Arthritis Care & Research
Vol. 68, No. 10, October 2016, pp 1530–1537
DOI 10.1002/acr.22844
© 2016, The Authors. Arthritis Care & Research published by Wiley Periodicals, Inc. on behalf of the American College of Rheumatology.
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

ORIGINAL ARTICLE

Development and First Validation of a Disease Activity Score for Gout

CARLO A. SCIRÈ,[1] GRETA CARRARA,[1] CINZIA VIROLI,[2] MARCO A. CIMMINO,[3] WILLIAM J. TAYLOR,[4] MARIA MANARA,[1] MARCELLO GOVONI,[5] FAUSTO SALAFFI,[6] LEONARDO PUNZI,[7] CARLOMAURIZIO MONTECUCCO,[8] MARCO MATUCCI-CERINIC,[9] GIOVANNI MINISOLA,[10] AND THE STUDY GROUP FOR THE KICK-OFF OF THE ITALIAN NETWORK FOR GOUT STUDY

Objective. **To develop a new composite disease activity score for gout and provide its first validation.**
Methods. **Disease activity has been defined as the ongoing presence of urate deposits that lead to acute arthritis and joint damage. Every measure for each Outcome Measures in Rheumatology core domain was considered. A 3-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 the surrogate criteria of high disease activity. Baseline and 12-month followup data of 446 patients included in the Kick-Off of the Italian Network for Gout 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 analog scale (VAS) of pain, VAS global activity assessment, swollen and tender joint counts, and a 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 a composite measure of disease activity in gout. Further testing is required to confirm its generalizability, responsiveness, and usefulness in assisting with clinical decisions.**

INTRODUCTION

Disease activity is a challenging concept in gout. Based on the currently accepted model of the disease, high serum levels of urate 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.

The Kick-Off of the Italian Network for Gout Study was supported by the Italian Society for Rheumatology as part of the Epidemiology Unit Development Programme.

[1]Carlo A. Scirè, MD, PhD, Greta Carrara, PhD, Maria Manara, MD: Italian Society for Rheumatology, Milan, Italy; [2]Cinzia Viroli, PhD: University of Bologna, Bologna, Italy; [3]Marco A. Cimmino, MD: University of Genoa, Genoa, Italy; [4]William J. Taylor, MD: University of Otago, Wellington, New Zealand; [5]Marcello Govoni, MD: University of Ferrara, Ferrara, Italy; [6]Fausto Salaffi, MD: Polytechnic University of Marche, Ancona, Italy; [7]Leonardo Punzi, MD: University of Padua, Padua, Italy; [8]Carlomaurizio Montecucco, MD: IRCCS San Matteo Foundation, Pavia, Italy; [9]Marco Matucci-Cerinic, MD, PhD: University of Florence, Florence, Italy; [10]Giovanni Minisola, MD: San Camillo Hospital, Rome, Italy.

Dr. Scirè has received advisory board membership fees from AstraZeneca and Novartis (less than $10,000 each). Dr. Cimmino has received consulting fees from Menarini International (less than $10,000). Dr. Taylor has received consulting and advisory board membership fees from AstraZeneca and Pfizer (less than $10,000 each). Dr. Punzi has received consulting and/or speaking fees from Abbott, Fidia, Menarini, Merck, and Pfizer (less than $10,000 each).

Address correspondence to Carlo A. Scirè, MD, PhD, Via Turati 40, 20121 Milano, Italy. E-mail: c.scire@reumatologia.it.

Submitted for publication August 6, 2015; accepted in revised form January 12, 2016.

Significance & 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 data sets is now required to clarify the responsiveness and clinical utility.

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

Tissue deposits of uric acid (UA) are the best candidate to measure disease activity in gout because they reflect a still ongoing active disease process (1). Serum UA (sUA) approximates tissue levels well, and it is the main target of treatment in gout (2). The relevance of using sUA as the 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 a 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 (HRQOL) in patients with gout (2).

Therefore, 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 Outcome Measures in Rheumatology (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 relate mainly to the consequences of the disease (6). Interfering with sUA might not be 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 effective and more comprehensive way to describe 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, and tophi burden (5). Beyond its clinical (and statistical) relevance, the main methodologic concerns rely on the face validity of the items to be included, the assumption of a compensation among 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, analyzing data collected in a multicenter observational study, including a random sample of gout patients with complete baseline and followup clinimetric data. Candidate disease activity scores were developed according to a well-recognized data-driven process and externally validated.

