF-36 were examined. It was hypothesized that an adequate measure of disability should be strongly related to other aspects of physical and general health, moderately to aspects that may be a combination of physical and psychosocial factors and weakly to psychosocial aspects of health. For both analyses, a Spearman correlation coefficient of 0.20–0.39 was regarded as weak, 0.40–0.59 as moderate, 0.60–0.79 as strong and 0.80–1.0 as very strong [18]. Secondly, the relative efficiency of the scales in discriminating between groups based on self-reported general health (Item 1 from the SF-36) was examined using one-way analysis of variance (ANOVA) tests [19]. Given that only two and five patients reported excellent or poor health, respectively, responses of excellent and very good and fair and poor were collapsed into single categories. It was hypothesized that poorer levels of health should be associated with increasingly worse disability scores. For the purposes of comparison, relative validity (RV) coefficients for the HAQ-II and PF-10 as compared with the HAQ-DI were computed. Finally, distributional properties of the scale scores were examined to identify possible floor and ceiling effects. Floor or ceiling effects were considered to be present if  $>15\%$  of the patients scored the worst or best possible physical function score, respectively.

## Results

Mean (s.d.) age and disease duration of the patients (females/males: 17/80) were 58.9 (12.5) and 7.2 (8.9) years, respectively. The mean most recent serum urate value was 0.45 (0.12) mmol/l and 33% of the patients had tophi at the time of the study. Average NRS pain and general health scores were 4.46 (3.19) and 4.52 (2.74), respectively. Mean scores on the HAQ-DI, HAQ-II and PF-10 were 0.61 (0.73), 0.73 (0.74) and 61.94 (29.33), respectively.

Internal consistency was high for all measures, with Cronbach's  $\alpha = 0.93$  for the eight categories of the HAQ-DI ( $\alpha = 0.97$  for the 20 items) and  $\alpha = 0.94$  for the HAQ-II and PF-10. The measures were strongly to very strongly intercorrelated (Table 1). Additionally, they demonstrated a pattern of correlations similar to other aspects of health (Table 1). Most correlations were of the expected magnitude, although the association with role-emotional problems was somewhat stronger than expected for all measures. Likewise, all measures were able to significantly discriminate between different levels of general health (Table 2). As expected, disability scores on the HAQ versions gradually increased in patient groups with worse levels of health. The PF-10 performed somewhat worse, as represented by a slightly higher mean score in the good health category than in the excellent–very good health category and a lower RV value.

**TABLE 1** Spearman correlations between the HAQ-DI, HAQ-II and PF-10 and other dimensions of health status

Measures	Expected correlation	Observed correlation*		
		HAQ-DI	HAQ-II	PF-10
Concurrent validity				
HAQ-DI	-	-	-	-
HAQ-II	≥0.60	0.87	-	-
PF-10	≥0.60	-0.75	-0.79	-
Construct validity				
SF-36 role-physical	0.60-0.79	-0.67	0.72	0.68
SF-36 general health	0.60-0.79	-0.63	-0.65	0.56
SF-36 bodily pain	0.40-0.59	-0.42	-0.45	0.46
SF-36 vitality	0.40-0.59	-0.55	-0.54	0.54
SF-36 social functioning	0.40-0.59	-0.54	-0.56	0.43
SF-36 role-emotional	0.20-0.39	-0.42	-0.45	0.48
SF-36 mental health	0.20-0.39	-0.41	-0.35	0.30

\*All correlations significant at  $P < 0.01$ .

**TABLE 2** Ability of HAQ-DI, HAQ-II and PF-10 scores to discriminate between different levels of self-reported general health

Measures	Excellent/very good ( $n = 18$ )	Good ( $n = 46$ )	Fair/poor ( $n = 33$ )	$F^*$ (2, 94)	RV
HAQ-DI	0.16 (0.32) <sub>a</sub>	0.33 (0.51) <sub>a</sub>	1.25 (0.73) <sub>b</sub>	32.21	1.00
HAQ-II	0.28 (0.46) <sub>a</sub>	0.44 (0.55) <sub>a</sub>	1.39 (0.66) <sub>b</sub>	32.87	1.02
PF-10	71.91 (32.30) <sub>a</sub>	74.27 (22.15) <sub>a</sub>	39.33 (23.12) <sub>b</sub>	21.18	0.66

Values are represented as mean (s.d.); \*All  $F$ -values are significant at  $P < 0.001$ . Means in the same row that do not share the same subscript differ at  $P < 0.05$  using Bonferroni *post hoc* tests.  $F = F$ -statistic from one-way ANOVA; RV: RV (ratio of  $F$ -statistics compared with HAQ-DI).

