Treatment Target and Followup Measures for Patients with Gout: A Systematic Literature Review

Mariano Andrés, Francisca Sivera, Louise Falzon, Désirée M. van der Heijde, and Loreto Carmona

ABSTRACT. Objective. To systematically review the validity of serum uric acid (SUA) as a treatment target for patients with gout, and the clinimetric properties of the potential tools for monitoring these patients.

Methods. A search was performed in Medline, Embase and the Cochrane Library from inception to October 2011, and the 2010–2011 American College of Rheumatology and European League Against Rheumatism meeting abstracts. Studies evaluating different SUA levels or SUA reduction with the achievement of outcomes, and studies assessing clinimetric properties of instruments used to follow patients with gout were selected. Intervention studies were also included in order to estimate responsiveness. Titles and abstracts of the identified references were screened, and included articles were reviewed in detail and data collected using ad hoc standard forms.

Results. In total, 4575 articles were retrieved, 120 articles reviewed in detail, and 54 articles were included in the systematic literature review. SUA reduction was significantly associated with a reduction in acute attacks (6 studies), tophus regression (2 studies), and crystal clearance (3 studies). SUA 6.0 mg/dl was used as cutoff point in most of studies, but this level was found to be arbitrary. For followup of patients with gout, tophus measurement by caliper and ultrasound, the physical component of the Medical Outcomes Study Short Form-36 Survey, and Health Assessment Questionnaire have shown excellent clinimetric properties for this purpose.

Conclusion. Reducing SUA is a valid treatment target for patients with gout, but the target level of reduction (cutoff point) is not clear. Some tools were found suitable for following patients with gout.

Key Indexing Terms:
GOUT   SERUM URIC ACID   IMAGING   VALIDITY   PSYCHOMETRIC PROPERTIES

Gout is an inflammatory joint disorder caused by deposition of monosodium urate (MSU) crystals as a consequence of persistently elevated serum uric acid (SUA) levels. These deposits are reversible; crystals dissolve if SUA is reduced to normal levels, making the inflammatory manifestations disappear, so gout is now considered curable¹. Experts agree that the main treatment target for patients with gout should be the normalization of SUA levels²–⁷. Different SUA cutoff points have been recommended, but the level of 6 mg/dl, stated by the 2006 European League Against Rheumatism (EULAR) recommendations² and the 2012 American College of Rheumatology (ACR) guidelines³, is widely used in trials and clinical practice. The ACR guidelines also affirmed that 5 mg/dl might be appropriate in some patients, in order to durably improve signs and symptoms of gout. British guidelines⁵ also recommended a level of 5 mg/dl, a value based on the median SUA concentration of British men.

This article is part of the 3e (Evidence, Expertise, Exchange) Initiative on Diagnosis and Management of Gout⁸. The objective of the current work was to systematically review the available literature concerning 1 of 10 selected questions as an evidence base for generating recommendations. The question was: “What should the treatment target be and how should patients with gout be followed [i.e., with which measures (patient-reported outcomes, clinical, biochemical, and/or imaging)]?” The proposed clinical question was a complex one: The treatment target in any disease is obtaining either cure or control; these are abstract concepts making measurement extremely difficult. Instead, we commonly use markers highly associated with the cure/control as the target in the...
management of the disease. As mentioned, experts consider that SUA normalization is the main objective of the management of gout. This review will therefore focus on evidence of the different cutoff points for SUA and the clinimetric properties of measures proposed for monitoring patients with gout.

MATERIALS AND METHODS

PICO debriefing. The original question was rephrased to make it correspond to the PICO9 (Population, Intervention, Comparison, Outcome) formula for systematic literature reviews (SLR) as suggested by the Cochrane Collaboration. We built 2 different PICO strategies (see online supplementary material, available from www.3egout.com) as our question had 2 different types of outcomes: (A) For the efficacy of the SUA cutoff levels, we searched for intervention and observational studies that examined the association of SUA levels with relevant outcome domains, such as those established by OMERACT for chronic gout10, plus cure of the disease, resolution of inflammation, and clearance of crystals; and (B) for the followup measures we predefined 5 categories of instruments: laboratory (i.e., inflammatory markers), patient-reported outcomes (quality of life questionnaires), imaging [ultrasound (US)], adherence (patient adherence to treatment), and tophi burden. We searched for studies assessing any of their clinimetric or psychometric properties (as defined by the COSMIN group11). In short, these concepts refer to how close to the truth the instrument is (validity), how stable it is when repeated (reliability), how precise it is in detecting changes (responsiveness or sensitivity to change), and how easy it is to use (feasibility).