PATIENTS AND METHODS

Study design, setting, and recruitment. This is a longitudinal analysis of a multicenter cohort study (Kick-Off of the Italian Network for Gout [KING] Study, promoted by the Italian Society for Rheumatology [NCT01549210]), 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 (see Appendix A for members of the KING Study Group). A probability sample was drawn from clinical registers as previously described, and clinical diagnoses were validated by the participating 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 centers. 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 months, and 12 months all patients underwent a full clinical evaluation that 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 counts on 66 of 68 joints, measurement of tophi (count and tape measure of all the clinically evaluable tophi) (13), 0–10 on a visual analog scale (VAS) for pain, patient's global assessment of disease activity (PtGA), general health 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) health survey, and the Gout Impact Scale (GIS) (15–17). In the external validation sample, a patient acceptable symptom state (PASS) VAS (0–100) 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 (see Supplementary Figure 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/

acr.22844/abstract). As a consequence of this concept of disease activity, the decision to change treatment at the baseline visit (a new prescription or increasing dose/switch of urate-lowering treatment not related to adverse drug reactions, and/or prescription of symptomatic drugs, such as colchicine, nonsteroidal antiinflammatory drugs, or intraarticular/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 of >200 subjects is appropriate to develop response criteria according to the recommendations of the ACR Subcommittee on Classification and Response Criteria of the ACR 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 to the OMERACT domains for studies on gout (6) (pain, sUA, joint swelling, joint tenderness, flare of gout, tophus burden, HRQOL, PhGA, PtGA, functional disability, and work disability), 30 rheumatologists 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, and 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 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 rounds. Selected outcomes were then filtered based on feasibility, reliability, and availability in the data set.

Two different methods to derive composite scores were applied. A 3-step method followed the methodology applied to other composite disease scores in rheumatic disease (19–21). Variables selected from the Delphi exercise were standardized. 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 normalized and used to weight the transformed variables in order to compute the GAS at every time point.

Table 1. Baseline patients' characteristics*

Characteristic	Value
Sex, no. males/total no. (%)	368/406 (90.6)
Age, mean ± SD years	64 ± 11.5
Current smokers, no./total no. (%)	68/405 (16.8)
BMI, mean ± SD	28.0 ± 3.9
Comorbidities, median (IQR)	3 (2–4)
Hypertension, no. (%)	287 (70.7)
Renal failure, no. (%)	47 (11.6)
Osteoarthritis, no. (%)	226 (55.7)
Cardiovascular disorders, no. (%)	105 (25.9)
Diabetes mellitus, no. (%)	58 (14.3)
Liver disorders, no. (%)	32 (7.9)
Neoplasms, no. (%)	25 (6.2)
Fulfills preliminary ACR criteria, no./total no. (%)	373/405 (91.9)
Disease duration (years), median (IQR)	3.8 (1.6–10.2)
Joint involvement, no. (%)	
Monoarticular (1 joint)	87 (21.4)
Oligoarticular (2–4 joints)	239 (58.9)
Polyarticular (>4 joints)	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, no. (%)	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), no. (%)	120 (29.6)
Serum urate level	
Mean ± SD mg/dl†	6.3 ± 1.8
Mean ± SD mmoles/liter	0.37 ± 0.11
Previous corticosteroids	117 (28.8)
Current NSAIDs or colchicine	172 (42.4)
Current allopurinol, no. (%)	279 (68.7)
Current febuxostat, no. (%)	55 (13.6)

* BMI = body mass index; IQR = interquartile range; ACR = American College of Rheumatology; VAS = visual analog scale; NSAIDs = nonsteroidal antiinflammatory drugs.
† Cutoff is 7 mg/dl.

Given the low number of items to be evaluated in the factor analysis that might have threatened the validity of the 3-step approach, a second simplified method included the a priori selected list of items in a discriminant function analysis. Discriminant coefficients were then normalized and used to derive a second score. Internal validation was performed by bootstrap (1,000 samples) and cross-validation (20 samples) for the 3-step and the 1-step method, respectively.

