# DEVELOPMENT AND FIRST VALIDATION OF A DISEASE ACTIVITY SCORE FOR GOUT

Carlo A. Scirè[1] MD PhD, Greta Carrara[1] PhD, Cinzia Viroli[2] PhD, Marco A. Cimmino[3] MD, William J. Taylor[4] MD, Maria Manara[1] MD, Marcello Govoni[5] MD, Fausto Salaffi[6] MD, Leonardo Punzi[7] MD, Carlomaurizio Montecucco[8] MD, Marco Matucci-Cerinic[9] MD PhD, Giovanni Minisola[10] MD, KING Study Group

1. *Epidemiology Unit, Italian Society for Rheumatology, Milano, Italy;*

2. *Department of Statistical Sciences, University of Bologna, Bologna, Italy;*

3. *Research Laboratory and Academic Unit of Clinical Rheumatology, Department of Internal Medicine (DI.M.I.), University of Genoa, Genoa, Italy;*

4. *Department of Medicine, University of Otago, Wellington, New Zealand;*

5. *Department of Medical Sciences, University of Ferrara, Ferrara, Italy;*

6. *Department of Rheumatology, Polytechnic University of Marche, Ancona, Italy;*

7. *Department of Rheumatology, University of Padova, Padova, Italy;*

8. *Division of Rheumatology, IRCCS San Matteo Foundation, Pavia, Italy;*

9. *Department of Internal Medicine, University of Florence, Florence, Italy;*

10. *Department of Rheumatology, San Camillo Hospital, Rome, Italy.*

**Running title**: gout activity score

**Correspondence to**:

Carlo Alberto Scirè

Via Turati 40 – 20121 Milano – Italy

Tel. +39 (0)2 65563677

Fax. +39 (0)2 87152033

c.scire@reumatologia.it

## ABSTRACT

**Objective**: To develop a new composite disease activity score for gout and provide its first validation.

**Methods**: Disease activity has been defined as ongoing presence of urate deposits that lead to acute arthritis and joint damage. Every measure for each OMERACT core domain was considered. A three-step approach (factor analysis, linear discriminant analysis and linear regression) was applied to derive the gout activity score (GAS). Decision to change treatment or 6-month flare count were used as surrogate criterion of high disease activity. Baseline and 12-month follow-up data of 446 patients included in the Kick-off of the Italian Network for Gout (KING) cohort were used. Construct and criterion-related validity were tested. External validation on an independent sample is reported.

**Results**: Factor analysis identified 5 factors: patient-reported outcomes, joint examination, flares, tophi and serum uric acid (sUA). Discriminant function analysis resulted in a correct classification of 79%. Linear regression analysis identified a first candidate GAS including: 12 month flare count, sUA, visual analogue scale (VAS) of pain, VAS global activity assessment, swollen and tender joint count and cumulative measure of tophi. Alternative scores were also developed. The developed GAS demonstrated a good correlation with functional disability (criterion validity) and discrimination between patient- and physician-reported measures of active disease (construct validity). The results were reproduced in the external sample.

**Conclusion**: This study developed and validated composite measure of disease activity in gout. Further testing is required to confirm its generalisability, responsiveness and usefulness in assisting with clinical decisions.

**Keywords**:          gout,          disease          activity,          patient          perspective

2

## SIGNIFICANCE AND INNOVATIONS

- There are currently no composite disease activity measures available for gout. This analysis provides a clinically useful instrument to define disease activity beyond the use of biomarkers as the sole guidance for clinical decisions.

- Based on the results of this study, a composite disease activity measure including relevant domains (serum urate levels, acute flares, tophus burden, patient reported outcomes) able to discriminate between relevant states of disease is proposed.

- Further testing in different datasets is now required to clarify the responsiveness and clinical utility.

3

Disease activity is a challenging concept in gout. Based on the currently accepted model of the disease, high serum levels of urate (UA) lead to joint deposits of uric acid crystals, which may provoke acute arthritis. Acute inflammation resolves but eventually evolves into chronic arthropathy, with development of disability and impairment of quality of life.

Accurate measure of disease activity is an important component of appropriate and targeted treatment strategies, as well as of evaluation of new treatments.

The best candidate to measure disease activity in gout are tissue deposits of UA, because they reflect a still ongoing active disease process (1). Serum level of UA (sUA) well approximates tissue levels and it is the main target of treatment in gout (2). The relevance of using sUA as outcome measure in gout is strengthened by the consideration that drugs without influence on sUA do not have any plausible effect on the disease process (3). Furthermore, sUA levels associate with the risk of acute arthritis in a dose-dependent manner, and their persistent control leads to reduction of tissue deposits and risk of flare (4). However, sUA is only weakly associated with other relevant outcomes such as disability and health-related quality of life in patients with gout (2).

Thus, sUA is a necessary component of the measurement of the activity of the disease process, but other components might play a role (5). As recognized by the OMERACT, several domains should be assessed when evaluating the outcome of patients with gout, some of them relating more to the current disease activity while others mainly to the consequences of the disease (6). Interfering with sUA might be not sufficient to control the activity of the disease process, and to achieve this goal we need a feasible, reliable and valid measure to apply in practice and clinical trials.