The statistical analyses used to test validity were correlation coefficients (r or rho) and regression analyses (r²). r or rho > 0.6 were considered good. For reliability we used Cronbach’s α, intraclass correlation coefficients (ICC), Cohen’s κ, and Bland-Altman plots. For Cronbach’s α, ICC, and Cohen’s κ, the results are considered high when they approach 1. Low dispersion of measures in Bland-Altman plots also represents high reliability. For responsiveness, we focus on measures of effect size [i.e., standardized effect sizes (SES), standardized response means (SRM), and Guyatt statistic]. Although no recommendation or gold standard on how to measure and report responsiveness is available, the results are usually graded as low change (or low sensitive to change) if SES or SRM is < 0.2; moderate change if SES or SRM is 0.2–0.79, and high change when SES or SRM is ≥ 0.8. A Guyatt statistic with a magnitude ≥ 1.00 is also considered indicative of a highly responsive scale. When no validation studies of a specific instrument assessed responsiveness, we estimated SES or SRM from intervention studies that used this instrument, if available. Feasibility was assessed by 3 authors (MA, FS, LC) in a 0–2 scale, where 0 = easy and/or clear to be used, and 2 = hard and/or inaccessible to be used. An average between the 3 reviewers was calculated, and a score of 0–0.5 was considered easy, 0.6–1.4 was considered as medium, and 1.5–2 was considered hard.

Search strategy. Searches were conducted in Medline (from 1948), Embase (from 1980), and the Cochrane Central Register of Controlled Trials (CENTRAL: up to October 16, 2011). In addition, hand searches of the reference list of the selected articles, and of abstracts presented at the 2010 and 2011 ACR and EULAR scientific meetings, were performed. The search strategy was developed along with a librarian with expertise in SLR (LF). For the full search strategy see online supplementary material, available from www.3egout.com. Two independent reviewers (MA and FS) screened the titles and abstracts of all citations identified by the searches, reviewed potentially relevant articles in full text for inclusion according to aforementioned criteria, and performed data extraction of the selected studies. When discrepancies arose and no consensus could be reached, a third author (LC) acted as arbiter. We restricted articles to those published in English or in a language in which at least 1 member of the 3e Initiative bibliographic group was fluent (Dutch, French, German, Spanish). Standardized tools were used to assess the risk of bias of included studies:

RESULTS

From the 4575 articles retrieved from bibliographic databases, and the 2 articles found through the manual reference search, 120 articles were reviewed in detail, and we ultimately selected 54 articles for the SLR (Figure 1). No meeting abstract was included. The full list of included and rejected articles can be found as online supplementary material, available from www.3egout.com. Because our review comprises 2 different questions, we will separate the results referring to SUA cutoff levels from those referring to followup instruments.

SUA as a Treatment Target

Table 1 shows the relevant studies assessing SUA levels. The included studies assessed the association of a SUA reduction, or a specific cutoff point, with reduction in acute attacks, tophi regression, and disappearance of crystals. Six studies found that lower SUA levels were significantly associated with fewer gouty attacks15,16,17,18,19,20. The overall risk of bias was moderate to high. One study15 noted that the recurrence of gout attacks after withdrawal of urate-lowering therapy (ULT) took longer in those patients with lower SUA levels (either during treatment or at discontinuation). Other studies found an association of gouty attacks and high SUA levels compared to lower levels; the study by Sarawate, et al16 showed that the risk of flares was higher in those with SUA > 6.0 mg/dl [odds ratio (OR) 1.59 (95% CI 1.21–2.09)] compared to those with SUA < 6.0. Halpern, et al17 and Wu, et al18 reported similar findings.

Two observational studies addressed the association of SUA levels with changes in tophi size. In the study by Perez-Ruiz, et al,21 all tophi disappeared in all patients after 6 to 60 months of ULT. Interestingly, a strong, inverse correlation between the SUA achieved and the velocity of reduction of the tophi (r = −0.62) was noted; patients with an average SUA below 4.0 mg/dl showed a quicker reduction in tophi compared to those with higher SUA levels (1.52 ± 0.67 mm/month in patients with SUA < 4.0 mg/dl versus 0.53 ± 0.59 mm/month in patients with SUA > 6.1 mg/dl). The results of McCarthy, et al22 were similar: retrospectively, patients whose tophi decreased presented a significantly lower SUA level (6.2 mg/dl) compared to those patients whose tophi did not change or increased (8.2 mg/dl).