GAS scores, developed in a training sample at baseline, were then calculated on the overall study sample at baseline, 6 months, and 12 months.

Criterion validity (concurrent and predictive) was tested by comparing the GAS scores against the HAQ score, SF-36 physical component summary and mental component summary subscales by Spearman's correlation coefficients in the overall KING data set. A further internal validation was performed evaluating the consistency of predictive

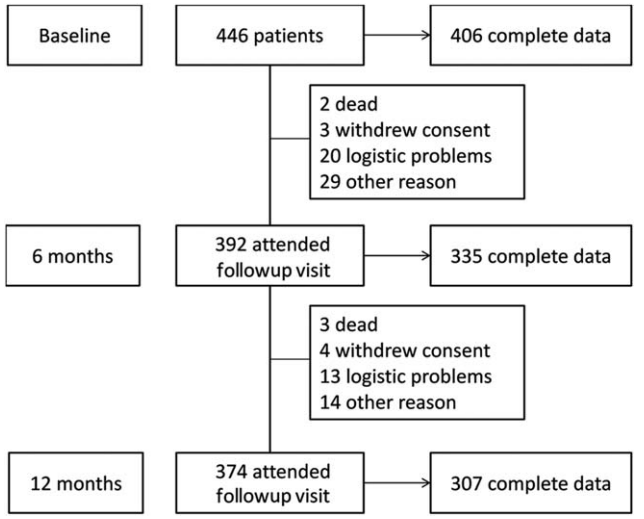

Figure 1. Study flow chart.

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 standardized mean difference (21) against the decision to change treatment, categorized 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 data sets for external validation purpose and to practice, simplified scores were developed following the same methodology.

The GAS scores were categorized according to the best cutoff using preliminary remission criteria (10) as the classification variable, using the maximal sum of sensitivity and specificity from analysis of the receiver operating characteristic (ROC) curve at 12 months.

Sensitivity analyses explored different recall periods for the count of the number of flares (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 (release 11) and R Statistical Software.

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 rounds, respectively. After 3 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, and sUA) covering 7 domains (flare of gout, pain, PtGA, PhGA, joint inflammation, tophus burden, and sUA) were selected to be included in the analyses (see Supplementary Table 1, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.22844/abstract).

KING data set. 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.

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 a 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 a crystal-proven diagnosis of gout and acute symptoms.

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

$$\begin{aligned}
\mathrm{GAS}_{3\ \mathrm{steps-a}} &= 0.09 \times \text{last 12 month attacks} + 1.03 \times \sqrt{\mathrm{sUA}} + 0.16 \times \text{VAS pain} + 0.17 \\
&\quad \times \text{VAS patient} + 0.07 \times \sqrt{\mathrm{TJC}} + 0.09 \times \sqrt{\mathrm{SJC}} + 0.32 \times \ln(1 + \text{tophi dimension}) \\
\mathrm{GAS}_{1\ \mathrm{step-a}} &= 0.10 \times \text{last 12 month attacks} + 0.27 \times \mathrm{sUA} + 0.31 \times \text{VAS pain} + 0.07 \\
&\quad \times \text{VAS patient} + 0.03 \times \mathrm{TJC} + 0.01 \times \mathrm{SJC} + 0.09 \times \text{tophi dimension} \\
\mathrm{GAS}_{3\ \mathrm{step-b}} &= 0.08 \times \text{last 12 month attacks} + 1.05 \times \sqrt{\mathrm{sUA}} + 0.16 \times \text{VAS pain} + 0.17 \\
&\quad \times \text{VAS patient} + 0.07 \times \sqrt{\mathrm{TJC}} + 0.08 \times \sqrt{\mathrm{SJC}} + 0.47 \times \ln(1 + \text{tophi number}) \\
\mathrm{GAS}_{1\ \mathrm{step-b}} &= 0.10 \times \text{last 12 month attacks} + 0.28 \times \mathrm{sUA} + 0.31 \times \text{VAS pain} + 0.07 \\
&\quad \times \text{VAS patient} + 0.03 \times \mathrm{TJC} + 0.006 \times \mathrm{SJC} + 0.20 \times \text{tophi number} \\
\mathrm{GAS}_{3\ \mathrm{step-c}} &= 0.09 \times \text{last 12 month attacks} + 1.01 \times \sqrt{\mathrm{sUA}} + 0.34 \times \text{VAS patient} + 0.53 \times \ln(1 \\
&\quad + \text{tophi number}) \\
\mathrm{GAS}_{1\ \mathrm{step-c}} &= 0.11 \times \text{last 12 month attacks} + 0.25 \times \mathrm{sUA} + 0.36 \times \text{VAS patient} + 0.20 \\
&\quad \times \text{tophi number}
\end{aligned}$$

