# The Health Assessment Questionnaire Disability Index is a valid measure of physical function in gout

W.J. Taylor[1], K. Colvine[2], K. Gregory[3], J. Collis[2], F.M. McQueen[3], N. Dalbeth[3]

*[1]Department of Medicine, University of Otago, Wellington; [2]Counties Manukau District Health Board, Auckland; [3]Department of Medicine, University of Auckland, Auckland, New Zealand.*

## Abstract

### Objective

*There are no disability instruments that have specifically been validated for gout. The aim of this study was to determine the construct validity of the Health Assessment Questionnaire Disability Index (HAQ-DI) in gout and the internal validity using Rasch analysis.*

### Methods

*An observational cohort study of two groups of clinic patients with gout (n=20, n=53), in which clinical and functional measures were correlated with HAQ-DI scores. Rasch analysis was used to determine the internal validity of summated scores as a measure of physical disability.*

### Results

*The HAQ-DI items fitted a Rasch measurement model, confirming internal validity of the scale, although there was evidence of disordered thresholds and rescoring items as a 3-option response rather than a 4-option response improved model fit and resolved the disordered thresholds. HAQ-DI scores showed a bimodal distribution and evidence of floor effects. Clinical indices correlated highly with HAQ-DI scores in gout patients, particularly other measures of physical function. A strong relationship between days of sick leave and HAQ-DI was observed in gout patients ($r^2$=0.44, $p<0.001$).*

### Conclusions

*It is concluded that HAQ-DI has good construct and internal validity in gout but a modified scoring approach better fits a Rasch model.*

**Key words**

Gout, HAQ, psychometrics, validity.

*William J. Taylor, PhD;*
*Keith Colvine, MBChB;*
*Kate Gregory; Julie Collis, DipOT;*
*Fiona M. McQueen, MD;*
*Nicola Dalbeth, MD.*

*Nicola Dalbeth is funded by the Health Research Council of New Zealand. Kate Gregory was the recipient of a Maurice and Phyllis Paykel Summer studentship. This project was funded by the Auckland Regional Rheumatology Trust.*

*Please address correspondence and reprints requests to: Dr. W.J. Taylor, Department of Medicine, University of Otago, P.O. Box 7343, Wellington, New Zealand.*
*E-mail: will.taylor@otago.ac.nz*

*Received on September 10, 2007; accepted in revised form on December 14, 2008.*



*Competing interests: none declared.*

## Introduction

Gout is a prevalent disorder of uric acid metabolism (1), characterised by recurrent flares of self-limiting arthritis, and in the presence of persistent hyperuricaemia, formation of tophaceous disease and chronic deforming arthritis.

A number of new therapies have been recently developed for use in gout (2-5). Testing of new agents for management of gout requires careful assessment of clinical endpoints. In addition to relevant endpoints such as serum uric acid levels, frequency of gout flares and index tophus size, well-validated endpoints are needed to analyse the impact of these therapies on functional disability (6).

Functional status has not been often assessed in studies of gout. The original development of the Steinbrocker scale occurred in a population of patients with rheumatoid arthritis (RA), osteoarthritis and gout (7), but few studies since that time have reported functional measures in patients with gout. A recent observational study from Mexico measured physical function using the Health Assessment Questionnaire Disability Index (HAQ-DI) in a study of chronic gout to identify predictors of disability. This showed that nearly half of the sample had some level of disability and that HAQ-DI was associated with tophi, hypertriglyeridemia and ischaemic heart disease (8). A study of patients with severe gout screened for inclusion in a Phase 1 study of pegylated uricase found high scores on the modified Health Assessment Questionnaire (MHAQ), with a mean (95% confidence interval) score of 1.19 (0.94 to 1.44) (9). However, these instruments have not been specifically validated in patients with gout.

The HAQ-DI is the most widely used tool for assessment of functional status in rheumatology practice and research. Since the development of the HAQ-DI in 1980, this questionnaire has been extensively validated for use in rheumatoid arthritis (RA) (10, 11). The HAQ-DI is a key patient reported outcome in the majority of rheumatology clinical trials, and is frequently used as part of the core set of the ACR clinical response criteria in rheumatoid arthritis (RA) clinical trials (12). In patients with RA, the HAQ-DI is a strong predictor of health care utilisation (13), work disability (14), morbidity and mortality (15).