Five studies evaluated if lowering SUA is associated with clearance of MSU crystals from joints. The presence of crystals can be detected indirectly — through gout-specific US findings, such as the double-contour sign23 — or directly, by means of joint aspiration and observation with polarized light microscopy. The only study found on US24, with high risk of bias, showed disappearance of the
double-contour sign in all 3 patients who achieved a sustained SUA < 6.0 mg/dl, while the sign persisted in two with SUA > 6.0. Four studies used joint aspiration\textsuperscript{24,25,26,27}. Pascual, \textit{et al}\textsuperscript{25} found that after 3 to 33 months of ULT, MSU crystals disappeared from synovial fluid (SF) in all patients, as SUA diminished. The study by Li-Yu, \textit{et al}\textsuperscript{26}, a cross-sectional analysis of a prospective cohort, showed that patients with sustained SUA < 6.0 mg/dl were associated with a less frequent presence of crystals in SF (44%) compared to those with SUA > 6.0 (88%; OR 0.11, 95% CI 0.02–0.26). So, lower SUA levels appear to be associated with disappearance of MSU crystals from SF.

**Followup Measures**

Table 2 summarizes our findings on clinimetric properties of proposed tools for following patients with gout. As noted above, instruments were divided into 5 different categories, but due to space constraints, results from only 2 categories highly specific to the disease (tophi measurement and gouty bone damage assessment) are presented in detail.

**Tophi measurement.** In our search, 5 different methods for measuring tophi were found (Table 3). Tape measuring showed a high reliability in a single study\textsuperscript{29} (average percentage difference was 29% for the intrarater reliability and 32% for interrater reliability), but no data about validity and responsiveness were found. Two studies tested the properties of a caliper for tophus measurement. In the study by Dalbeth, \textit{et al}\textsuperscript{30}, the intra- and interrater reliability found were very high (ICC = 1). This instrument was also used in the aforementioned study for patients with tophaceous gout\textsuperscript{21}. From mean largest diameter of 18.7 mm at baseline, the final diameter was 0 mm, as tophi resolved in all patients. This is highly representative of the responsiveness of the instrument (if we tried to calculate the effect size, the result would be infinite).

Clinimetric properties of tophus measurement by US were extensively assessed in a prospective cohort\textsuperscript{31}. The validity of US was shown in 2 different ways. First, tophi identified by US were confirmed through crystal identification (face validity). Later, the maximal diameter measured by US correlated (r = 0.65) with the same measurement by magnetic resonance imaging (MRI). The intrarater reliability of the minimal diameter, the maximal diameter, and volume measurements was very high (ICC 0.95, 0.96 and 0.98, respectively). The minimal and maximal diameters also showed high interrater reliability (ICC 0.71 and 0.83, respectively). Responsiveness was assessed after 12 months of followup. The effect size was calculated using the Guyatt statistic for the maximal diameter and volume, and for both measurements the US can be considered as highly sensitive to change (maximal diameter 1.7; volume 1.93).

The reliability of MRI tophus measurement has been
tested in a single study. The results of the intra- and inter-rater reliability were excellent. No study focused on validity, and responsiveness was not identified in our search. Regarding tophus measurement by computed tomography (CT), the study by Dalbeth, et al addressed validity and reliability of the technique in gouty patients. CT measurement strongly correlated with measurement by caliper (r = 0.91), and the intra- and interrater reliability showed very high results (ICC 1 and 0.989, respectively). The responsiveness of these measurements has not been tested to date.