Figure 2. Formula for the calculation of different candidate gout activity scores. GAS = Gout Activity Score; sUA = serum uric acid (mg/dl); VAS = visual analog scale (0–10 cm); TJC = tender joint count (0–68); SJC = swollen joint count (0–66).

Table 2. Criterion validity: cross-sectional and longitudinal correlation between gout activity scores and functional disability or health-related quality of life*

	Baseline			6 months			12 months		
	HAQ	SF-36 PCS	SF-36 MCS	HAQ	SF-36 PCS	SF-36 MCS	HAQ	SF-36 PCS	SF-36 MCS
Baseline $GAS_{3\text{-step-a}}$	0.50	−0.57	−0.31				0.38	−0.24	−0.34
Baseline $GAS_{1\text{-step-a}}$	0.50	−0.57	−0.31						
6 months $GAS_{3\text{-step-a}}$	0.35	−0.36	−0.31	0.43	−0.48	−0.33			
6 months $GAS_{1\text{-step-a}}$	0.34	−0.36	−0.31	0.41	−0.47	−0.31			
12 months $GAS_{3\text{-step-a}}$	0.38	−0.34	−0.24	0.40	−0.41	−0.25	0.55	−0.60	−0.32
12 months $GAS_{1\text{-step-a}}$	0.38	−0.36	−0.24	0.39	−0.40	−0.26	0.53	−0.60	−0.33
Baseline $GAS_{3\text{-step-b}}$	0.50	−0.57	−0.31						
Baseline $GAS_{1\text{-step-b}}$	0.50	−0.57	−0.30						
6 months $GAS_{3\text{-step-b}}$	0.35	−0.36	−0.31	0.42	−0.48	−0.33			
6 months $GAS_{1\text{-step-b}}$	0.35	−0.36	−0.30	0.41	−0.46	−0.32			
12 months $GAS_{3\text{-step-b}}$	0.38	−0.34	−0.24	0.41	−0.40	−0.25	0.54	−0.59	−0.33
12 months $GAS_{1\text{-step-b}}$	0.39	−0.37	−0.23	0.39	−0.40	−0.27	0.53	−0.59	−0.33
Baseline $GAS_{3\text{-step-c}}$	0.45	−0.51	−0.29						
Baseline $GAS_{1\text{-step-c}}$	0.44	−0.50	−0.28						
6 months $GAS_{3\text{-step-c}}$	0.30	−0.30	−0.27	0.38	−0.43	−0.32			
6 months $GAS_{1\text{-step-c}}$	0.29	−0.30	−0.27	0.37	−0.41	−0.31			
12 months $GAS_{3\text{-step-c}}$	0.34	−0.30	−0.21	0.36	−0.36	−0.20	0.50	−0.53	−0.30
12 months $GAS_{1\text{-step-c}}$	0.34	−0.30	−0.20	0.35	−0.35	−0.23	0.49	−0.53	−0.30

* All Spearman's rho coefficient P values < 0.001. HAQ = Health Assessment Questionnaire; SF-36 = Short Form 36 health survey; PCS = physical component summary; MCS = mental component summary; GAS = Gout Activity Score.

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% 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 1-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 was constructed using both 3- and 1-step methodology (Figure 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 data sets, and clinical practice.

Validity of candidate measures. Candidate measures showed significant correlations with functional disability and HRQOL, both cross-sectionally, at baseline and

Table 3. Discriminant validity of the Gout Activity Scores (GAS), and single items: SMD 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)
$GAS_{3\text{-step-a}}$	1.18†	1.26†	1.00†
$GAS_{3\text{-step-b}}$	1.18†	1.25†	0.99†
$GAS_{3\text{-step-c}}$	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†
$\sqrt{sUA}$	0.49‡	0.57‡	0.70‡
VAS pain	1.00†	0.92†	0.55‡
VAS patient global	0.97†	1.16†	0.77†
$\sqrt{TJC}$	0.84†	0.77†	0.27
$\sqrt{SJC}$	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

* SMD = standardized mean difference; VAS = visual analog scale; GIS = Gout Impact Scale; sUA = serum uric acid; TJC = tender joint count; SJC = swollen joint count.
† $P \leq 0.001$.
‡ $P \leq 0.01$.
§ $P \leq 0.05$.