The HAQ-DI was comprehensively reviewed in 2003 and cites 183 publications relevant to the HAQ since 1996 (11). It has been employed in studies of patients with HIV-AIDS, normal aging populations, disabled workers, National Health and Nutrition Examination Survey (NHANES), osteoarthritis, systemic lupus erythematosus, fibromyalgia, psoriatic arthritis and systemic sclerosis. It has been translated into more than 60 languages. Normative values are available from a Finnish population (16). It is easily available and easy to use. The HAQ-DI consists of 20 activities (items) rated as to the degree of difficulty with 4 response options ranging from "no difficulty" to "unable to do". The items are aggregated into 8 categories and the highest scoring item contributes to the score of that category. The HAQ-DI raw score is expressed as the average of the 8 category scores (scale range 0 to 3). In addition the use of any aids for any activity increase the relevant category score to at least "with much difficulty".

Nonetheless, there have been criticisms of the HAQ-DI. The HAQ-DI was not able to detect short-term changes in physical performance in a trial of exercise in patients with RA (17). The HAQ-DI raw scores are ordinal rather than interval and therefore not strictly suitable for parametric statistical analysis (18). Scoring all 20 items of the HAQ-DI, rather than the highest scoring item within each category improves person separation, that is, it distinguishes more clearly between people of different disability levels (19). Rasch analysis identified an 8-item questionnaire (derived from the items of HAQ-DI) that had better measurement properties that the HAQ-DI for people with fibromyalgia (20).

The aim of this study was to determine whether the HAQ-DI is a valid tool for assessment of disability in patients with gout. This work is necessary to determine the 'truth' aspect of the OMERACT filter (21): does the HAQ-DI really measure the construct of physical disability in gout?

Valid self-report measures need to conform to adequate psychometric standards in order for the results of the questionnaire to be trusted. When questionnaire data are expressed numerically, there is even greater need to ensure that the numbers are meaningful, especially when mathematically manipulated as in parametric statistics. This has clinical relevance both to the practitioner using the questionnaire in the clinic and to the practitioner when assessing the results of clinical research that have used such questionnaires.

## Methods

*Subjects*

Two groups of patients with gout completed the HAQ-DI and were analysed in this study. All participants met the Wallace preliminary classification criteria for gout (22). The first group comprised of 20 patients with gout recruited from rheumatology outpatient clinics. This group contributed to a study of hand function in gout, which included completion of the Medical Outcomes Study Short Form 36 (SF-36) questionnaire (23), assessment of the American College of Rheumatology (ACR) functional class, measurement of hand function using the Disability of Arm, Shoulder and Hand (DASH) questionnaire (24), and the Sollerman hand function test (25).

A further group of 53 consecutive patients, attending rheumatology outpatient clinics for management of recurrent acute gout or chronic tophaceous gout was also studied.

For all patients, tender joint count (/68), swollen joint count (/66), pain score (visual analogue scale 0-100mm), patient global assessment (Likert scale 0-5), physician global assessment (Likert scale 0-5), number of days off work due to arthritis in preceding six months were recorded.

This study was approved by the local ethics committee and all patients gave written, informed consent to participate.

*Assessment tools*

All patients completed the 20-item HAQ-DI in English as described by Fries (26). This self-report questionnaire examines eight key domains of activities of daily living (dressing and grooming, arising, eating, walking, hygiene, reach, grip, activities). Each domain is scored 0-3 ("no difficulty" to "unable to do"), with the sum of all domains divided by 8 to provide a HAQ-DI score ranging from 0 (no disability) to 3 (severe disability) on an ordinal scale.

The SF-36 is a validated self-report questionnaire that measures eight domains (physical functioning, physical roles, emotional roles, bodily pain, general health, vitality, social functioning and mental health). Each domain ranges from 0 (severe disability) to 100 (no disability) (23).

The DASH is a self-reported series of 30 questions relating to symptoms and general function of the upper extremities, with a score of 100 indicating maximum disability (24).

The Sollerman hand function test is administered by an occupational therapist. In this test, patients are examined on their ability to perform 20 everyday tasks, testing seven of the eight most common hand grips. Activities include doing up buttons, turning a key and closing a zipper, and are scored on time taken and grip style used with a total of 80 indicating full hand function (25). This instrument is a useful measure of hand function in RA (27).