An attractive way to describe in a more comprehensive way the activity of gout is to derive a composite measure of disease activity that includes and weights relevant variables, such as sUA, joint inflammation, pain measures, tophi burden (5). Beyond its clinical (and statistical) relevance, the main methodological concerns rely on the face validity of the items to be included, the assumption of a compensation among

4

items and the relative response to treatment of single items within the composite score (7). Previous

studies explored the relative relevance of items belonging to the OMERACT domains to be included in a

composite outcome measure, showing substantial disagreement among clinicians, researchers and patients

(5,8,9). Nevertheless, more recently, a preliminary definition of remission has been proposed, suggesting

the potential validity of measuring a state within the continuum of disease activity in gout (10).

This analysis aims to derive a new composite disease activity score for gout and to provide its first

validation. For this purpose we applied a data-driven approach, analysing data collected in a multicentre

observational study including a random sample of patients with gout with complete baseline and follow-up

clinimetric data. Candidate disease activity scores were developed according to a well-recognised data-

driven process and externally validated.

## PATIENTS AND METHODS

### Study design, setting and recruitment

This is a longitudinal analysis of a multicentre cohort study including a nationwide representative sample of patients referred to 30 rheumatology clinics across Italy, with a clinical diagnosis of gout, recruited between June 2011 and January 2012 (*Kick-off of the Italian Network for Gout* - KING Study, promoted by the Italian Society for Rheumatology SIR, NCT01549210). A probability sample was drawn from clinical registers as previously described and clinical diagnoses validated by the participant rheumatologists (11). All the patients were assessed at baseline, 6 months and 12 months. The study protocol was approved by the local ethics committees of the participating centres. A validation sample of consecutive patients with crystal proven gout and acute symptoms was recruited between September 2015 and November 2015 at 4 KING sites.

### Variables

At baseline, 6 and 12 months all patients underwent full clinical evaluation which followed a structured case report form including both general health and disease-specific variables.

Gout-related variables included symptoms duration, classification according to the 1977 preliminary American College of Rheumatology (ACR) criteria (12), disease-related comorbidities, previous and current treatment for gout, swollen and tender joint count on 66/68 joints, measurement of tophi (count and tape measure of all the clinically evaluable tophi) (13), 0-10 visual analogue scale (VAS) for pain, patient's global assessment of disease activity (PtGA), general health (GH) and physician's assessment of response to treatment (PhGA), flare occurrence (patient-reported acute and significant worsening of joint pain) (14) and sUA levels. All patients completed the Italian versions of the Health Assessment Questionnaire (HAQ),the Short-Form-36 (SF-36), and the Gout Impact Scale(GIS) (15–17). In the external validation sample a 0-100

6

patient acceptable symptom state (PASS) VAS was also measured along with variables to be included in the activity scores and the HAQ.

The underlying construct of disease activity was defined as the presence of tissue deposits leading to acute or chronic inflammation, and as a consequence to pain, joint damage and functional disability (Supplementary material - Figure S1).

As a consequence of this concept of disease activity, decision to change treatment at baseline visit (a new prescription or increasing dosage/switch of urate lowering treatment not related to adverse drug reactions, and/or prescription of symptomatic drugs such as colchicine, NSAIDs or intra-articular/systemic glucocorticoids) or the occurrence of a gout flare within the following 6 months was set as the criterion marker of disease activity to develop the disease activity score.

For the purpose of these analyses the included sample size > 200 subjects is appropriate to develop response criteria according to recommendations of the classification and response criteria subcommittee of the American College of Rheumatology committee on quality measures (18). Therefore, a subsample of 214 patients from the study sample was drawn for the development of the scores. For the purpose of external validation 60 subjects were deemed as sufficient to detect correlation coefficients >0.35.

## Selection of items and statistical methods

Starting from a list of items belonging from the OMERACT domains for studies on gout (6) (pain, serum uric acid, joint swelling, joint tenderness, flare of gout, tophus burden, HRQoL, physician global assessment, patient global assessment, functional disability, work disability), 30 rheumatologist actively involved in the KING study were asked to participate in a Delphi exercise to identify items suitable to be included in a measurement of disease activity for gout.

After a brief introduction to the aim of the survey, participants were asked to assign a 1-10 score of appropriateness to each item to describe different constructs: disease activity, disease-related damage, general health. Invitations to the second and third rounds were sent only to participants who had completed the first round of the survey. Items were selected if more than 70% of the responders thought

7

they were more suitable to describe disease activity rather than damage or general health status, and

rejected if less than 30%. Items with an intermediate level of agreement were proposed again in the second

and third round. Selected outcomes were then filtered based on feasibility, reliability and availability in the

dataset.

Two different methods to derive composite scores were applied.

A three-step method followed the methodology applied to other composite disease scores in rheumatic

disease (19–21). Variables selected from the Delphi exercise were standardised. A factor analysis was then

applied and factors explaining more than 80% of the cumulative variance and eigenvalues > 0.8 were

retained. Factor loadings were then fitted in a linear discriminant function analysis, using the criterion of

low/high disease activity as the classification rule. A classification table was used to evaluate the

misclassification rates of the model. Discriminant scores were then computed for each subject and used as

the dependent variable of a multivariate hierarchical linear regression analysis including transformed

variables as regressors. The combination of variables explaining more than 0.95 of variation of the

discriminant score were selected to be included in the final disease activity score for gout (Gout Activity

Score, GAS). Finally, beta coefficients were normalised and used to weight the transformed variables in

order to compute GAS at every time point.

