RHEUMATOLOGY

Original article

doi:10.1093/rheumatology/kez064

Development and validation of a patient-reported gout attack intensity score for use in gout clinical studies

Carly A. Janssen[1], Martijn A. H. Oude Voshaar[1], Peter M. ten Klooster[1], Harald E. Vonkeman[1,2] and Mart A. F. J. van de Laar[1,2]

Abstract

Objective. Inflammation-related symptoms such as pain, swelling and tenderness of the affected joint are frequently assessed using 5-point diary rating scales in gout clinical trials. Combining these into a single gout attack symptom intensity score may be a useful summary measure for these data, which is potentially more responsive to change compared with the individual components. The objective of this study was to develop a patient-reported gout flare intensity score, the Gout Attack Intensity Score (GAIS), for use in clinical studies, that includes components for gout-related pain, swelling and tenderness.

Methods. Data from a randomized controlled trial comparing anakinra to standard of care for the treatment of acute gout attacks were used for this study. A 7-day flare diary was completed by patients, including questions relating to intensity of pain, swelling and tenderness (5-point rating scales). Scalability of these items was assessed using Mokken Scale Analysis, and reliability using greatest lower bound reliability coefficients. Known-groups validity was evaluated, as well as the responsiveness to change and the presence of floor and ceiling effects.

Results. Scalability of the single items was supported, and GAIS scores were reliable (greatest lower bound >0.80). GAIS scores demonstrated responsiveness to change with high effect sizes (>0.8), and discriminated better between responders and non-responders compared with its single-item components. No floor and ceiling effects were found.

Conclusion. The GAIS seems to be a reliable and responsive instrument for assessing patient-reported gout attack intensity that may be used in gout clinical studies.

Key words: gout attack intensity, pain, swelling, tenderness, outcome measure

Rheumatology key messages

- The Gout Attack Intensity Score assesses gout flare symptom intensity in clinical studies.
- The Gout Attack Intensity Score includes components for patient-reported pain, swelling and tenderness.
- The Gout Attack Intensity Score is a reliable, responsive instrument for assessing gout flare intensity.

Introduction

In gout clinical studies, flare diaries are often administered to patients to assess the severity of the various symptoms they may experience as a result of their gout attack [1–4]. Such diaries typically include 5-point rating scales that ask patients to rate the intensity of their joint pain, joint swelling and joint tenderness [5]. Despite their popularity in study protocols and the endorsement of the individual symptoms as relevant outcome domains by the OMERACT Gout Working Group, reports of clinical trials in gout rarely present outcomes for symptoms other than pain in the main paper [6, 7]. Possibly, this is because the different individual symptom measures are considered to provide somewhat redundant information, or because limited evidence is currently available with respect to the validity of these scores [8].

Combining the individual scores of each symptom rating scale included in a gout attack diary (i.e. pain, swelling, tenderness) may summarize the information provided by the individual rating scales into a single score that represents a patient's overall level of gout attack intensity. In addition, combining information from multiple rating

[1]Arthritis Center Twente, Department of Psychology, Health and Technology, University of Twente and [2]Department of Rheumatology and Clinical Immunology, Medisch Spectrum Twente, Enschede, The Netherlands

Submitted 28 November 2018; accepted 4 February 2019

Correspondence to: Carly A. Janssen, Department of Psychology, Health and Technology, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands. E-mail: C.A.Janssen@utwente.nl

scales that assess a single latent variable also serves to reduce the contribution of measurement error to the variability of the scores. This should theoretically increase the ability of the combined score to measure change over time and to discriminate between responders and non-responders, relative to the individual rating scales. This notion is supported by a previous study which already showed that more reliable results can be obtained with composite scores including two to nine pain intensity items, compared with a single 24-h recall pain intensity item, using data from a published randomized controlled clinical trial [9].

In the current study we propose a new three-item Gout Attack Intensity Score, referred to as the GAIS, for use in flare diaries in clinical trials or observational studies in patients with acute gout attacks, and evaluate its measurement properties.