*Statistical analysis*

The performance of each of the eight items of the HAQ-DI was determined using Rasch analysis (28). This approach proposes a measurement model in which a response to an item depends on two parameters: the ability of the respondent (level of functional disability) and the difficulty of the item (level of functional difficulty). Each parameter can be estimated independently, importantly implying that item parameters are not sample-dependent. A probabilistic relationship (that takes the form of the logistic function) defines the Rasch model in terms of item difficulty and respondent ability. Measures that fit the Rasch model have many useful properties. Ideally, items within a measurement instrument should be as easy or as difficult for people with different characteristics (other than ability on the scale of interest). For example, for the people with the same level of physical disability, males should find the same item as difficult as females. The extent to which this is not true is termed "differential item functioning".

In simple terms, the Rasch model permits rational summation of item scores from a multiple item questionnaire. This is what we expect from all self-report questionnaires that have more than one question. For summated items to result in a single score that is actually meaningful, a number of conditions must be met. One of these conditions is that each item is measuring roughly the same attribute (uni-dimensionality); another is that the response to the item by a particular person should relate to the difficulty of the item (easier items should be endorsed with less of the attribute than more difficult items and more disabled people should endorse more functionally difficult items). Using the parameters of the Rasch model that are estimated from the observed data, Rasch modelled scores can be calculated. These scores are felt to be truer measures of the underlying attribute than the raw summated score and are strictly linear across the scale. The Rasch modelled scores are expressed in 'logits', which are units of measurement relevant to the logistic form of the Rasch model so that each logit unit change represents a factor of 2.718 (Napier's constant) in the amount of the attribute that the person possesses. In addition, the difficulty of each item is measured on the same logit scale, so a direct comparison between the targeting of the items and the ability of the sample can be easily undertaken. Because of the form of the Rasch model, when the item-difficulty is the same as the person-ability the probability of that item being endorsed is 50%. Rasch models originally developed from an educational context, in which student examinations were being assessed as valid representations of the knowledge that was being tested. In a rheumatology context, Wolfe *et al*. developed the HAQ-II from Rasch analysis of original HAQ-DI items plus additional items. The HAQ-II questionnaire was

shown to have better psychometric properties because it better fitted the Rasch model (29).

A rating scale polychotomous Rasch model was used for this analysis, expressing the probability of person answering in a certain way, given the ability of the person, the average difficulty of the item and the threshold of difficulty between each response option. This model was used after checking a likelihood-ratio test comparing the rating scale model with an unrestricted partial-credit model ($\chi^2_{df=13}$ 15.19, $p$=0.296).

Fit of the observed data to the Rasch model was assessed using a chi-square statistic of the residuals between modelled and observed scores. The calculated residual corresponds to an unweighted mean-square fit statistic. A sample size of 64 to 144 patients has been determined to estimate item location, with Type 1 error rate of 5%, to within 0.5 logits (30).

Differential item functioning (DIF) was assessed using ANOVA of group-specific residuals. In this analysis, 2-way ANOVA is conducted across levels of patient factors and levels of the variable (3 classes of HAQ-DI scores). Uniform DIF is indicated by a significant main effect for the patient factors and non-uniform DIF is indicated by an interaction effect between patient factor and class interval. The patient factors tested were: ethnicity (Caucasian, Non-Caucasian). RUMM2020 was used to perform the analysis using the rating scale polychotomous Rasch model (31). Bonferroni corrections were made to adjust for multiple statistical tests.

Internal consistency of the HAQ was measured by Cronbach's alpha reliability coefficient. Concurrent validity was measured by determining the Pearson's correlation between the Rasch modelled HAQ scores and other measures of function. A priori predictions were made regarding the approximate size of the correlation (low <0.30, medium 0.30 to 0.60, high >0.60). SPSS 12.0.1 was used to perform this analysis.

## Results

*Sample characteristics*

The characteristics of patients are shown in Table I. Since the sample was predominantly male, we did not test for differential item functioning (DIF) by gender.