Given the low number of items to be evaluated in the factor analysis that might have threatened the

validity of the three-step approach, a second, simplified, method included the a priori selected list of items

in a discriminant function analysis. Discriminant coefficients were then normalised and used to derive a

second score. Internal validation was performed by bootstrap (1000 samples) and cross-validation (20

samples) for the three-steps and the one step method, respectively.

GAS scores - developed in a training sample at baseline - were then calculated on the overall study sample

at baseline, 6 and 12 month.

Criterion validity (concurrent and predictive) was tested by comparing the GAS scores against the HAQ

score, SF-36 PCS and MCS subscales by Spearman's correlation coefficients in the overall KING dataset. A

8

further internal validation was performed evaluating the consistency of predictive validity at different time points (e.g. prediction of HAQ from baseline to 6 months and from 6 months to 12 months).

Construct (discriminant validity) was tested computing the standardised mean difference (21) against the decision to change treatment, categorised physician VAS of response to treatment and patient's perception of being in remission, as coded in the GIS (17).

In order to increase applicability to other existing datasets for external validation purpose and to practice, simplified scores were developed, following the same methodology.

The GAS scores were categorised according to the best cut-off using preliminary remission criteria (10) as the classification variable, using the maximal sum of sensitivity and specificity from analysis of the ROC curve at 12 months.

Sensitivity analyses explored different recall periods for the count of the number of flare (past 3 or 6 months), and used only the 'therapeutic' criterion as the classification rule.

In the external validation sample, Spearman's correlation coefficients between GAS score and HAQ score, physician VAS, and PASS were estimated.

Analyses were performed using STATA software package (StataCorp, 2009, release 11, TX, USA) and R Statistical Software (Foundation for Statistical Computing, Vienna, Austria).

9

## RESULTS

### Delphi exercise

Twenty-six of the 30 solicited investigators completed the first round of the survey, and 24 and 23 completed the second and third round, respectively. After three rounds, 8 items (number of recent flares, VAS pain, VAS global patient, VAS response to treatment physician, tender joint count on 68, swollen joint count on 66, measurement of tophi, sUA) covering 7 domains (flare of gout, pain, patient global assessment, physician global assessment, joint inflammation, tophus burden, serum uric acid) were selected to be included in the analyses. (see supplementary material – Table S1).

### KING dataset

Baseline characteristics of the study cohort are given in table 1. The study flow-chart in Figure 1 reports the study cohort included in these analyses.

Out of the 214 subjects selected to develop the scores, 94 had a change in treatment not related to adverse events or reductions in therapy; further 36 subjects without change of treatment experienced a flare within 6 months from baseline. As a result 130 patients (60.7%) were classified as 'active'.

The external validation sample included 61 more patients (15 women) with crystal-proven diagnosis of gout and acute symptoms.

### Three-step derivation

In the first step, factor analysis using standardized variables revealed 5 components which described the following domains: patient-reported outcomes, joint involvement, flares, tophi burden, serum urate levels. These 5 factors explained 93.1% of the total variability.

In the second step, the loadings of every single factor were used as independent variables in the linear discriminant function analysis. The linear discriminant analysis resulted in the correct classification of 79.1%

10

of the cases. The discriminant function using the canonical discriminant function coefficients was used to calculate individual scores.

In the third step, individual discriminant scores were used as dependent variables of a hierarchical linear regression using transformed original variables as independent variables. This model estimated the relative weights for each variable. Bootstrap validation of these coefficients showed absolute bias values <0.01.

### Single step derivation

Variables to be included in the one-step process were selected on the basis of the consensus, when the score related to the concept of disease activity was the highest compared with scores related with disease severity and general health.

Original variables (untransformed) were used as independent variables in a linear discriminant function analysis. The linear discriminant analysis resulted in the correct classification of 78.5% of the cases. Cross-validation confirmed the robustness of this finding (cross-validated correct classification of 79.9%).

### Candidate scores

After normalization of the coefficients, a first (a) and second (b) candidate activity score for gout were constructed using both 3- and 1-step methodology (Table 2). A simplified 4-variables score (c) was also derived, selecting variables with higher contribution to variance and easier to be measured and retrievable on other existing datasets, and clinical practice.

### Validity of candidate measures

Candidate measures showed significant correlations with functional disability, HRQoL, both cross-sectionally – at baseline and follow-up visits - (concurrent validity) and prospectively (predictive validity) (Table 3).

Discriminant validity of the candidate scores and single items were evaluated by the standardised mean difference (Table 4). For all the developed instruments composite scores showed higher discriminatory ability when compared with single variables, for every external criterion of disease activity.

11

### Cut-offs

Based on the operating characteristics of every cut-point – as derived from the ROC table – the cut-off associated with the lowest misclassification rate of patients in remission according to the de Lautour's definition of remission (10) were identified for each candidate score (Table 5).

### Sensitivity analyses

In order to evaluate the robustness of the results, in a first sensitivity analysis, we varied the time-frame for the evaluation of previous flares from 12 to 6 and 3 months, with no major impact on the results (supplementary Table S2-S7). A second set of sensitivity analyses were done using different criteria for the estimation of the coefficient of the disease activity scores using only decision to change treatment as criterion of disease activity, showing only a slight increase of weighting of swollen joint (Supplementary table S8).