Gouty Bone Damage Assessment
Joint damage in gout was assessed by 2 different techniques:

radiography and CT. A summary of data is shown in Table 4. No data regarding other imaging techniques were found. Two studies evaluated radiography for assessing gout bone damage. From the study by Bloch, et al, the validity of radiography was evaluated compared with control status of the disease (defined by authors as the virtual absence of attack episodes and SUA normalization). We were able to extract data from 40 patients with gouty changes on radiography and data about control status: gouty changes tended to improve in 10 patients with controlled gout and in 1 patient with uncontrolled gout; gouty lesions on radiography remained unchanged in 6 patients whose disease was under control, as well as in 2 with uncontrolled disease; and disease progression was found in 18 patients with recurrent
attacks and high SUA as well as in 3 patients whose disease was under control. With these data, we performed a kappa test, with good agreement between radiographs and clinical disease status (0.70). This result can be considered as supporting construct validity. In the study by Dalbeth, et al, a radiographic damage index for gouty patients was proposed. To assess its validity, the index was compared to different radiographic scores commonly used for different rheumatic diseases (Table 4). The strongest correlation was found with combined Sharp/van der Heijde erosion and narrowing score ($r = 0.88$). The authors also evaluated the reliability of the index, with excellent results (ICC for intrarater agreement $= 0.998$; ICC for interrater agreement $= 0.963$). Responsiveness has not yet been assessed.

In another study, Dalbeth, et al proposed a CT score to determine gouty bone erosion in the feet of patients with tophaceous gout. In a cross-sectional validation study, with unclear risk of bias, the score was compared to several features of the disease to assess validity: the correlation was strong with the radiographic damage index ($r = 0.86$) and presence of clinical tophi ($r = 0.82$), moderate with the duration of disease ($r = 0.42$), and poor with SUA levels ($r = 0.14$). Reliability was assessed as agreement in detection of erosions (Cohen $\kappa = 0.68$) and in the final total erosion score (ICC $= 0.96$), with excellent results. Responsiveness has not yet been assessed.

**Other Instruments**

Results from other categories can be found separately (see online supplementary material, available from www.3egout.com). In short, included studies showed excellent clinimetric properties for the Medical Outcomes Study Short Form-36 (SF-36) and the Health Assessment Questionnaire (HAQ). The physical component summary of SF-36 can be taken as a tool to assess disability.

**DISCUSSION**

There is a clear agreement in guidelines and sets of recommendations that reducing SUA levels should be the treatment target for patients with gout as a surrogate marker of the disease. We have found evidence supporting this,
reductions in SUA are significantly associated with the achievement of other desirable gout outcomes, such as control of acute episodes of inflammation, regression of tophi, and clearance of MSU crystals. However, the level up to which SUA must be reduced (cutoff point) remains unclear. Most studies retrieved in our search used 6.0 mg/dl (360 µmol/l), which is an arbitrary level. One study used 5.05 mg/dl as the cutoff point, because this was the median SUA level while patients were receiving ULT, and this level is associated with a low rate of acute attacks. Also, Perez-Ruiz, et al noted in another study that the velocity of tophi reduction depends on the SUA level achieved, being double when SUA levels reaches 4.0 mg/dl.

Despite these data, most experts keep the SUA target at 6.0 mg/dl. As noted, the 2012 ACR guidelines consider 5.0 mg/dl as the cutoff point.

### Table 3. Instruments for tophus measurement.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Property</th>
<th>Study</th>
<th>Design</th>
<th>No. Patients Included</th>
<th>Results</th>
<th>Risk of Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tape</td>
<td>Reliability</td>
<td>Schumacher, 2005</td>
<td>Cross-sectional</td>
<td>52</td>
<td>Intrarater reliability: APD (area): 29% ± 33; Interrater reliability: APD (area): 32% ± 27; Bland-Altman plot: low dispersion</td>
<td>Low</td>
</tr>
<tr>
<td>Caliper</td>
<td>Reliability</td>
<td>Dalbeth, 2007</td>
<td>Cross-sectional</td>
<td>47</td>
<td>Intrarater reliability: ICC 0.996; Interrater reliability: ICC 0.985</td>
<td>Low</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>Perez-Ruiz, 2002</td>
<td>Prospective cohort</td>
<td>63</td>
<td>Baseline largest diameter: 18.7 ± 10.2 mm; Final diameter: 0 mm; (all tophi disappeared at end of study)</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>CT</td>
<td>Validity Reliability</td>
<td>Dalbeth, 2007</td>
<td>Cross-sectional</td>
<td>47</td>
<td>Correlation with caliper: r = 0.91</td>
<td>Low</td>
</tr>
<tr>
<td>Ultrasound</td>
<td>Validity Reliability</td>
<td>Perez-Ruiz, 2007</td>
<td>Prospective cohort</td>
<td>25</td>
<td>Tophi were confirmed by crystal examination; Correlation of maximal diameter by US with MRI: ( r = 0.65 )</td>
<td>Low</td>
</tr>
<tr>
<td>MRI</td>
<td>Reliability</td>
<td>Schumacher, 2006</td>
<td>Cross-sectional</td>
<td>28</td>
<td>Intrarater reliability: APD (volume): 17.2% ± 25; Interrater reliability: ICC 0.989</td>
<td>Unclear</td>
</tr>
</tbody>
</table>

APD: average percentage difference; ICC: intraclass correlation coefficient; US: ultrasound; CT: computed tomography; MRI: magnetic resonance imaging; min: minimum; max: maximum.