Methods

Study database

Data obtained during an investigator-initiated randomized, double blinded, double dummy, active placebo-controlled non-inferiority trial carried out in The Netherlands across seven rheumatology clinics were used for this study (NTR5234) [10]. In short, adult patients with crystal-proven acute gout attacks recruited between February 2016 and February 2018 were randomized to treatment with the IL-1 inhibitor anakinra, or to treatment with usual care (colchicine, NSAID or CSs) for the treatment of their acute gout attack. Results of this study showed that anakinra was non-inferior to standard of care in the treatment of acute gout flares.

Patients were asked to fill in a 7-day flare diary starting at baseline (day 1), including items about perceived pain [5-point rating scale and 10-point numeric rating scale (NRS)], tenderness (5-point rating scale), swelling (5-point rating scale), treatment response (8-point rating scale) and their global assessment of overall wellbeing [patient global assessment (PGA)] (10-point NRS). For the present study, PGA scores were recoded so that higher scores represent worse wellbeing. A copy of the flare-diary including all its items is included in supplementary material, section Gout Flare Diary, available at *Rheumatology* online. At day 1 and day 7, levels of CRP were measured. The study was performed in accordance with the Declaration of Helsinki and was approved by both an independent ethics committee and the institutional review board of each participating centre. All participants provided written informed consent.

For the current study, we used the daily data of the 7-day flare diary of patients in both treatment groups. The GAIS was obtained by taking the mean of the patient-reported 5-point rating scale pain, rating scale swelling and rating scale tenderness. Since the GAIS contains only three items, only patients who had no missing values for the three single items used to calculate the GAIS were included.

Scaling properties

Scaling properties were examined using the model of monotone homogeneity (MMH), using the Mokken package in R×64 version 3.4.2. The MMH is a non-parametric item response theory model. The model is based on the assumption that there exists a latent variable (θ) on which a scale's items as well as the persons responding to the items can be ordered. The model can be considered a probabilistic version of polytomous Guttman scaling. In polytomous Guttman scaling, each item with m response categories is broken down into m-1 item steps, which represent the location on the latent variable where the probability that a person selects the higher of two adjacent response categories jumps from 0 to 1. Violations of the Guttman model, or Guttman errors, occur if a patient endorses an item step that has a higher location on the latent variable compared with another item step that this patient fails to endorse. In the MMH, the probability that a person will select the higher of two adjacent response categories does not jump from 0 to 1 at a particular location on the latent variable, like in polytomous Guttman scaling, but is instead described for different values of the latent variable by the item step response functions (ISRF). ISRF are defined as $P(X_i \geqslant x_i|\theta$, where X_i is a random variable that refers to the score on item i, which takes on values: $x_i = 0,. m$.

The MMH applies if all i items in a scale measure the same latent variable, i.e. the scale is unidimensional and the ISRF are monotonically non-decreasing throughout the latent variable [11]. If the model applies, it supports that higher scores on the scale reflect a higher level of gout attack intensity.

Monotonicity was tested by inspecting plots of the ISRF of each item (e.g. swelling) over the summed score continuum of the two remaining items (e.g. pain and tenderness). Deviations from monotonicity were statistically tested, using group sizes of 5, 10 and 20, with the check.monotonicity function of the Mokken R package. Monotonicity was considered to apply if the plots of ISRF were non-decreasing, the number of statistically significant deviations from monotonicity were zero and the magnitude of the violations, as indicated by the crit statistic (critical value for model violations statistic) was $\leqslant$40 [12]. Unidimensionality was tested using Loevinger's scalability coefficients, which take on lower values as the number of Gutmann errors increase. Both item-level scalability (H_i), as well as scale-level scalability coefficients (H) were assessed. Scale-level scalability coefficients between 0.30 and $<$0.40 suggest weak scalability, between 0.40 and $<$0.50 suggest a moderate scale and $\geqslant$0.5 suggests a strong scale [13, 14]. A lower bound of 0.3 was considered the cut-off point for item-level scalability, according to a rule of thumb [15]. Data from each day of the gout flare diary were used for the analyses.

