OMERACT 7 Special Interest Group

Outcome Measures for Acute and Chronic Gout

H. RALPH SCHUMACHER Jr, LAWRENCE N. EDWARDS, FERNANDO PEREZ-RUIZ, MICHAEL BECKER, LAN X. CHEN, DANIEL E. FURST, NANCY JOSEPH-RIDGE, NAOMI SCHLESINGER, ZEB HOROWITZ, KENNETH SAAG, JUDITH A. BOICE, and HISASHI YAMANAKA

ABSTRACT. Gout provides some unique challenges in classification and measurement of outcomes. Our aim was to evaluate criteria for classification and to develop and validate optimal instruments to measure outcomes for acute and chronic gout. A planning committee and interested attendees met to propose classification criteria and domains for outcomes. Seven of the current American Rheumatism Association preliminary criteria for classification were proposed as the best current criteria for acute gouty arthritis, pending further studies. The presence of gout is best established by crystal identification, although this technique has limitations. Five domains for acute gout outcomes and 9 for chronic gout were identified along with proposed instruments for testing and validation. The unique problems of gout evaluation can and will be addressed. (J Rheumatol 2005;32:2452–5)

Key Indexing Terms: GOUT CLASSIFICATION OUTCOMES INSTRUMENTS

Introduction
When the special interest group on gout first convened in October 2003, they identified 4 clear objectives for the meeting: 1. Clarify diagnostic criteria for gout to be used in various settings 2. Establish outcome measures to be used in evaluation of resolution of acute attacks 3. Assess outcomes to be used in the evaluation of chronic gout: Can we validate that lowering serum uric acid to a given level is an outcome that will correlate with clinical significance?

Classification
Our committee, mainly physicians from academic settings and scientists from industry, strongly recommended that the definitive diagnosis of gout as a disease should be made only on the basis of identification of monosodium urate crystals (MSU) from a joint or tophus. Twenty-six percent of clinically suspected diagnoses of gouty arthritis were changed after arthrocentesis and synovial fluid analysis in the study by Eisenberg, et al. This disease has a definitive finding that should be used to establish diagnosis. However, other important questions remain unanswered: (1) Should there be a set of criteria for “probable gout” based on other criteria? (2) Do we also need criteria for acute gouty arthritis? (The latter will likely be needed for clinical trials evaluating treatment for acute gout since the degree of severity of inflammation will need to be standardized.)

Classification criteria for gout were noted to be needed for several purposes, including epidemiology, clinical diagnosis, and evaluation of therapy. Moreover, no prospective
study has been done comparing any criteria with crystal identification. Previous studies used expert opinion for diagnosis. To date, criteria have only been tested by comparing cases of gout versus rheumatoid arthritis (RA), septic arthritis, and pseudogout established by expert diagnosis. How would inclusion of psoriatic, reactive, unclassified, apatite, and palindromic arthritis alter results? Self-diagnosis of gout was substantiated by New York or Rome criteria in a Sudbury, MA, USA, study in only 44% of cases.

A working recommendation was agreed upon: For acute gout treatment trials, 7 instead of 6 of the American Rheumatism Association (ARA) preliminary criteria should be used and then tested. By using 7 criteria instead of 6, sensitivity would decrease to 74.1%, but only 4.4% of other diseases considered would be misdiagnosed as gout. With 6 criteria, sensitivity would be 87.6%, but 19.5% of other diseases would be misclassified as gout. Future prospective studies are proposed to evaluate and test each criterion versus some derivative of the New York and Rome criteria, such as a history of at least 2 attacks of painful limb joint swelling of abrupt onset, with the initial attack having resolved within 2 weeks. If using ARA preliminary criteria could criteria be weighted?

The committee recommended, as a possible gold standard, prospective trials at centers where it is routine to perform joint aspiration, using MSU crystal identification plus signs of acute inflammation (and a course not consistent with infection). Problems of sensitivity and specificity of crystal identification are a concern and will receive further discussion.

For chronic gout treatment, proof of MSU crystals should be required.

It was noted that for epidemiology other classification criteria may be needed. Interestingly, in a recent diet study by Choi, et al., by applying more specific criteria such as synovial fluid MSU crystal identification, correlations between diet and gout increased.

We discussed sources of variation and bias, as reviewed by Whiting, et al., to be considered in testing diagnostic criteria for gout. For example, in comparing clinical criteria for gout with a reference standard such as MSU crystal presence, the following sources should be considered:

- Patient selection: Do we include only those with successful joint aspirations?
- Reference execution: How accurate is MSU crystal identification?
- Test execution. There are subjective aspects of criteria.
- Interpretation and analysis. Is blinding possible?