**Table I.** Sample characteristics.

| | Gout hand function study n=20 | Gout clinic patients n=53 |
|---|---|---|
| Age, years (median, range) | 58 (38-78) | 57 (29-83) |
| Male gender, n (% | 19 (95%) | 41 (77%) |
| Ethnicity, n (%) | | |
| Maori/Pacific | 11 (55%) | 36 (68%) |
| Other | 9 (45%) | 17 (32%) |
| Disease duration, years (median, range) | 17 (1-50) | 11 (0.5-35) |
| Tophaceous disease, n (%) | 16 (80%) | 38 (72%) |

*Rasch analysis*

The HAQ-DI fitted the Rasch model with no evidence of item-trait interaction (item residual-0.35 SD 0.76; $\chi^2_{df=16}$ 20.6, $p$=0.19). However, all items showed evidence of disordered thresholds both in the gout. The pattern of thresholds suggest that patients are confused or inconsistent in how they score an item as "with much difficulty" or "unable to do". A repeat analysis after rescoring items by combining the "with much difficulty" and "unable to do" responses into a single category showed better overall fit (mean fit residual-0.10 SD 1.04) and there was no evidence of disordered thresholds. A principal components factor analysis of person-item residuals confirmed uni-dimensionality of the modified HAQ-DI scores with 4.00% (95%CI 0 to 9) of $t$-tests that compared positively loading with negatively loading items on the first component of the factor analysis, being significant at the 5% level. The fit of the modified HAQ-DI to the Rasch model is shown in Table II.

Items from an ideal measure should perform in the same way in different groups of subjects. That is, the difficulty estimates of each item should be the same in non-Caucasian as Caucasian. The extent to which this does not occur is termed DIF. There was no evidence of DIF by ethnicity in this dataset, using the modified scoring.

*Internal consistency*

Internal consistency was assessed by Cronbach's alpha reliability coefficient, which represents the average inter-item correlation. For the HAQ-DI scores in gout (n=73), Cronbach alpha measured 0.94, indicating excellent internal consistency.

*Concurrent validity*

The HAQ-DI is used to measure physical function. Therefore, we hypothesized that the HAQ-DI should strongly

**Table II.** Fit of HAQ-DI (modified scoring) to the Rasch model in gout patients.

| | Item difficulty (SE)* | Residual ($p$-value)† |
|---|---|---|
| Eating | 1.16 (0.26) | 0.63 (0.72) |
| Dressing | 0.47 (0.25) | -0.88 (0.49) |
| Walking | 0.11 (0.24) | 0.83 (0.69) |
| Reaching | -0.004 (0.24) | -1.52 (0.26) |
| Grip | -0.27 (0.24) | 1.02 (0.99) |
| Arising | -0.36 (0.24) | 0.90 (0.20) |
| Hygiene | -0.48 (0.24) | -0.95 (0.23) |
| Activity | -0.63 (0.24) | -0.87 (0.34) |

*Item-difficulty refers to the amount of functional disability required for a person to be likely to endorse that particular item, as estimated by the Rasch model. Values are given in logits (with standard errors). The lower item-difficulty estimates represent functional harder items.

†$p$-value for a Chi-square test, Bonferroni corrected significance level 0.00625. The residual represents the difference between the observed score and the Rasch modelled score.

**Table III.** Correlation between HAQ-DI and other clinical measures in gout.

| Indicator | Expected correlation | Observed correlation |
|---|---|---|
| ACR functional class [a] | +++ | 0.94* |
| SF - physical function [a] | +++ | -0.83* |
| SF - physical role limitations [a] | +++ | -0.60* |
| SF - emotional health [a] | + | -0.54 |
| SF - emotional role limitations [a] | + | -0.59 |
| SF - vitality [a] | ++ | -0.44* |
| SF - pain [a] | ++ | -0.60* |
| SF - general health [a] | ++ | -0.73* |
| SF - social [a] | + | -0.48 |
| DASH score [a] | +++ | 0.81* |
| Sollerman score [a] | +++ | -0.79* |
| Number of gout flares in the past 6 months [a] | + | 0.41 |
| Pain VAS [b] | ++ | 0.70 |
| Patient global assessment [b] | +++ | 0.73* |
| Physician global assessment [b] | ++ | 0.77 |
| Number of days of sick leave in past six months [b] | +++ | 0.68 (n=41)* |
| Swollen joint count [b] | ++ | 0.62 |
| Tender joint count [b] | ++ | 0.58* |