Reliability

Reliability of the GAIS observed on days 1–7 was assessed using greatest lower bound (GLB) coefficients,

since this statistic has been shown to yield a more realistic estimate of the lower bound value for test reliability compared with the more frequently applied Cronbach's alpha [16]. For tests with small numbers of items, such as in our study, GLB has been shown to be relatively unaffected by sampling bias, even with small sample sizes [17]. Reliability coefficients of 0.80 were considered to be satisfactory for use in clinical research settings [18]. The Psych and Rcsdp package of the statistical programme R x64 version 3.4.2 and Rstudio were used for the reliability analyses. The inter-item covariance matrix was obtained in SPSS Version 22.

Validity

Known-groups validity was examined by determining whether groups with different levels of self-reported treatment response could be distinguished based on their scores on the GAIS, joint swelling, joint tenderness, pain and PGA, at baseline. For this analysis, we partitioned the total sample in two groups based on their level of self-reported treatment response: patients who scored within the range of 1 (completely resolved) to 4 (somewhat improved) were considered responders, and patients who scored in the range 5 (unchanged) to 8 (very much worse) were considered non-responders. Score differences between responders and non-responders were compared using one-way analysis of variance in SPSS, version 22. It was expected that patients who reported having a poorer treatment response (non-responders) would have significantly higher scores on all the instruments compared with the responders. To compare the discriminative ability between different instruments, the relative efficiency (RE) of each instrument was compared with that of the GAIS [19]. It was hypothesized that the discriminative ability of the GAIS would be greater than that of the single items of which it consists, represented by a higher RE.

Responsiveness and sensitivity to change

Responsiveness to change was compared between different instruments by calculating Cohen's *d* effect sizes (ES) as [mean day 1 – mean *t* days]/pooled S.D. In the analyses, only cases for whom on both day 1 and day 5 (or day 1 and day 7 for CRP) data were available for each instrument were included. In calculating the ES for CRP, log-transformed data were used. An ES of 0.2 was considered a small effect, 0.5 a moderate effect and 0.8 a large effect [20]. Since anti-inflammatory treatment at recommended dosages was administered to all patients, and gout attacks are generally known to have a self-limiting course of limited duration, we expected to observe large improvements in clinical status over time, and thus large effect sizes for all instruments. Analyses were done using SPSS, version 22, and Microsoft Excel.

Floor and ceiling effects

Floor (ceiling) effects, defined as the proportions of patients scoring the worst (best) possible score were compared between the GAIS and the individual component scores, the NRS pain and the NRS PGA of the gout flare diary, at baseline. We hypothesized that ceiling effects (lowest possible score) would be absent for all instruments, considering that at baseline, gout attack symptoms would be expected to be at their worst. Floor or ceiling effects were considered to be problematic in case >15% of patients scored the worst or best possible score, respectively [21]. Analyses were done using SPSS, version 22.

Results

Baseline population

A total of 88 patients were enrolled in the study. The baseline population consisted predominantly of middle-aged men suffering from mono-articular gout. Other baseline characteristics of the entire sample are listed in Table 1. At baseline, 76 (86%) patients completed the gout flare diary. For the following days, diaries for which the summed score could be calculated were available for 82, 85, 83, 86, 79 and 82 patients on days 2–7, respectively.

Scalability

The item-level scalability coefficients (H_i) for pain, tenderness and swelling were all greater than the lower bound cut-off value of 0.3 for days 1 through 7, ranging from 0.61–0.83, 0.63–0.84 and 0.44–0.75, respectively (Table 2). Scale-level scalability coefficients for the GAIS revealed strong unidimensionality ($H \geqslant 0.50$), with H coefficients ranging from 0.56–0.80 across the seven days. Statistical testing of monotonicity showed no significant deviations from zero at group sizes 5, 10 and 20 for days 1 through 7, as well as no critical values >40. These results support that the ISRFs and the expected mean items score were monotonically increasing over the latent variable (Table 2 and supplementary Table S1, available at *Rheumatology* online). This was also confirmed by visual inspection of the ISRF plots. Overall, the results of the Mokken scaling analysis support the conclusion that all items relate to the same latent variable, and that higher scores on GAIS indicate a higher level of gout attack intensity.

Reliability

The median reliability coefficient for days 1–7 was 0.84, with GLB coefficients ranging between 0.78 and 0.91. Only one GLB was <0.80. These results suggest that scores on the GAIS are generally sufficiently reliable for use in clinical trials (Table 2).