Outcomes

Discussion about the feasibility of placebo groups in studies of acute gout was felt to be appropriate. Outcome measures in placebo trials might differ from those in studies using an active comparator. The only published placebo controlled trial in acute gout compared colchicine against placebo. A possible approach for placebo based studies is a time-to-rescue model. Attendees felt (10 to 2) that this would be ethical but not feasible. Studies using control groups taking low dose agents, acetaminophen, or nonpharmacologic therapies were also considered. The consensus was, however, that active comparator trials would be predominant. For non-inferiority trials information on effect size and response rate would be needed.

All outcomes selected for consideration still need to be assessed for the OMERACT filter of truth, discrimination, and feasibility. The major discussions focused on identification of core set domains that we could propose and test for outcome measures.

**Acute gouty arthritis.** For attacks of acute gouty arthritis we identified 5 domains (Table 1). Physician global assessment was considered, but not selected in our group vote. There was also discussion about the possible need to identify disease subsets that might alter outcomes. We reviewed the contemporary randomized controlled trials (RCT) that might be used to validate any outcomes and noted 339 subjects in trials from 2002–2004 comparing coxibs to indomethacin.

Instruments to measure selected domains received preliminary discussion and will be the basis for ongoing work by expanded committees.

Pain can be assessed on visual analog scales (VAS) or Likert scales as absolute pain at different times or percentage improvement. Measures included set times such as 2, 4, 8, 12, 36, 48 hours, or queries about time to first evidence of any relief, meaningful relief, and complete relief. Instruments used for dental pain or in studies of rheumatoid arthritis would need some adaptation for acute gout. Recording time of onset of pain and treatment as well as the time of maximal pain were identified as important. Patients generally would enter studies within 48 hours after onset of acute gouty attacks but might have widely differing courses without treatment.

Inflammation as an outcome could be scored as swelling or tenderness on 0–3 point scales as have been used in rheumatoid arthritis. Erythema and heat might be recorded as only positive or negative. Possible uses of systemic markers such as C-reactive protein, interleukin 6, and tumor necrosis factor-α were also discussed, as were ultrasound and other imaging modalities.

**Table 1. Outcome measures for acute gout.**

<table>
<thead>
<tr>
<th>Proposed Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain</td>
</tr>
<tr>
<td>Inflammation</td>
</tr>
<tr>
<td>Function</td>
</tr>
<tr>
<td>Patient Global</td>
</tr>
<tr>
<td>Safety</td>
</tr>
</tbody>
</table>
Function of the target joint might be assessed as follows: 3 = total disability, 2 = movement possible, 1 = weight bearing possible, and 0 = painless full function. Patient global assessment also needs consideration. We proposed to examine utility of the modified Study Short Form-36 (SF-36), VAS, or Likert global, and an experimental gout questionnaire developed by TAP Pharmaceuticals.

Safety will be the final outcome to consider. Adverse effects recording can be done in a variety of ways.

We will also assess whether there are important clinical subsets that influence response, for example: one of a few first attacks or one of a series in chronic disease; involvement of single or multiple joints; which joint is involved; are there associated diseases such as renal insufficiency; what treatments are being used for gout or other diseases.

Chronic gout. Prior to the session committee members reviewed possible outcome domains for chronic gout. In contrast to acute attack, chronic gout is defined in terms of aspects of longterm management including, but not limited to, residual arthritis, tophi, and effects of longterm hyperuricemia. Contemporary randomized controlled trials could be used to validate outcomes, for example, a study of a new selective inhibitor of xanthine oxidase performed on about 2000 subjects and 56 others with a PEG uricase.

After discussion, a list of proposed domains for chronic gout outcome measures was developed (Table 2). Other domains considered but not included in this working list were renal calculi, status of comorbid conditions, pain, and physician global assessment. Difficulties perceived in measuring possible, and 0 = painless full function. Patient global assessment should also be considered in chronic gout, as in acute gout. Participation, which may be closely related to other domains, was unfamiliar but felt to be an attractive outcome to assess impact on all aspects of life.

Optimal instruments to accurately assess safety and to collect adverse drug reactions and interactions are controversial and are being studied by other groups.
Research Agenda
This meeting allowed us to recommend next steps for current and new working group members as we expand participation to include many who attended and expressed interest.

Pending completion of our evaluation of classification criteria, we recommend using the 7 ARA preliminary criteria in trials of acute gout; and MSU crystal confirmation in the diagnosis of chronic gout.

After further discussion we proposed to evaluate the suggested outcome domains via Delphi techniques and then further ranking.

Instruments to examine selected domains will receive the major attention, with discussions on how best to validate and test these with the OMERACT filter. How reliable are measures over time and between individuals? Which of the various measures can best establish the truth (validity, face content, construct, etc.) about an outcome? Are they applicable to all subsets of patients? Are times or rates of achieving outcomes important in chronic as well as acute gout? How long should chronic gout studies be? How measurable is effect size?

ACKNOWLEDGMENT

REFERENCES