[a]Hand function sample, n=20; [b]Total sample, n=73
+low correlation (0 to 0.29); ++ moderate correlation (0.3 to 0.59); +++ strong correlation (more than 0.60).
*indicates where predicted correlation was actually observed

correlate with other measures of functional status. Rasch transformed scores were used in this analysis. That is, the estimate of physical function derived from the HAQ-DI is indicated by its logit score from the Rasch analysis. This gives an interval-level measure that is the best estimate of physical functioning and is suitable for parametric analysis. A logit represents the unit of measurement relevant to the logistic form of the Rasch model, so that one logit change represents a factor 2.718 (Napier's constant). The relationship between HAQ-DI and other indicators of functional status as shown in Table III together with the predicted correlation based on the behaviour of these measures in other populations. In general, the correlations between clinical indicators and HAQ-DI in gout were stronger than predicted. In the 20 patients participating in the hand function study, there was a very clear relationship between HAQ-DI score and ACR functional class, with a correlation of 0.94 ($p<0.001$). Similarly, there was a strong correlation between other measures of physical disability in this group, such as the SF-physical function, SF-physical role limitations, DASH and Sollerman test. Of 18 predicted correlations, 10 were observed in gout. There was a strong relationship between the HAQ-DI and the number of sick days ($r^2=0.45$, $p<0.001$).

*Floor and ceiling effects*

Ideally, a measure should cover the range of potential ability. The floor and ceiling effects of the HAQ were analysed by plotting the distribution of HAQ scores for the entire gout cohort. This analysis demonstrated that the distribution of HAQ scores in the gout cohort was skewed with 20.5% of scores of 0 (Fig. 1). Rasch transformation normalised the distribution somewhat but there was still clear evidence of floor effects. Although the physical function subscale of SF-36 was more normally distributed, there was still a floor effect with 25% of the gout group scoring 90 or more.

## Discussion

The HAQ-DI performs well as a measure of physical disability in assessment of patients with gout. There is excellent correlation with other measures of physical disability, *e.g.*, DASH and Sollerman as has been described in RA (32). Good correlation with work disability was also observed, a potential

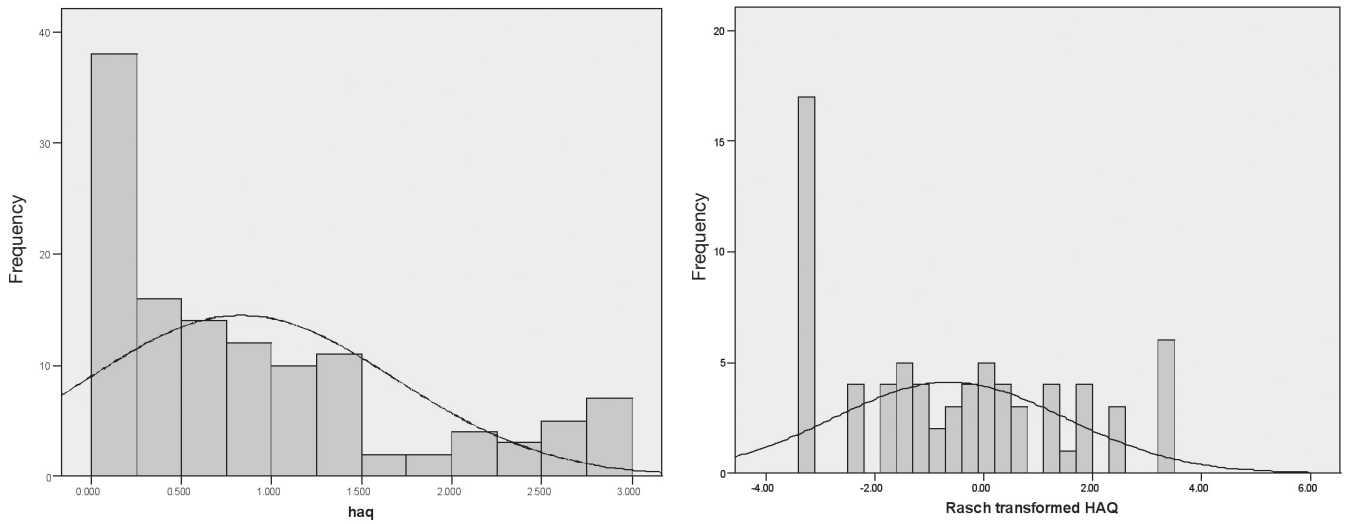


**Fig. 1.** Distribution of HAQ-DI scores (original scoring and Rasch modelled scores). Showing floor and ceiling effects.

consequence of poor functional ability. This is all consistent with adequate construct validity of the HAQ-DI in gout. We did find improved internal validity of the HAQ-DI by collapsing response categories from 4 options to 3 options. It would be of interest to assess whether similar findings are observed in other studies of the HAQ-DI. Floor effects mean that small amounts of disability may not be adequately measured. However, this is similar in RA (33, 34).