Validity

All instruments, with the exception of the NRS PGA, were able to significantly ($P < 0.05$) discriminate between the two groups of self-reported treatment response, showing mean scores in the non-responders group that were higher than the responders for all instruments, as hypothesized (Table 3). As expected, the RE of the single items pain (0.41), swelling (0.81) and tenderness (0.66) were all lower than the GAIS (1.00), implying the

Table 1 Baseline characteristics

Characteristic	Score range of measure	Value	*n*
Age, mean (S.D.), years	–	61.6 (12.8)	88
Male sex, *n* (%)	–	83 (94.3)	88
BMI, mean (S.D), kg/m^2	–	29.1 (4.1)	86
Systolic blood pressure, mean (S.D.), mmHg	0–999	143.3 (22.9)	79
Diastolic blood pressure, mean (S.D.), mmHg	0–999	85.2 (14.5)	80
SUA, median (Q1, Q3), mmol/l	0–9	0.51 (0.44, 0.59)	82
CRP, median (Q1, Q3), mg/l	0–999	15.0 (6.0, 32.0)	82
GAIS, median (Q1, Q3)	1–5	3.67 (3, 4)	76
5-point rating scale pain, median (Q1, Q3)	1–5	4 (3, 4)	76
5-point rating scale tenderness, median (Q1, Q3)	1–5	4 (3, 4)	76
5-point rating scale swelling, median (Q1, Q3)	1–5	3 (3, 4)	76
NRS pain, mean (S.D.)	0–10	6.47 (1.70)	76
Gout classification[a], *n* (%)			
Monoarticular	–	53 (60.2)	88
Oligoarticular	–	28 (31.8)	88
Polyarticular	–	7 (8.0)	88
Number of gout attacks in previous $\leqslant$12 months, median (Q1, Q3)	0–999	3.0 (1.0, 4.0)	88
Intermittent gout, *n* (%)[b]	–	71 (80.7)	88
Comorbidities, *n* (%)			
Diabetes mellitus	–	8 (15.1)	53
Hypertension	–	27 (50.9)	53
Cardiovascular disease	–	29 (54.7)	53
Renal disorders	–	5 (9.4)	53
Musculoskeletal disease[c]	–	10 (18.9)	53
Gastrointestinal disorders	–	10 (18.9)	53
Neurological disorders	–	4 (7.5)	53

[a]Monoarticular implies one joint has been affected by gout; oligoarticular, more than one but fewer than five joints have been affected by gout; polyarticular, five or more joints have been affected by gout. [b]Compared with patients having chronic gout. [c]Diseases other than gout. SUA: serum urate acid; GAIS: Gout Attack Intensity Score; NRS: numeric rating scale; *n*: total number of patients; Q1, Q3: first and third quartile respectively.

Table 2 Scalability and reliability coefficients, of the single-component items and GAIS

		Unidimensionality, *H* (S.E.)				Monotonicity[a]			Reliability[b]
		Item level			Scale level GAIS				
Day	*n*	Pain	Tenderness	Swelling		Pain	Tenderness	Swelling	
1	76	0.61 (0.07)	0.63 (0.07)	0.44 (0.10)	0.56 (0.07)	0 (0)	0 (28)	0 (17)	0.78
2	82	0.67 (0.06)	0.73 (0.05)	0.62 (0.07)	0.68 (0.06)	0 (0)	0 (0)	0 (0)	0.83
3	85	0.66 (0.08)	0.72 (0.06)	0.59 (0.09)	0.66 (0.07)	0 (0)	0 (0)	0 (0)	0.80
4	83	0.71 (0.07)	0.69 (0.06)	0.64 (0.07)	0.68 (0.06)	0 (0)	0 (0)	0 (0)	0.84
5	86	0.83 (0.05)	0.81 (0.05)	0.75 (0.07)	0.80 (0.05)	0 (–2)	0 (0)	0 (0)	0.89
6	79	0.78 (0.05)	0.76 (0.05)	0.70 (0.07)	0.75 (0.05)	0 (0)	0 (0)	0 (0)	0.88
7	82	0.83 (0.05)	0.84 (0.04)	0.72 (0.07)	0.80 (0.05)	0 (0)	0 (0)	0 (0)	0.91

[a]Given as the number of statistically significant deviations from monotonicity (critical value for violations), at group size = 10. [b]Reliability coefficient, *r*, according to the greatest lower bound. *H*: scalability coefficient for item-level scalability (*Hi*), and scale-level scalability (*H*); *n*: sample size; GAIS: Gout Attack Intensity Score.