### External validation

Based on its metric characteristics, the 3-step methodology showed the best results, and based on feasibility the 4 variable-GAS (GAS -3c) score was evaluated in the validation sample. The 4-variable GAS score showed a significant association with HAQ (Spearman's rho 0.33, p<0.05), VAS Pain (Spearman's rho 0.61, p<0.05), and VAS PASS (Spearman's rho -0.37, p<0.05).

## DISCUSSION

In this paper we report the development of a new composite disease activity measure for gout to be used in clinical practice and clinical research. We derived the gout activity score following a data-driven methodology consistent with that used for other composite measures in rheumatic diseases and in accordance with relevant recommendations (18–21).

Disease activity has been conceptually defined as presence of urate tissue deposits that lead to acute arthritis, which may evolve in chronic arthropathy with development of joint damage and functional disability. This definition is not merely focused on acute symptoms but also on the patient-related consequences of a persistently active disease process. The list of possible measures to be included in the composite score derived from those whose validity, reliability and responsiveness were evaluated by the OMERACT Gout Special Interest Group (22). Experts then chose by consensus those measures that better defined the concept of disease activity compared to other related constructs such as disease severity and general health: sUA, flares, patient reported outcomes, tophi, and joint inflammation. The candidate items comply with the most important ones identified by other studies, as well as the novel definition of clinical remission for gout (3,5,10). The most relevant difference is the inclusion of joint inflammation and the exclusion of measures of function. These differences might be due to the fact that items were entirely derived from the experts and were specifically chosen to discriminate disease activity from other constructs, particularly disease severity. The inclusion of an extensive joint count might threaten the feasibility of our instrument. Also because of the lack of gout-specific restricted joint count, GAS scores without joint count were developed. In the absence of fully validated gout-specific instruments for the evaluation of disability in gout (23), current instruments may potentially underestimating the impact of the disease on function (24). Similarly the clinical measurement of tophi is not the most sensitive method available (e.g. ultrasonography, dual energy computerised tomography), but still the most feasible in clinical practice (13). Acute phase reactants would be of interest as additional item to be included, but because they are not routinely collected in gout databases they were excluded.

13

Despite the experts reached a consensus about the most relevant measures to include in a composite disease activity score, this does not overcome the substantial lack of consensus among clinicians and patients about their relative importance, as shown by recent robust qualitative researches (3,5). Building on these results we applied a methodology to weight variables, following the approach proposed by Van Der Heijde in the first development of the disease activity score for rheumatoid arthritis (19). This is a data driven-approach that statistically identifies the combination of variables and their relative weights that best discriminate between disease states, using an external definition of active disease. In absence of a gold standard of definition of active disease - by analogy from other disease activity measures - we opted for a relevant decision point in patient management: the moment when the rheumatologist considered gout sufficiently active that the patient had to start treatment with or switch or increase the dosage of urate-lowering drugs or symptomatic drugs. This choice reflects the perception of the physician of a poor control of the disease process. Such reference standard described the real process of decision making in practice because the rheumatologists were unaware that their therapeutic decisions were part of the investigation. The choice of this classification criterion for disease activity alone would be quite arbitrary and may be associated with a high misclassification. Alternatively, though the occurrence of flare has a high positive predictive value (almost 100%) in identifying active patients, all patients with no flares in the follow-up might be misclassified as 'not active'. For this reason the performance of a score developed only on the risk of flare is likely to be insufficiently accurate. Based on statistical and conceptual considerations we combined the 'treatment criterion' and 'flare criterion' in order to increase the performance of classification and to cover all the aspects of our definition of disease activity. Sensitivity analyses support the robustness of this approach, showing only a slight decrease of weight for sUA and increase for joint inflammation in the scores developed using only 'treatment criterion' as the classification rule.

Using the KING dataset it was possible to develop the gout activity score because of the prospective follow up of a large number of patients, belonging to the entire spectrum of disease and with a large number of variables prospectively collected. This is a rheumatology practice-based multicentre cohort that ensures high reliability in the assessment but selection of more severe or refractory disease. Of note we included a

14

relatively higher proportion of men than expected in a general population sample (25). Further validation in a primary care setting is worthwhile to define its generalisability.

We developed a first gout activity score including all the relevant variables. The robustness of the classification was also demonstrated by the consistency of two different methodologies, by internal validations, sensitivity analyses and external validation. Alternative scores were developed to make variable collection and score calculations more feasible, without major loss of discriminating ability.

A first validation of the score showed significant correlation with functional disability and physical function both cross-sectionally and longitudinally. All the scores also perform better than single variables in discriminating between physician and patient-reported disease activity, and in predicting the risk of flare throughout the follow-up. Though essentially interchangeable based on the available data, among the set of scores we developed those following the three-step methodology had the best metric properties, as also showed by the consistency of remission cut-offs. The 4-variable ($GAS_{3c}$) is the best candidate to be fully validated and to be applied to clinical practice. The results of the external validation confirm the same correlations observed in the development sample. However results on the responsiveness of the different GAS will drive a more informed choice of the best instrument.

In conclusion, we developed and validated a new instrument to measure disease activity in gout. In practice the gout activity score might be used to assess the disease activity of an individual patient and determine objectively when to modify treatment. In addition, the efficacy of therapeutic strategies might be determined using an outcome measure that incorporates relevant patient-reported outcomes.