Table 4. Cutoffs of clinical remission estimated at 12 months*

Score	Best cutoff	AUC (95% CI)
$GAS_{3\text{-step-a}}$	<2.7	0.864 (0.811–0.916)
$GAS_{3\text{-step-b}}$	<2.7	0.865 (0.813–0.917)
$GAS_{3\text{-step-c}}$	<2.5	0.858 (0.806–0.909)

* AUC = area under the curve; 95% CI = 95% confidence interval; GAS = Gout Activity Score.

followup visits (concurrent validity), and prospectively (predictive validity) (Table 2).

Discriminant validity of the candidate scores and single items were evaluated by the standardized mean difference (Table 3). For all the developed instruments, composite scores showed higher discriminatory ability when compared with single variables, for every external criterion of disease activity.

Cutoffs. Based on the operating characteristics of every cut point, as derived from the ROC table, the cutoff associated with the lowest misclassification rate of patients in remission, according to the definition of remission by de Lautour et al (10), were identified for each candidate score (Table 4).

Sensitivity analyses. In order to evaluate the robustness of the results in a first sensitivity analysis, we varied the timeframe for the evaluation of previous flares from 12 months to 6 months and 3 months, with no major impact on the results (see Supplementary Tables 2–7, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.22844/abstract). A second set of sensitivity analyses was done using different criteria for the estimation of the coefficient of the disease activity scores using only decision to change treatment as a criterion of disease activity, showing only a slight increase of weighting of swollen joint (see Supplementary Table 8, available on the *Arthritis Care & Research* web site at http://onlinelibrary.wiley.com/doi/10.1002/acr.22844/abstract).

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 the 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 article 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 the 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 underestimate 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 computerized tomography), but is still the most feasible in clinical practice (13). Acute-phase reactants would be of interest as an additional item to be included, but because they are not routinely collected in gout databases they were excluded.

Despite the experts having 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 research (3,5). Building on these results, we applied a methodology to weight variables, following the approach proposed by van der Heijde et al 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 the absence of a gold-standard 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 to or increase the dose of urate-lowering drugs or symptomatic drugs. This choice reflects the perception of the physician of a poor control of the disease process. Such a 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, although the occurrence of flare has a high positive predictive value (almost 100%) in identifying active patients, all patients with no flares in the followup 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 data set it was possible to develop the GAS because of the prospective followup 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 multicenter cohort that ensures high reliability in the assessment but selection of more severe or refractory disease. Of note, we included a 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 generalizability.

We developed a first GAS, including all the relevant variables. The robustness of the classification was also demonstrated by the consistency of 2 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 followup. Though essentially interchangeable based on the available data, among the set of scores we developed, those following the 3-step methodology had the best metric properties, as also showed by the consistency of remission cutoffs. 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.

ACKNOWLEDGMENTS

The authors acknowledge the contribution of the entire KING staff, including other collaborators and research nurses. They are also grateful to the rheumatologists of the KING centers for their former data collection and for making available their registers of patients with gout.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be submitted for publication. Dr. Scirè had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Scirè, Carrara, Cimmino.

Acquisition of data. Scirè, Cimmino, Manara, Govoni, Salaffi, Punzi, Montecucco, Matucci-Cerinic, Minisola.

Analysis and interpretation of data. Scirè, Carrara, Viroli, Taylor.