GAIS had a greater discriminative ability. Only the NRS pain had a higher RE than the GAIS, which makes this instrument a more suitable instrument for discriminating between the groups when solely measuring pain.

Responsiveness, and floor and ceiling effects

As expected, after 5 days improvements were seen in mean scores on all the variables, with all ES being >0.8. Notably,

Table 3 Ability of instruments to discriminate between difference groups of self-reported treatment response

Instrument	Groups of treatment response on day 1: Non-responders ($n = 44$), mean (S.D.)	Responders ($n = 27$), mean (S.D.)	F	RE
GAIS	3.69 (0.64)	3.10 (0.66)	13.92*	1.00
Joint pain[a]	3.66 (0.78)	3.22 (0.70)	5.72*	0.41
Joint tenderness[a]	3.80 (0.85)	3.15 (0.91)	9.20*	0.66
Joint swelling[a]	3.61 (0.87)	2.93 (0.78)	11.31*	0.81
NRS joint pain[b]	7.11 (1.38)	5.56 (1.55)	19.32*	1.39
NRS PGA[b]	5.68 (2.09)	4.85 (1.35)	3.39	0.24

[a]Measured on a 5-point rating scale. [b]A 10-point scale. *Significant at $P < 0.05$. GAIS: Gout Attack Intensity Score; NRS: numeric rating scale; PGA: patient global assessment; F: F-statistic from one-way analysis of variance; RE: relative efficiency (ratio of F-statistics compared with the GAIS).

Table 4 Responsiveness to change of self-reported measures between day 1 and day 5

	n	Mean change (S.D)[a]	ES
GAIS	74	1.60 (0.70)	2.32
Joint pain[b]	74	1.66 (0.75)	2.21
Joint swelling[b]	74	1.49 (0.84)	1.76
Joint tenderness[b]	74	1.66 (0.85)	1.95
NRS joint pain[c]	74	4.37 (1.82)	2.40
NRS PGA[c]	74	2.50 (2.12)	1.18
CRP[d]	71	1.12 (1.20)	0.94

[a]Pooled S.D. [b]Measured on a 5-point rating scale. [c]A 10-point scale. [d]Calculated using log-transformed CRP values. GAIS: Gout Attack Intensity Score; ES: effect size; PGA: patient global assessment; NRS: numeric rating scale.

the sensitivity to change of the GAIS (2.32) was greater compared with other single-item instruments, and only the NRS pain had a slightly higher ES of 2.40 (Table 4). CRP proved to be the weakest measure of change, although the improvement still constituted a large effect (>0.8).

At baseline, no floor effects, defined as >15% of patients with the worst possible score, were observed for the GAIS (1.3%), or for its component scores pain (6.6%), tenderness (13.2%) and swelling (10.5%). Floor effects were also largely absent for the NRS pain and NRS PGA, with 2.6% and 0%, respectively. As expected, ceiling effects were absent (0%) for the GAIS, rating scale pain, rating scale swelling, NRS pain and the NRS PGA, and was only 3.9% for the rating scale tenderness.

Discussion

In this study we propose a simple, patient-reported gout attack symptom intensity score (GAIS) for use in gout clinical trials. We aimed to develop a new score that includes information that is already frequently collected in clinical trials, with the hope that future studies will report more comprehensive data on gout attack symptom intensity, besides pain alone. Also, by providing a standardized approach to assessing and scoring these gout-related symptoms, it may be easier to compare outcomes across different studies.

Results of our study suggest that combining the single-item components joint pain, swelling and tenderness into an average score, GAIS, yields reliable scores and leads to a more responsive instrument to assess patient-reported gout attack intensity compared with the individual rating scales it is composed of. In fact, responsiveness of the GAIS was rather similar to the NRS pain, which is usually the most responsive instrument in gout clinical trials [8]. Furthermore, the results of our Mokken scaling analysis support the proposed GAIS scoring rule of simply summing the individual rating scales. Specifically, these findings imply that the GAIS may be interpreted as an ordinal scale. That is, patients can be rank-ordered with respect to the degree of gout attack intensity they experience using the GAIS total score.