15

## ACKNOWLEDGMENTS

16

and Luigi Sinigaglia, Struttura Complessa di Reumatologia DH, Istituto Ortopedico G. Pini, Milano (MI);

Marco Atteritano, Unit of Rheumatology, Department of Clinical and Experimental Medicine University of

Messina, Messina (ME); Marco Broggini, SSD di Reumatologia, A.O. Ospedale di Circolo e Fondazione

Macchi, Varese (VA); Marta Caprioli, UO di Medicina Interna, Istituto Clinico Beato Matteo, Vigevano (PV);

Marta Favero, Rheumatology Unit, Department of Medicine - DIMED, University Hospital of Padova, Padova

(PD); Salvatore Sallì, Rheumatology Unit, Policlinico Universitario Paolo Giaccone, Palermo (PA); Marco

Scarati and Simone Parisi, Rheumatology Unit A.O. Città della Salute e della Scienza di Torino, Torino (TO);

Nazzarena Malavolta and Stefania Corvaglia, Rheumatology Unit, Azienda Ospedaliera Universitaria di

Bologna - Policlinico S. Orsola Malpighi, Bologna (BO); Salvatore Scarpato and Vittorio Veneto,

Rheumatology Unit, Ospedale 'M. Scarlato' ASL Salerno, Scafati (SA).

17

## REFERENCES

1. Stamp LK, Khanna PP, Dalbeth N, Boers M, Maksymowych WP, Schumacher HR, et al. Serum urate in chronic gout--will it be the first validated soluble biomarker in rheumatology? *JRheumatol* 2011;38:1462–1466.

2. Stamp LK, Zhu X, Dalbeth N, Jordan S, Edwards NL, Taylor W. Serum urate as a soluble biomarker in chronic gout-evidence that serum urate fulfills the OMERACT validation criteria for soluble biomarkers. *Semin Arthritis Rheum* 2011;40:483–500.

3. Taylor WJ, Brown M, Aati O, Weatherall M, Dalbeth N. Do Patient Preferences for Core Outcome Domains for Chronic Gout Studies Support the Validity of Composite Response Criteria? *Arthritis Care Res* 2013;65:1259–1264.

4. Perez-Ruiz F. Treating to target: a strategy to cure gout. *Rheumatology* 2009;48:ii9–ii14.

5. Taylor WJ, Singh JA, Saag KG, Dalbeth N, MacDonald PA, Edwards NL, et al. Bringing it all together: a novel approach to the development of response criteria for chronic gout clinical trials. *JRheumatol* 2011;38:1467–1470.

6. Schumacher HR, Taylor W, Edwards L, Grainger R, Schlesinger N, Dalbeth N, et al. Outcome Domains for Studies of Acute and Chronic Gout. *J Rheumatol* 2009;36:2342–2345.

7. Felson DT, Anderson JJ. Methodological and statistical approaches to criteria development in rheumatic diseases. *Baillières Clin Rheumatol* 1995;9:253–266.

8. Prowse RL, Dalbeth N, Kavanaugh A, Adebajo AO, Gaffo AL, Terkeltaub R, et al. A delphi exercise to identify characteristic features of gout - opinions from patients and physicians, the first stage in developing new classification criteria. *JRheumatol* 2013;40:498–505.

9. Taylor WJ, Schumacher HR, Baraf HS, Chapman P, Stamp L, Doherty M, et al. A modified Delphi exercise to determine the extent of consensus with OMERACT outcome domains for studies of acute and chronic gout. *AnnRheumDis* 2008;67:888–891.

10. Lautour H de, Taylor WJ, Adebajo A, Alten R, Burgos-Vargas R, Chapman P, et al. Development of Preliminary Remission Criteria for Gout Using Delphi and 1000Minds® Consensus Exercises. *Arthritis Care Res* 2015:n/a–n/a.

11. Scirè CA, Manara M, Cimmino MA, Govoni M, Salaffi F, Punzi L, et al. Gout impacts on function and health-related quality of life beyond associated risk factors and medical conditions: results from the KING observational study of the Italian Society for Rheumatology (SIR). *Arthritis Res Ther* 2013;15:R101.

12. Wallace SL, Robinson H, Masi AT, Decker JL, McCarty DJ, Yü TF. Preliminary criteria for the classification of the acute arthritis of primary gout. *Arthritis Rheum* 1977;20:895–900.

13. Dalbeth N, Schauer C, Macdonald P, Perez-Ruiz F, Schumacher HR, Hamburger S, et al. Methods of tophus assessment in clinical trials of chronic gout: a systematic literature review and pictorial reference guide. *AnnRheumDis* 2011;70:597–604.

18

14. Taylor WJ, Shewchuk R, Saag KG, Schumacher HR, Singh JA, Grainger R, et al. Toward a valid definition of gout flare: results of consensus exercises using Delphi methodology and cognitive mapping. *Arthritis Rheum* 2009;61:535–543.

15. Apolone G, Mosconi P. The Italian SF-36 Health Survey: translation, validation and norming. *JClinEpidemiol* 1998;51:1025–1036.

16. Ranza R, Marchesoni A, Calori G, Bianchi G, Braga M, Canazza S, et al. The Italian version of the Functional Disability Index of the Health Assessment Questionnaire. A reliable instrument for multicenter studies on rheumatoid arthritis. *Clin Exp Rheumatol* 1993;11:123–128.