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? J Rheumatol 2011;38:1462–6.
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 (Hoboken) 2013;65:1259–64.
4. Perez-Ruiz F. Treating to target: a strategy to cure gout. Rheumatology (Oxford) 2009;48:ii9–14.
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. J Rheumatol 2011;38:1467–70.
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;36:2342–5.
7. Felson DT, Anderson JJ. Methodological and statistical approaches to criteria development in rheumatic diseases. Baillières Clin Rheumatol 1995;9:253–66.
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. J Rheumatol 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. Ann Rheum Dis 2008;67:888–91.
10. De Lautour H, 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 (Hoboken) 2016;68:667–72.
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. Ann Rheum Dis 2011;70:597–604.
14. Taylor WJ, Shewchuk R, Saag KG, Schumacher HR Jr, 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;61:535–43.
15. Apolone G, Mosconi P. The Italian SF-36 health survey: translation, validation and norming. J Clin Epidemiol 1998; 51:1025–36.
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;11:123–8.
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 and Response Criteria Subcommittee of the American College of Rheumatology Committee on Quality Measures. Development of classification and response criteria for rheumatic diseases. Arthritis Rheum 2006;55:348–52.
19. Van der Heijde DM, van 't Hof MA, van Riel PL, Theunisse LA, Lubberts EW, van Leeuwen MA, 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;49:916–20.
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;72:986–91.
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. J Rheumatol 2009;36:2346–55.
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. J Rheumatol 2011;38:1452–7.
24. Ten Klooster PM, Vonkeman HE, Voshaar MA, Bode C, van de Laar MA. Experiences of gout-related disability from the patients' perspective: a mixed methods study. Clin Rheumatol 2014;33:1145–54.
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. Ann Rheum Dis 2013;72:694–700.

APPENDIX A: KING STUDY GROUP

Members of the KING Study Group are as follows: Alarico Ariani (Polytechnic University of Marche, Ancona), Alessandra Galossi (UOC Medicina Interna ad indirizzo Epato-Gastroenterologico, AO S, Giuseppe Marino), Ciro Lauriti (Ospedale Civile Pescara, Pescara), Elena Fracassi, Luca Idolazzi (Azienda Integrata Ospedale, Università di Verona, Verona), Marco Bardelli, Enrico Selvi (Policlinico "Le Scotte," Siena), Enrico Tirri (UOSD di Reumatologia, Ospedale San Giovanni Bosco, Napoli), Federica Furini (University of Ferrara, Azienda Ospedaliero, Universitaria S. Anna, Ferrara), Flora Inverardi (IRCCS Fondazione Policlinico San Matteo, Università di Pavia, Pavia), Andrea Calabrò, Francesco Porta (SOD Reumatologia, AOUC, University of Firenze, Firenze), Raffaele Bittelli, Francesco Venturino (Casa di Cura Reumatologica Oliveti Cotronei), Franco Capsoni (Istituto Ortopedico Galeazzi, University of Milan, Milano), Immacolata Prevete, Giandomenico Sebastiani (Azienda Ospedaliera San Camillo, Forlanini, Roma), Carlo Selmi, Gianluigi Fabbriciani (Humanitas Clinical and Research Center, Rozzano), Giovanni D'Avola (ASP3 Catania, Catania), Giulia Botticella (University of Genova, Genova), Francesca Serale, Giulia Seminara (ASO S Croce e Carle Cuneo), Giuseppe D'Alessandro (SS di Malattie Apparato Osteoarticolare ed Immunoproliferative, Azienda Ospedaliero-Universitaria 'Ospedali Riuniti' di Foggia, Foggia), Leonardo Santo (ASL BT - DSS 4 Barletta, Barletta), Lorena Longato (Ambulatorio di Reumatologia, Poliambulatorio ASL, ASL Biella, Biella), Eleonora Zaccara, Luigi Sinigaglia (Struttura Complessa di Reumatologia DH, Istituto Ortopedico G. Pini, Milano), Marco Atteritano (University of Messina, Messina), Marco Broggini (SSD di Reumatologia, AO Ospedale di Circolo e Fondazione Macchi, Varese), Marta Caprioli (UO di Medicina Interna, Istituto Clinico Beato Matteo, Vigevano), Marta Favero (University Hospital of Padova, Padova), Salvatore Sallì (Policlinico Universitario Paolo Giaccone, Palermo), Marco Scarati, Simone Parisi (Città della Salute e della Scienza di Torino, Torino), Nazzarena Malavolta, Stefania Corvaglia (Azienda Ospedaliera Universitaria di Bologna, Policlinico S. Orsola Malpighi, Bologna), Salvatore Scarpato, Vittorio Veneto (Ospedale 'M. Scarlato' ASL Salerno, Scafati).