As expected, GAIS was able to better discriminate responders from non-responders, compared with its single-item components. This property is of special importance in clinical trials where the objective is usually to differentiate treatment groups by their level of achieved response. Therefore, our results suggest that researchers intending to use the 5-point pain rating scale as a primary endpoint in their studies, which is not uncommonly done, may consider the GAIS instead [1, 3, 22, 23]. This would allow for a more comprehensive assessment of gout attack-related symptoms, and would allow treatment effects to be demonstrated using a smaller numbers of patients.

The current study is also among the first to report on the psychometric properties of all three patient-reported individual symptom rating scales of pain, tenderness and swelling. Thus far, only one previous study examined the measurement properties of the patient-reported 4-point Likert scale pain using data from four randomized controlled trials and one observational study [8]. From this study it was concluded that the construct validity and discriminative ability, both the within-group and between-group discrimination, of the 4-point Likert scale pain measure was supported by sufficient evidence. However, in the same study, only instruments assessing physician-reported joint swelling and joint tenderness on a Likert scale were evaluated, while no information for these outcomes assessed by patients were available. In fact, as far as we know, no previous study has done an evaluation of the psychometric properties of self-reported joint swelling and tenderness.

The GAIS is not the first patient-reported outcome measure developed specifically for the gout population. Besides the Tophus Impact Questionnaire [24], Colwell

et al. developed the Gout Assessment Questionnaire, followed by a second version of this instrument, the Gout Assessment Questionnaire 2.0, by Hirsch *et al.* [25–27]. The Gout Assessment Questionnaire 2.0 aims to retrospectively assess the impact of gout on health-related quality of life. It also includes two NRS that aim to assess pain and disease activity over the last 4 weeks. As such, it is particularly well suited for cross-sectional studies or to assess change over a relatively longer period of time. The GAIS on the other hand is intended to be administered during a gout attack to collect daily prospective information on gout flare symptom intensity, as part of a gout flare diary. Its intended purpose is to help assess effectiveness of anti-inflammatory treatments in clinical trials or observational studies in patients with acute gout attacks.

A strength of this study is that strict selection criteria were maintained for the trial. Also, only patients with crystal-proven gout, the diagnostic gold-standard for gout, were included. However, this study also had some limitations, including the relatively small sample size. Moreover, most patients were recruited from the rheumatology department at hospital centres in the Netherlands. Therefore, whether these results will hold for populations in, for example, primary care where generally less severe gout patients are seen, remains to be determined in future studies. Another limitation of this study was that we were not yet able to relate GAIS scores to other well-known instruments such as the HAQ. Future studies will be needed to demonstrate that GAIS scores relate to these measures in expected ways to further support its construct validity.

In conclusion, the present study introduces the GAIS, an average score of three patient-reported outcome measures, that may be used in clinical studies for assessing gout attack symptom intensity. The GAIS seems to be a reliable and responsive instrument that represents the intensity of gout attacks by considering patient-reported joint pain, tenderness and swelling.

Funding: This work was supported by a grant from The Netherlands Organisation for Health Research and Development (ZonMw) under its programme Rational Pharmacotherapy [836031015].

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

Supplementary data

Supplementary data are available at *Rheumatology* online.

References

1 Terkeltaub RA, Schumacher HR, Carter JD *et al*. Rilonacept in the treatment of acute gouty arthritis: a randomized, controlled clinical trial using indomethacin as the active comparator. Arthritis Res Ther 2013;15:R25.

2 Dalbeth N, Jones G, Terkeltaub R *et al*. Lesinurad, a selective uric acid reabsorption inhibitor, in combination with febuxostat in patients with tophaceous gout: a phase III clinical trial. Arthritis Rheumatol 2017;69:1903–1913.

3 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.

4 Schlesinger N, De Meulemeester M, Pikhlak A *et al*. Canakinumab relieves symptoms of acute flares and improves health-related quality of life in patients with difficult-to-treat Gouty Arthritis by suppressing inflammation: results of a randomized, dose-ranging study. Arthritis Res Ther 2011;13:R53.