### Table 4. Instrument to evaluate gouty joint damage.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Property</th>
<th>Study</th>
<th>Design</th>
<th>No. Patients Included</th>
<th>Results</th>
<th>Risk of Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-rays</td>
<td>Validity</td>
<td>Bloch, 1980</td>
<td>Retrospective cohort</td>
<td>466</td>
<td>Radiological gouty changes tended to improve or remain unchanged more often in patients with gout under control ((p &lt; 0.001))</td>
<td>Unclear</td>
</tr>
<tr>
<td>Dalbeth, 2007</td>
<td>Validity &amp; Reliability</td>
<td>Cross-sectional</td>
<td>35</td>
<td>Radiograph damage index. Correlations: S-vdH erosion score: ( r = 0.825 ); S-vdH narrowing score: ( r = 0.766 ); S-vdH erosion + narrowing score: ( r = 0.881 ); S-vdH erosion + Ratingen score: ( r = 0.831 ); Ratingen score: ( r = 0.718 ); Steinbrocker score: ( r = 0.86 )</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>CT</td>
<td>Validity</td>
<td>Dalbeth, 2011</td>
<td>Cross-sectional</td>
<td>25</td>
<td>CT erosion score. Correlations: Radiograph damage score: ( r = 0.86 ); Presence of tophi: ( r = 0.82 ); Disease duration: ( r = 0.42 ); SUA: ( r = 0.14 )</td>
<td>Unclear</td>
</tr>
</tbody>
</table>

S-vdH: Sharp-van der Heijde score; ICC: intraclass correlation coefficient; CT: computed tomography; \( \kappa \): Cohen’s kappa.

---

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2014. All rights reserved.
mg/dl for some patients to “improve signs and symptoms of gout,” and in the 2007 British guidelines 5.0 mg/dl is also advised5, but because it is the average SUA level in British males. Gout is associated with an increased risk of coronary heart disease, stroke, and increased mortality in patients with chronic kidney disease61. The mechanism behind this seems to be related to the persistent, subclinical inflammation associated with MSU crystals42,43. Therefore, establishing a very low SUA cutoff that makes crystals disappear faster might be beneficial. To date, no prospective studies have assessed this issue, but favorable opinions are emerging44,45. However, it is important to note that markedly low uric acid levels have been linked to the development of neurological disorders46,47 and even to a higher all-cause mortality48.

Regarding followup of patients with gout, tophus measurement by caliper and US has shown excellent clinical properties for this purpose. A radiological joint damage index appears to be a valid and reliable tool, but data about responsiveness are absent and doubts arise whether it is feasible to use it in clinical practice. In the review of other instruments [see online supplementary material, available from www.3egout.com], the physical component of the SF-36 and the HAQ was noted to be useful for monitoring patients with gout.

In this review, SUA was found to be a valid measure. Reliability of SUA has also been comprehensively reviewed recently49. The problem arises when trying to assess the responsiveness of the SUA measurement. Responsiveness is the capacity of an instrument to detect changes when patients have actually changed. As SUA is the treatment target, changes in patients are demonstrated by reduction in SUA levels. Effect size may even be calculated, but interpretation of the results is tricky, due to its circularity. Taking this into account, we finally decided not to evaluate responsiveness of SUA.

Reducing SUA levels is a valid treatment target for patients with gout, but what remains unclear is to what level (cutoff point) they should be reduced; low SUA levels might bring certain advantages unless proven more risky. For followup of patients with gout, we consider that tophus measurement by caliper and US and the SF-36 (physical component summary) and HAQ have shown excellent clinical properties for this purpose.

REFERENCES

Andrés, et al: Gout monitoring systematic review


24. Thiele RG, Schlesinger N. Ultrasonography shows disappearance of monosodium urate crystal deposition on hyaline cartilage after sustained normouricemia is achieved. Rheumatol Int 2010;30:495-503.