High ceiling effects were observed for the HAQ-DI and HAQ-II with, respectively, 34.0 and 25.8% of the patients scoring no disability, compared with only 7.2% of the patients scoring perfect functioning on the PF-10. Floor effects were negligible for all scales (HAQ-DI: 1.0%; HAQ-II: 1.0%; PF-10: 3.1%).

## Discussion

Few patient-reported outcome measures have been thoroughly validated for use in patients with gout. Recently, the Gout Assessment Questionnaire (GAQ) was developed to measure gout-related quality of life [20]. Although the GAQ demonstrated promising psychometric properties, it does not contain a separate subscale for functional disability. Moreover, in order to compare the impact of gout with that of other conditions, more generic measures are still needed. In this study, we compared the reliability and validity of the generic HAQ-DI, HAQ-II and PF-10.

The scales were highly intercorrelated, indicating that they measure a similar construct. Moreover, they demonstrated sufficient reliability for use in individual-level analyses. Finally, all measures largely demonstrated the expected pattern of correlations with other aspects of health and the ability to discriminate between different

levels of health. Consequently, the results provide preliminary support for their reliable and valid use in measuring physical disability in gout. Psychometric properties of the HAQ-DI and HAQ-II were extremely similar. Given that the HAQ-II is shorter and easier to score, it may be an attractive alternative to the HAQ-DI.

However, the ceiling effects of both HAQ versions suggests that they may be measuring too many activities of daily living that are not relevant or too easy to perform for patients with gout. The proportion of patients scoring perfect functioning was well above the commonly accepted criterion of 15% for both scales [9]. This also makes these measures theoretically less suitable for intervention studies, because they may not be able to detect improvement in a large proportion of patients. The observed ceiling effect of the HAQ-DI is consistent with previous findings in gout [5-8]. More surprisingly, the HAQ-II, which was specifically developed to overcome this problem by including more difficult activities, also demonstrated a high ceiling effect.

The PF-10 did not have a ceiling effect, but demonstrated lower validity in discriminating between levels of health, and mean scores did not decrease as expected between the excellent-very good and good categories of health. Although this could be caused by the non-proportional number of patients in each category, both

HAQ scores did show steadily increasing disability scores. Consequently, the PF-10 may also not be optimally responsive in clinical intervention studies. Therefore, it is suggested that future studies examine the relevance and comprehensiveness of all three scales in patients with gout. Qualitative studies (e.g. focus groups, cognitive testing) or combinations of qualitative and quantitative approaches such as those performed in RA [21] may provide more information about the construct of physical disability in gout and the content validity of the current measures.

It should be noted that a cross-sectionally observed ceiling effect alone does not necessarily point to reduced responsiveness, especially when a considerable number of patients truly have no disability [22]. It becomes problematic, however, when scores remain stable when functional status changes by other measures. Moreover, we did not specifically examine whether patients were experiencing a flare-up of the disease at the time of the study. It is likely that ceiling effects will be less pronounced during activity flares, when the disease has its greatest impact on functioning. Consequently, the responsiveness of the measures during flares should be further examined in longitudinal trials. Finally, the present study design did not allow us to examine other important psychometric properties such as test-retest reliability and minimal important differences. Future studies should also examine these properties before the scales can be fully endorsed as adequate outcome measures of physical disability in gout.

In conclusion, the present study suggests that the HAQ-DI, HAQ-II and PF-10 are similarly reliable and valid measures of physical disability in patients with gout. However, more research is needed to examine whether their content validity can be improved and to evaluate their test-retest reliability and responsiveness to change.