Gout disease activity fluctuates more than in RA, and significant changes in functional capacity may occur at the time of acute flares. Of note, many of the patients included in this study had chronic tophaceous gout with long disease duration, and this group may respond differently to those with early disease and recurrent acute gout. However, when duration of disease was entered into a regression model, there was no significant effect of this variable on the relationship between HAQ-DI and the number of sick-leave days.

Other aspects of the OMERACT filter such as responsiveness over time and to therapy have not been analysed in this cross-sectional study. Further analysis in clinical trials of new agents will help to determine the responsiveness and discrimination between active and placebo treatment arms of the HAQ-DI. It may be prudent to evaluate newer, more robust versions of the HAQ-DI such as the 10-item HAQ-II (34) in further studies of functional status in patients with gout.

## Conclusions

The HAQ-DI is a valid instrument for measuring physical disability in people with gout. Internal validity is improved by a collapse of response options to each item from 4 to 3.

## References


1. KLEMP P, STANSFIELD SA, CASTLE B, ROBERTSON MC: Gout is on the increase in New Zealand. *Ann Rheum Dis* 1997; 56: 22-6.
2. SCHUMACHER HR JR, BOICE JA, DAIKH DI *et al*.: Randomised double blind trial of etoricoxib and indometacin in treatment of acute gouty arthritis. *BMJ* 2002; 324: 1488-92.
3. BECKER MA, SCHUMACHER HR JR, WORTMANN RL *et al*.: Febuxostat, a novel non-purine selective inhibitor of xanthine oxidase: A twenty-eight-day, multicenter, phase II, randomized, double-blind, placebo-controlled, dose-response clinical trial examining safety and efficacy in patients with gout. *Arthritis Rheum* 2005; 52: 916-23.
4. BECKER MA, SCHUMACHER HR, JR, WORTMANN RL *et al*.: Febuxostat compared with allopurinol in patients with hyperuricemia and gout. *N Eng J Med* 2005; 353: 2450-61.
5. SUNDY JS, GANSON N, KELLY SJ, SCARLETT EL, HERSHFIELD MS: A Phase I study of pegylated-uricase (Puricase) in subjects with gout. *Arthritis Rheum* 2004; 50: S337-38.
6. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Center for Devices and Radiological Health (CDRH). Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. In: *Food and Drug Administration*; 2006.
7. STEINBROCKER O: Prognosis for employability in the major arthritides rheumatoid arthritis, osteoarthritis and gout. *Pa Med* 1969; 72: 82-5.
8. ALVAREZ-NEMEGYEI J, CEN-PISTE JC, MEDINA-ESCOBEDO M, VILLANUEVA-JORGE S: Factors associated with musculoskeletal disability and chronic renal failure in clinically diagnosed primary gout. *J Rheumatol* 2005; 32: 1923-27.
9. GELETKA RC, HERSHFIELD MS, SCARLETT EL, SUNDY JS: Severe gout is associated with impaired quality of life and functional status. *Arthritis Rheum* 2004; 50: S340-41.
10. FRIES JF, SPITZ PW, YOUNG DY: The dimensions of health outcomes: the health assessment questionnaire, disability and pain scales. *J Rheumatol* 1982; 9: 789-93.
11. BRUCE B, FRIES JF: The Stanford Health Assessment Questionnaire: A review of its history, issues, progress, and documentation. *J Rheumatol* 2003; 30: 167-78.
12. FELSON DT, ANDERSON JJ, BOERS M *et al*.: The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. *Arthritis Rheum* 1993; 36: 729-40.
13. MICHAUD K, MESSER J, CHOI HK, WOLFE F: Direct medical costs and their predictors in patients with rheumatoid arthritis - A three-year study of 7, 527 patients. *Arthritis Rheum* 2003; 48: 2750-62.