17. Hirsch JD, Terkeltaub R, Khanna D, Singh J, Sarkin A, Shieh M, et al. Gout disease-specific quality of life and the association with gout characteristics. *Patient Relat Outcome Meas* 2010;1:1–8.

18. Classification, Measures RCS of the AC of RC on Q. Development of classification and response criteria for rheumatic diseases. *Arthritis Care Res* 2006;55:348–352.

19. Heijde DM van der, Hof MA van 't, Riel PL van, Theunisse LA, Lubberts EW, Leeuwen MA van, et al. Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score. *Ann Rheum Dis* 1990;49:916–920.

20. Helliwell PS, FitzGerald O, Fransen J, Gladman DD, Kreuger GG, Callis-Duffin K, et al. The development of candidate composite disease activity and responder indices for psoriatic arthritis (GRACE project). *Ann Rheum Dis* 2013;72:986–991.

21. Lukas C, Landewé R, Sieper J, Dougados M, Davis J, Braun J, et al. Development of an ASAS-endorsed disease activity score (ASDAS) in patients with ankylosing spondylitis. *Ann Rheum Dis* 2009;68:18–24.

22. Grainger R, Taylor WJ, Dalbeth N, Perez-Ruiz F, Singh JA, Waltrip RW, et al. Progress in measurement instruments for acute and chronic gout studies. *JRheumatol* 2009;36:2346–2355.

23. Singh JA, Taylor WJ, Simon LS, Khanna PP, Stamp LK, McQueen FM, et al. Patient-reported outcomes in chronic gout: a report from OMERACT 10. *JRheumatol* 2011;38:1452–1457.

24. Klooster PM ten, Vonkeman HE, Voshaar MAHO, Bode C, Laar MAFJ van de. Experiences of gout-related disability from the patients' perspective: a mixed methods study. *Clin Rheumatol* 2014;33:1145–1154.

25. Trifirò G, Morabito P, Cavagna L, Ferrajolo C, Pecchioli S, Simonetti M, et al. Epidemiology of gout and hyperuricaemia in Italy during the years 2005-2009: a nationwide population-based study. *AnnRheumDis* 2012. Available at: PM:22736095.

19

## TABLES

**Table 1. Baseline patients' characteristics.**

| *Characteristic* | |
|---|---|
| **Gender (male) [n(%)]** | 368/406 (90.6) |
| **Age [mean (SD)]** | 64 (11.5) |
| **Current smokers [n(%)]** | 68/405 (16.8) |
| **BMI [mean (SD)]** | 28.0 (3.9) |
| **Comorbidities [Median (IQR)]** | 3 (2-4) |
| Hypertension [n(%)] | 287 (70.7) |
| Renal failure [n(%)] | 47 (11.6) |
| Osteoarthritis [n(%)] | 226 (55.7) |
| Cardiovascular disorders [n(%)] | 105 (25.9) |
| Diabetes [n(%)] | 58 (14.3) |
| Liver disorders [n(%)] | 32 (7.9) |
| Neoplasms [n(%)] | 25 (6.2) |
| **Fulfills preliminary ACR criteria [n(%)]** | 373/405 (91.9) |
| **Disease duration (years) [median (IQR)]** | 3.8 (1.6 – 10.2) |
| **Joint involvement [n(%)]** | |
| Monoarticular (1 jt) | 87(21.4) |
| Oligoarticular (2-4 jts) | 239 (58.9) |
| Polyarticular (>4jts) | 77 (19) |
| **VAS pain (0-10), median (IQR)** | 2 (0-5) |
| **VAS patient global (0-10), median (IQR)** | 1 (0-5) |
| **Swollen joints (0-66), median (IQR)** | 0 (0-1) |
| **Tender joints (0-68), median (IQR)** | 1 (0-3) |
| **Presence of tophi [number (%)]** | 79 (19.5) |
| **Tophi dimension, median (IQR)** | 1.5 (0.5-6.5) |
| **Number of flares (3 months) [median (IQR)]** | 0 (0-0) |
| **Flare (previous month) [n(%)]** | 120 (29.6) |
| **Serum urate level [mean (SD)] (mg/dl*) / mmol/l** | 6.3 (1.8) / 0.37 (0.11) |
| **Previous corticosteroids** | 117 (28.8) |
| **Current NSAIDs or colchicine** | 172 (42.4) |
| **Current Allopurinol [n(%)]** | 279 (68.7) |
| **Current Febuxostat [n(%)]** | 55 (13.6) |

*BMI: Body Mass Index; SD: standard deviation; IQR: interquartile range; ARA: American Rheumatism Association; NSAIDs: non-steroidal anti-inflammatory drugs; *cut-off 7mg/dl.*