#### Rheumatology key messages

- The HAQ-DI, HAQ-II and PF-10 are reliable and valid for descriptive studies in gout.
- The content validity and responsiveness of the HAQ-DI, HAQ-II and PF-10 need further study.

## Acknowledgements

We would like to thank the patients who participated in this study.

**Funding:** This work was supported by the Dutch Arthritis Association.

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

## References

- 1 Taylor WJ, Schumacher HR Jr, Singh JA, Grainger R, Dalbeth N. Assessment of outcome in clinical trials of gout—a review of current measures. *Rheumatology* 2007; 46:1751–6.
- 2 Taylor WJ, Schumacher HR Jr, Baraf HS *et al.* A modified Delphi exercise to determine the extent of consensus with OMERACT outcome domains for studies of acute and chronic gout. *Ann Rheum Dis* 2008;67: 888–91.
- 3 Schumacher HR, Taylor W, Joseph-Ridge N *et al.* Outcome evaluations in gout. *J Rheumatol* 2007;34: 1381–5.
- 4 Becker MA, Schumacher HR, Benjamin KL *et al.* Quality of life and disability in patients with treatment-failure gout. *J Rheumatol* 2009;36:1041–8.
- 5 Alvarez-Hernandez E, Pelaez-Ballesteros I, Vazquez-Mellado J *et al.* Validation of the health assessment questionnaire disability index in patients with gout. *Arthritis Rheum* 2008;59:665–9.
- 6 Taylor WJ, Colvine K, Gregory K, Collis J, McQueen FM, Dalbeth N. The health assessment questionnaire disability index is a valid measure of physical function in gout. *Clin Exp Rheumatol* 2008;26:620–6.
- 7 Khanna D, Ahmed M, Yontz D *et al.* The disutility of chronic gout. *Qual Life Res* 2008;17:815–22.
- 8 Kawata AK, Benjamin KL, Waltrip RW *et al.* Validation of the SF-36 and HAQ-DI in patients with treatment-failure gout. *Ann Rheum Dis* 2007;66:236.
- 9 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.
- 10 van Groen MM, ten Klooster PM, Taal E, van de Laar MA, Glas CA. Application of the health assessment questionnaire disability index to various rheumatic diseases. *Qual Life Res* 2010;19:1255–63.
- 11 Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23: 137–45.
- 12 Wolfe F, Michaud K, Pincus T. Development and validation of the health assessment questionnaire II: a revised version of the health assessment questionnaire. *Arthritis Rheum* 2004;50:3296–305.
- 13 Ware JE Jr. SF-36 health survey update. *Spine* 2000;25: 3130–9.
- 14 Haley SM, McHorney CA, Ware JE Jr. Evaluation of the MOS SF-36 physical functioning scale (PF-10): I. Unidimensionality and reproducibility of the Rasch item scale. *J Clin Epidemiol* 1994;47:671–84.
- 15 Ware JE, Kosinski M, Dewey JE. How to score version 2 of the SF-36 health survey. Lincoln, RI: QualityMetric Incorporated, 2000.
- 16 Fries JF. The health assessment questionnaire (HAQ) and the improved HAQ. Stanford: Stanford University School of Medicine, 2009.
- 17 Lohr KN. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res* 2002;11:193–205.
- 18 Campbell MJ, Swinscow TDV. *Statistics at square one*, 11th edition. London: Blackwell Publishing, 2009.
- 19 Hays RD, Revicki D. Reliability and validity (including responsiveness). In: Fayers PM, Hays RD, eds. *Assessing quality of life in clinical trials: methods*

- and practice. Oxford: Oxford University Press, 2005:25-39.
- 20 Hirsch JD, Lee SJ, Terkeltaub R *et al.* Evaluation of an instrument assessing influence of gout on health-related quality of life. *J Rheumatol* 2008;35: 2406-14.
- 21 Hewlett S, Smith AP, Kirwan JR. Values for function in rheumatoid arthritis: patients, professionals, and public. *Ann Rheum Dis* 2001;60:928-33.
- 22 Bruce B, Fries JF. The Stanford health assessment questionnaire: a review of its history, issues, progress, and documentation. *J Rheumatol* 2003;30:167-78.