14. WOLFE F, HAWLEY DJ: The longterm outcomes of rheumatoid arthritis: Work disability: a prospective 18 year study of 823 patients. *J Rheumatol* 1998; 25: 2108-17.
15. WOLFE F, MICHAUD K, GEFELLER O, CHOI HK: Predicting mortality in patients with rheumatoid arthritis. *Arthritis Rheum* 2003; 48: 1530-42.
16. KRISHNAN E, SOKKA T, HAKKINEN A, HUBERT H, HANNONEN P: Normative values for the health assessment questionnaire disability index - Benchmarking disability in the general population. *Arthritis Rheum* 2004; 50: 953-60.
17. VAN DEN ENDE CH, BREEDVELD FC, DIJKMANS BA, HAZES JM: The limited value of the Health Assessment Questionnaire as an outcome measure in short term exercise trials. *J Rheumatol* 1997; 24: 1972-7.
18. TENNANT A, HILLMAN M, FEAR J, PICKERING A, CHAMBERLAIN MA: Are we making the most of the Stanford Health Assessment Questionnaire? *Br J Rheumatol* 1996; 35: 574-8.
19. WOLFE F: Which HAQ is best? A comparison of the HAQ, MHAQ and RA-HAQ, a difficult 8 item HAQ (DHAQ) and a rescored 20 item HAQ (HAQ20): analyses in 2491 rheumatoid arthritis patients following leflunomide initiation. *J Rheumatol* 2001; 28: 982-9.
20. WOLFE F, HAWLEY DJ, GOLDENBERG DL, RUSSELL IJ, BUSKILA D, NEUMANN L: The assessment of functional impairment in fibromyalgia (FM): Rasch analyses of 5 functional scales and the development of the FM Health Assessment Questionnaire. *J Rheumatol* 2000; 27: 1989-99.
21. BOERS M, BROOKS P, STRAND V, TUGWELL P: The OMERACT filter for outcome measures in rheumatology. *J Rheumatol* 1998; 25: 198-9.
22. WALLACE SL, ROBINSON H, MASI AT, DECKER JL, MCCARTY DJ, YU TF: Preliminary criteria for the classification of the acute arthritis of primary gout. *Arthritis Rheum* 1997; 20: 895-900.
23. WARE JE JR, SHERBOURNE CD: The MOS 36-item Short-Form Health Survey (SF-36). *Med Care* 1992; 30: 473-83.
24. HUDAK PL, AMADIO PC, BOMBARDIER C *et al*.: Development of an upper extremity outcome measure: The DASH (Disabilities of the Arm, Shoulder, and Hand). *Am J Ind Med* 1996; 30: 372-372.
25. SOLLERMAN C, EJESKAR A: Sollerman hand function test. A standardised method and its use in tetraplegic patients. *Scand J Plast Reconstr Surg Hand Surg* 1995; 29: 167-76.
26. FRIES JF, SPITZ P, KRAINES RG, HOLMAN HR: Measurement of patient outcomes in arthritis. *Arthritis Rheum* 1980; 23: 137-45.
27. ZHENG S, ROBINSON E, YEOMAN S *et al*.: MRI bone oedema predicts eight year tendon function at the wrist but not the requirement for orthopaedic surgery in rheumatoid arthritis. *Ann Rheum Dis* 2006; 65.
28. BOND TG, FOX CM: Applying the Rasch model: fundamental measurement in the Human Sciences. Mahwah, NJ: Lawrence Erlbaum Associates; 2001.
29. 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.
30. LINACRE JM: Sample size and item calibration stability. *Rasch Measure Tran* 1994; 7: 28.
31. ANDRICH D, LYNE A, SHERIDAN B, LUO G: Rasch Unidimensional Measurement Models (RUMM2020 Version 4.0). In *Duncraig, Western Australia*, Rumm Laboratory Pty Ltd; 2003.
32. CHIARI, GRISAR C, KOLLER U, STAMM TA, WANIVENHAUS A, TRIEB K: Performance

of the disabilities of the arm, shoulder and hand outcome questionnaire and the Moberg picking up test in patients with finger joint arthroplasty. *Arch Phys Med Rehabil* 2006; 87: 203-6.

33. BENTON N, STEWART N, CRABBE J, ROBINSON E, YEOMAN S, MCQUEEN FM: MRI of the wrist in early rheumatoid arthritis can be used to predict functional outcome at 6 years. *Ann Rheum Dis* 2004; 63: 555-61.

34. 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.