20

**Table 2. Formulae for the calculation of different candidate gout activity scores.**

$$GAS_{3\,steps-a} = 0.09 * last\ 12\ month\ attacks + 1.03 * \sqrt{SUA} + 0.16 * VAS\ pain + 0.17 * VAS\ patient + 0.07 * \sqrt{TJC} + 0.09 * \sqrt{SJC} + 0.32 * \ln(1 + tophi\ dimension)$$

$$GAS_{1\,step-a} = 0.10 * last\ 12\ month\ attacks + 0.27 * SUA + 0.31 * VAS\ pain + 0.07 * VAS\ patient + 0.03 * TJC + 0.01 * SJC + 0.09 * tophi\ dimension$$

$$GAS_{3\,step-b} = 0.08 * last\ 12\ month\ attacks + 1.05 * \sqrt{SUA} + 0.16 * VAS\ pain + 0.17 * VAS\ patient + 0.07 * \sqrt{TJC} + 0.08 * \sqrt{SJC} + 0.47 * \ln(1 + tophi\ number)$$

$$GAS_{1\,step-b} = 0.10 * last\ 12\ month\ attacks + 0.28 * SUA + 0.31 * VAS\ pain + 0.07 * VAS\ patient + 0.03 * TJC + 0.006 * SJC + 0.20 * tophi\ number$$

$$\mathbf{GAS_{3\,step-c} = 0.09 * last\ 12\ month\ attacks + 1.01 * \sqrt{SUA} + 0.34 * VAS\ patient + 0.53 * \ln(1 + tophi\ number)}$$

$$GAS_{1\,step-c} = 0.11 * last\ 12\ month\ attacks + 0.25 * SUA + 0.36 * VAS\ patient + 0.20 * tophi\ number$$

*GAS: gout activity score; SUA: serum uric acid (mg/dl); VAS: visual analogue scale (0-10 cm); TJC: tender joint count (0-68); SJC: swollen joint count (0-66)*

Table 3. Criterion validity. Cross-sectional and longitudinal correlation between gout activity scores and functional disability or health-related quality of life.

|  | baseline | | 6 months | | 12months | |
|---|---|---|---|---|---|---|
|  | GAS $_{3\text{-steps-a}}$ | GAS $_{1\text{-step-a}}$ | GAS $_{3\text{-steps-a}}$ | GAS $_{1\text{-step-a}}$ | GAS $_{3\text{-steps-a}}$ | GAS $_{1\text{-step-a}}$ |
| *baseline* | | | | | | |
| HAQ score | 0.50 | 0.50 | | | | |
| SF36-PCS | -0.57 | -0.57 | | | | |
| SF36-MCS | -0.31 | -0.31 | | | | |
| *6 months* | | | | | | |
| HAQ score | 0.35 | 0.34 | 0.43 | 0.41 | | |
| SF36-PCS | -0.36 | -0.36 | -0.48 | -0.47 | | |
| SF36-MCS | -0.31 | -0.31 | -0.33 | -0.31 | | |
| *12 months* | | | | | | |
| HAQ score | 0.38 | 0.38 | 0.40 | 0.39 | 0.55 | 0.53 |
| SF36-PCS | -0.34 | -0.36 | -0.41 | -0.40 | -0.60 | -0.60 |
| SF36-MCS | -0.24 | -0.24 | -0.25 | -0.26 | -0.32 | -0.33 |

All Spearman's rho coefficient p values <0.001

|  | baseline | | 6 months | | 12months | |
|---|---|---|---|---|---|---|
|  | GAS $_{3\text{-steps-b}}$ | GAS $_{1\text{-step-b}}$ | GAS $_{3\text{-steps-b}}$ | GAS $_{1\text{-step-b}}$ | GAS $_{3\text{-steps-b}}$ | GAS $_{1\text{-step-b}}$ |
| *baseline* | | | | | | |
| HAQ score | 0.50 | 0.50 | | | | |
| SF36-PCS | -0.57 | -0.57 | | | | |
| SF36-MCS | -0.31 | -0.30 | | | | |
| *6 months* | | | | | | |
| HAQ score | 0.35 | 0.35 | 0.42 | 0.41 | | |
| SF36-PCS | -0.36 | -0.36 | -0.48 | -0.46 | | |
| SF36-MCS | -0.31 | -0.30 | -0.33 | -0.32 | | |
| *12 months* | | | | | | |
| HAQ score | 0.38 | 0.39 | 0.41 | 0.39 | 0.54 | 0.53 |
| SF36-PCS | -0.34 | -0.37 | -0.40 | -0.40 | -0.59 | -0.59 |
| SF36-MCS | -0.24 | -0.23 | -0.25 | -0.27 | -0.33 | -0.33 |

All Spearman's rho coefficient p values <0.001

|  | baseline | | 6 months | | 12months | |
|---|---|---|---|---|---|---|
|  | GAS $_{3\text{-steps-c}}$ | GAS $_{1\text{-step-c}}$ | GAS $_{3\text{-steps-c}}$ | GAS $_{1\text{-step-c}}$ | GAS $_{3\text{-steps-c}}$ | GAS $_{1\text{-step-c}}$ |
| *baseline* | | | | | | |
| HAQ score | 0.45 | 0.44 | | | | |
| SF36-PCS | -0.51 | -0.50 | | | | |
| SF36-MCS | -0.29 | -0.28 | | | | |
| *6 months* | | | | | | |
| HAQ score | 0.30 | 0.29 | 0.38 | 0.37 | | |
| SF36-PCS | -0.30 | -0.30 | -0.43 | -0.41 | | |
| SF36-MCS | -0.27 | -0.27 | -0.32 | -0.31 | | |
| *12 months* | | | | | | |
| HAQ score | 0.34 | 0.34 | 0.36 | 0.35 | 0.50 | 0.49 |
| SF36-PCS | -0.30 | -0.30 | -0.36 | -0.35 | -0.53 | -0.53 |

