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?
4. Expand the group to attract participants in validation studies of proposed domains and instruments.

Efforts in preparation for the OMERACT 7 meeting focused on literature review and extensive discussions of diagnostic criteria. Note that although previous American College of Rheumatology preliminary diagnostic criteria[1] were described for use in identifying acute gouty arthritis, criteria are also needed for gout at any stage. Other sets of criteria noted were results of conferences in Rome in 1963 and New York in 1966, as reviewed by O'Sullivan[2], that have served as criteria for identification of the disease, gout.

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*[3]. 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

From the University of Florida, Gainesville, Florida; Rheumatology Section, Hospital de Cruces, Pais Vasco, Spain; University of Chicago Medical School, Chicago, Illinois, USA; University of Pennsylvania School of Medicine; UCLA Medical School; IM/Rheumatology, TAP Pharmaceuticals; University of Medicine and Dentistry – Robert Wood Johnson Medical School, New Brunswick, New Jersey, USA; Savient Pharmaceuticals, Inc.; Clinical Immunology and Analgesia, Merck Research Laboratories; University of Alabama, Birmingham, Alabama, USA; and Tokyo Women's Medical University, Tokyo, Japan.

H.R. Schumacher Jr, MD, Professor of Medicine, University of Pennsylvania Medical School; L.N. Edwards, MD, Professor of Medicine, University of Florida; F. Perez-Ruiz, MD, Head Assistant, Rheumatology Section, Hospital de Cruces; M. Becker, MD, Professor of Medicine, University of Chicago Medical School; L.X. Chen, MD, PhD, Clinical Assistant Professor, University of Pennsylvania School of Medicine; D.E. Furst, MD, Professor of Medicine, UCLA Medical School; N. Schlesinger, MD, Director of Clinical Rheumatology, UMDNJ – Robert Wood Johnson Medical School; Z. Horowitz, MD, Senior Vice President and Chief Medical Officer, Savient Pharmaceuticals, Inc.; J.A. Boice, PhD, Clinical Immunology and Analgesia, Merck Research Laboratories; K.G. Saag, MD, Associate Professor, University of Alabama; H. Yamanaka, MD, Professor of Medicine, Tokyo Women's Medical University.

Address reprint requests to Dr. H.R. Schumacher Jr, VA Medical Center 151K, University and Woodland Aves., Philadelphia, PA 19104. E-mail: schumacr@mail.med.upenn.edu

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[2].

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*[4], 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*[5], 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[6]. 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 noninferiority 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[7]. 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)[8,9] 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.

Proposed Criteria
Pain
Inflammation
Function
Patient Global
Safety

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[10] 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 measurement and discrimination weighed heavily in exclusion of the first 3.

There was also limited discussion about measurement of total body urate by nuclear medicine[11-13] and depletion of urate crystals from joints by arthrocentesis[14,15]. Both were considered possible outcomes for use in limited studies, to be compared with easier to obtain and more practical outcomes.

Measurement of uric acid pool by infusion of radioisotope labelled uric acid may be considered the gold standard for monitoring reduction of uric acid pool during urate-lowering therapy. Two studies have shown a reduction in the pool size of uric acid[11,13]. In the first study, only one patient underwent pre and post-urate-lowering therapy pool size study. After 2-year treatment with allopurinol 600 mg/day the post-treatment pool became normal. In the second study, pool size was decreased in all 6 patients from 30% to 50% after allopurinol therapy, although data on individual doses and on time taking therapy were not included in the results.

Table 2. Outcome measures for chronic gout.

Proposed Core Set Domains
Serum urate
Gout Flare recurrence
Tophus regression
Joint damage imaging
Health related QOL
Musculoskeletal function
Patient global assessment
Participation
Safety and tolerability

Specialized groups have been able to show that evidence of urate crystal disappearance from synovial fluid is related to level and duration of lowering of serum uric acid[13,14] with lowering of urate to ≤ 6.0 mg/dl for at least 1 year still not totally eliminating crystals from effusions.

Instruments to evaluate proposed domains for chronic gout received preliminary discussion. These include serum uric acid, gout flare, work, tophus regression, radiographic joint damage, and musculoskeletal characteristics.

Consensus was expressed that serum uric acid was an important outcome that should be measured using a specific enzymatic assay, with hyperuricemia defined physiochemically as > 6.8 mg/dl for both men and women. Evidence for the required level for lowered serum urate was reviewed, but will need more discussion.

Gout flares are an important outcome as these are the most easily recognized concerns of patients. Work was felt to be needed to carefully define flares. What is an attack? Some can be quite mild. Discussion will be required on how to quantify incidence over time, severity, effect of response to symptomatic treatments, and monoarticular versus polyarticular flares.