5 Rubin BR, Burton R, Navarra S *et al*. Efficacy and safety profile of treatment with etoricoxib 120 mg once daily compared with indomethacin 50 mg three times daily in acute gout: a randomized controlled trial. Arthritis Rheum 2004;50:598–606.

6 Singh JA, Taylor WJ, Dalbeth N *et al*. OMERACT endorsement of measures of outcome for studies of acute gout. J Rheumatol 2014;41:569–73.

7 Schumacher HR, Evans RR, Saag KG *et al*. Rilonacept (interleukin-1 trap) for prevention of gout flares during initiation of uric acid-lowering therapy: results from a phase III randomized, double-blind, placebo-controlled, confirmatory efficacy study. Arthritis Care Res (Hoboken) 2012;64:1462–70.

8 Taylor WJ, Redden D, Dalbeth N *et al*. Application of the OMERACT filter to measures of core outcome domains in recent clinical studies of acute gout. J Rheumatol 2014;41:574–80.

9 Jensen MP, Hu X, Potts SL, Gould EM. Single vs composite measures of pain intensity: relative sensitivity for detecting treatment effects. Pain 2013;154:534–8.

10 Janssen CA, Oude Voshaar MAH, Vonkeman HE *et al*. Anakinra for the treatment of acute gout flares: a randomized, double-blind, placebo-controlled, active-comparator, non-inferiority trial. Rheumatology (Oxford) 2019; doi.org/10.1093/rheumatology/key402.

11 Sijtsma K, van der Ark LA. A tutorial on how to do a Mokken scale analysis on your test and questionnaire data. Br J Math Stat Psychol 2017;70:137–58.

12 van Schuur WH. Quantitative Applications in the Social Sciences: Ordinal item response theory: Mokken scale analysis. Thousand Oaks, CA: SAGE Publications, Inc., 2011: 128.

13 Wind SA. An instructional module on Mokken scale analysis. Educ Meas Issues Pract 2017;36:50–66.

14 Mokken R, Lewis C. A nonparametric approach to the analysis of dichotomous item responses. Appl Psychol Meas 1982;6:417–30.

15 Sijtsma K, Molenaar IW. Introduction to nonparametric item response theory. Thousand Oaks, CA: Sage Publications Inc., 2002: 176.

16 Sijtsma K. On the use, the misuse, and the very limited usefulness of Cronbach's alpha. Psychometrika 2009;74:107–20.

17 Ten Berge J, Sočan G. The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. Psychometrika 2004;69:613–25.

18 Nunnally JC. Psychometric theory, 2nd edn. New York, NY: McGraw-Hill, 1978.

19 Fayers PM, Machin D. Quality of life: the assessment, analysis and interpretation of patient-reported outcomes, 2nd edn. Chichester, England: Wiley, 2007.

20 Cohen J. Statistical power analysis for the behavioural sciences, 2nd edn. Hillsdale, NJ: Erlbaum, 1988.

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

22 Li T, Chen S, Dai Q *et al*. Etoricoxib versus indometacin in the treatment of Chinese patients with acute gouty arthritis: a randomized double-blind trial. Chin Med J (Engl) 2013;126:1867–71.

23 Xu L, Liu S, Guan M, Xue Y. Comparison of prednisolone, etoricoxib, and indomethacin in treatment of acute gouty arthritis: an open-label, randomized, controlled trial. Med Sci Monit 2016;22:810–7.

24 Aati O, Taylor WJ, Siegert RJ *et al*. Development of a patient-reported outcome measure of tophus burden: the Tophus Impact Questionnaire (TIQ-20). Ann Rheum Dis 2015;74:2144–50.

25 Colwell HH, Hunt BJ, Pasta DJ *et al*. Gout Assessment Questionnaire: initial results of reliability, validity and responsiveness. Int J Clin Pract 2006;60:1210–7.

26 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.

27 Hirsch JD, Terkeltaub R, Khanna D *et al*. Gout disease-specific quality of life and the association with gout characteristics. Patient Relat Outcome Meas 2010;1:1.