22

| | | | | | |
|---|---|---|---|---|---|
| SF36-MCS | -0.21 | -0.20 | -0.23 | -0.23 | -0.30 | -0.30 |

All Spearman's rho coefficient p values <0.001

**Table 4. Discriminant validity of the gout activity scores, and single items: standardised mean difference between patients defined as active or inactive according to internal and external constructs.**

| Variable | Disease Activity Criterion | VAS Physician (≤4 vs >4) | Patient reported remission (GIS item) |
|---|---|---|---|
| GAS$_{3\text{-steps a}}$ | 1.18[***] | 1.26[***] | 1.00[***] |
| GAS$_{3\text{-steps b}}$ | 1.18[***] | 1.25[***] | 0.99[***] |
| GAS$_{3\text{-steps c}}$ | 1.12[***] | 1.22[***] | 1.03[***] |
| GAS$_{1\text{-step a}}$ | 1.18[***] | 1.25[***] | 0.92[***] |
| GAS$_{1\text{-step b}}$ | 1.18[***] | 1.24[***] | 0.92[***] |
| GAS$_{1\text{-step c}}$ | 1.12[***] | 1.23[***] | 1.03[***] |
| 12-month attacks | 0.58[***] | 0.69[***] | 0.80[***] |
| √sUA | 0.49[**] | 0.57[**] | 0.70[**] |
| VAS pain | 1.00[***] | 0.92[***] | 0.55[**] |
| VAS patient global | 0.97[***] | 1.16[***] | 0.77[***] |
| √tender joint count | 0.84[***] | 0.77[***] | 0.27 |
| √swollen joint count | 0.75[***] | 0.77[***] | 0.17 |
| ln(1+tophi dimension) | 0.49[**] | 0.37[*] | 0.33 |
| ln(1+tophi number) | 0.48[**] | 0.30 | 0.26 |

**\*\*\* p-value≤0.001; \*\* p-value≤0.01; \* p-value≤0.05; √= square root**

**Table 5. Cut-offs of clinical remission estimated at 12 months.**

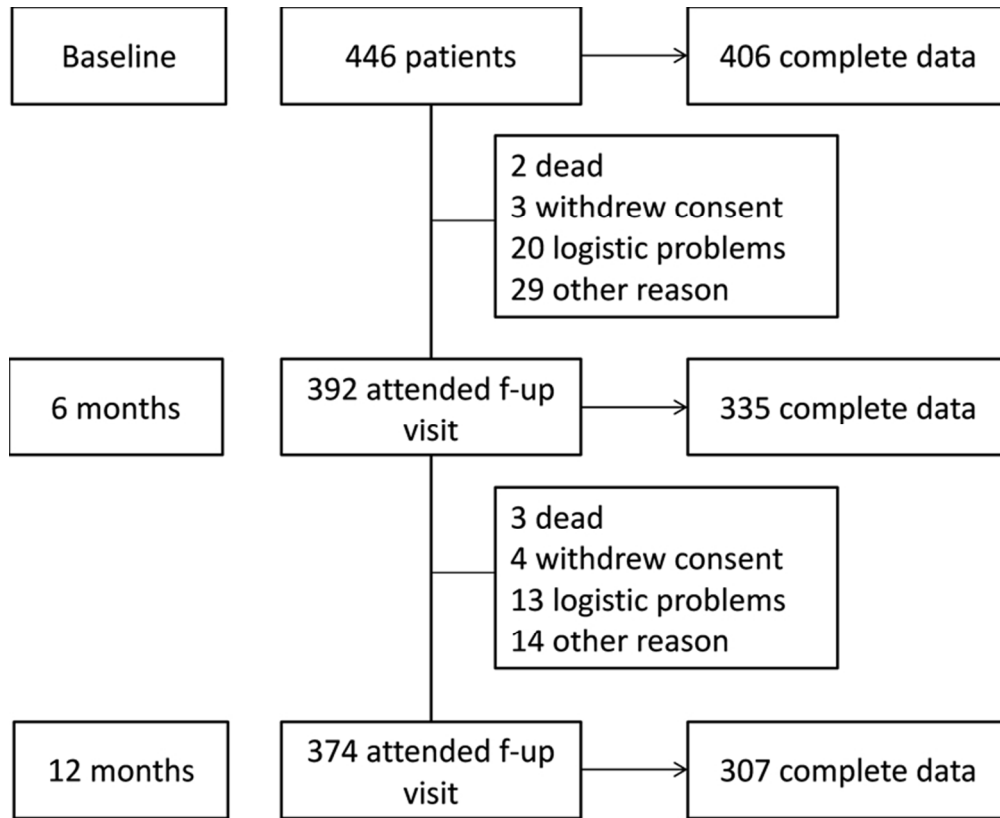| Score | Best cut-off | AUC (95%CI) |
|---|---|---|
| GAS$_{3\text{-steps a}}$ | <2.7 | 0.864 (0.811, 0.916) |
| GAS$_{3\text{-steps b}}$ | <2.7 | 0.865 (0.813, 0.917) |
| GAS$_{3\text{-steps c}}$ | <2.5 | 0.858 (0.806, 0.909) |

23

**FIGURE LEGEND**

**Figure 1. Study flow chart.**

Study flow-chart
145x121mm (150 x 150 DPI)