Tophus regression may be evaluated in several ways, including physical measurements with a caliper or tape as described by German[16], or using magnetic resonance imaging[17], ultrasound[18], or serial photographs. Whether any of these are reproducible and feasible will require discussion.

Joint damage imaging has been proposed. Radiographs or MRI might be able to show lack of deterioration or healing. Health related quality of life can be measured in a variety of ways, but none have been validated for gout. We can consider the health associated quality of life and SF-36 questionnaires. It was also proposed to develop a new scale based on querying gout patients about the most important areas of their activities affected by the disease. TAP Pharmaceuticals has been testing a gout assessment questionnaire in their chronic gout studies. Health utility by time tradeoff or standard gamble can also be considered.

Musculoskeletal function can be difficult to measure but is an area identified for further discussion. Might this include joint range of motion or work disability? 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

Others at the meeting in Asilomar, California, and participating in discussions: Manuela Berger, Joel Schiffenbauer, Jane Scott, Bert Spilker, Mark Saifer, Harold Paulus, Jennifer Anderson, Lee Simon, Ariella Kelman, Arthur Kavanaugh, Lachy McLean, Emilio Filippucci, Peter Balint, Richard Wakefield, Philip Conaghan, and William Taylor.

REFERENCES

1. Wallace SL, Robinson H, Masi AT, Decker JL, McCarty DJ, Yü T. Preliminary criteria for the classification of the acute arthritis of primary gout. Arthritis Rheum 1977;20:895-900.
2. O'Sullivan JB. Gout in a New England town. A prevalence study in Sudbury, Massachusetts. Ann Rheum Dis 1972;31:166-9.
3. Eisenberg JM, Schumacher HR, Davidson PK, Kaufman L. Usefulness of synovial fluid analysis in the evaluation. Use of threshold analysis and likelihood ratios to assess a diagnostic test. Arch Intern Med 1984;44:715-9.
4. Choi HK, Atkinson K, Karlson EW, Willet W, Corhan G. Purine rich foods, dairy and protein intake, and the risk of gout in men. N Engl J Med 2004;350:1093-103.
5. Whiting P, Rutjes AWS, Reitsma JB, Glas AS, Bossuyt PMM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy. Ann Intern Med 2004;40:189-202.
6. Ahern MJ, Reid C, Gordon TP, McCredie M, Brooks P, Jones M. Does colchicine work? The results of the first controlled study in acute gout. Aust NZ J Med 1987;17:301-4.
7. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for outcome measures in rheumatology. J Rheumatol 1998;25:198-9.
8. Schumacher HR, Boice JA, Daikh DI, et al. Randomized double blind trial of etoricoxib and indomethacin in treatment of acute gouty arthritis. BMJ 2002;324:1488-92.
9. Schlesinger N, Schumacher HR. Gout: How much of what we do is evidence based? Evidence Based Rheumatol 2003;65-9.
10. Schumacher HR Jr. Febuxostat: a non-purine, selective inhibitor of xanthine oxidase for the management of hyperuricaemia in patients with gout. Expert Opin Investig Drugs 2005;14:893-903.
11. Scott JT, Holloway VP, Glass HI, Arnott RN. Studies of uric acid pool size and turnover rate. Ann Rheum Dis 1969;28:366-73.
12. Loffler W, Grobner W, Medina R, Zollner N. Influence of dietary purines on pool size, turnover, and excretion of uric acid during balance conditions. Isotope studies using 15-N-uric acid. Res Exp Med 1982;181:113-23.
13. Edwards NL, Recker D, Aroizo D, Fox IH. Enhanced purine salvage during allopurinol therapy: an important pharmacologic property in humans. J Lab Clin Med 1981;98:673-83.
14. Li-Yu J, Clayburne G, Sieck M, et al. Treatment of chronic gout. Can we determine when urate stores are depleted enough to prevent attacks of gout? J Rheumatol 2001;28:577-80.
15. Pascual E, Batlle-Gualda E, Martinez A, Rosas J, Vela P. Synovial fluid analysis for diagnosis of intercritical gout. Ann Intern Med 1999;131:756-9.
16. German D. A simple measurement to monitor the size of gouty tophi and rheumatoid nodules. J Clin Rheumatol 1997;3:45-6.
17. Becker M, Schumacher HR, Gruber J, et al. Magnetic resonance imaging (MRI) in the quantitative assessment of gouty tophi [abstract]. Arthritis Rheum 2003:48 Suppl:S528.
18. Nalbant S, Corominas H, Hsu B, et al. Ultrasonography for assessment of subcutaneous nodules. J Rheumatol 2003;30:1191-5.