

Progress in Measurement Instruments for Acute and Chronic Gout Studies

REBECCA GRAINGER, WILLIAM J. TAYLOR, NICOLA DALBETH, FERNANDO PEREZ-RUIZ, JASVINDER A. SINGH, ROYCE W. WALTRIP, NAOMI SCHLESINGER, ROBERT EVANS, N. LAWRENCE EDWARDS, FRANCISCA SIVERA, CESAR DIAZ-TORNE, PATRICIA A. MACDONALD, FIONA M. MCQUEEN, and H. RALPH SCHUMACHER

ABSTRACT. Consensus exercises have identified and prioritized domains of measurement for studies in acute and chronic gout. In parallel, the technical properties of instruments for measurement in many of these domains have been assessed, with the main objective to consider the instruments in the context of the OMERACT filter of truth, discrimination, and feasibility. These data were presented and discussed at OMERACT 9 in the gout workshop, in breakout groups, and at informal meetings of the gout group. In acute gout, instruments for domains of pain, joint swelling, joint tenderness, and patient and physician global assessment have been assessed. In chronic gout, some validation exercises have been performed in instruments for domains serum urate, tophus measurement, health-related quality of life (HRQOL). In voting at OMERACT 9, the Medical Outcomes Study Short-Form 36 was endorsed as a valid instrument for measurement of HRQOL. Methods of tophus measurement were considered to have met some criteria of the OMERACT filter, but these require further work, particularly regarding sensitivity to change over shorter time periods. Priorities for future research include measurement of joint inflammation in acute gout and disability in acute and chronic gout. (J Rheumatol 2009;36:2346–55; doi:10.3899/jrheum.090371)

Key Indexing Terms:

GOUT

OUTCOME MEASURES

INSTRUMENTS

From the Department of Medicine, University of Otago, and the Malaghan Institute of Medical Research; Rehabilitation Teaching and Research Unit, University of Otago, Wellington; Department of Medicine, University of Auckland, Auckland, New Zealand; Sección de Reumatología, Hospital de Cruces, Vizcaya, Spain; Rheumatology Section and Center for Chronic Diseases Outcomes Research, Veterans Affairs Medical Centre Minneapolis, MN; Savient Pharmaceuticals, East Brunswick, NJ; Department of Medicine, UMDNJ-Robert Wood Johnson Medical School, New Brunswick, NJ; Regeneron Pharmaceuticals, Inc., Tarrytown, NY; Department of Medicine, University of Florida, Miami, FL, USA; Sección de Reumatología, Hospital General Universitario de Alicante, Alicante; Hospital de la Santa Creu I Sant Pau, Barcelona, Spain; TAP Pharmaceuticals Products Inc., Lake Forest, IL; Veterans Affairs Medical Center, Philadelphia, Pennsylvania, USA.

Rebecca Grainger is a recipient of a Clinical Research Training Fellowship from the Health Research Council of New Zealand.

R. Grainger, MBChB, FRACP, Department of Medicine, University of Otago, and Malaghan Institute of Medical Research; W. Taylor, PhD, MBChB, FRACP, Rehabilitation Teaching and Research Unit, University of Otago; N. Dalbeth, MD, Department of Medicine, University of Auckland; F. Perez-Ruiz, MD, Sección de Reumatología, Hospital de Cruces; J.A. Singh, MD, Rheumatology Section and Centre for Chronic Diseases Outcomes Research, Veterans Affairs Medical Centre Minneapolis; R.W. Waltrip, MD, Savient Pharmaceuticals; N. Schlesinger, MD, Department of Medicine, UMDNJ-Robert Wood Johnson Medical School; R. Evans, PharmD, Regeneron Pharmaceuticals, Inc.; N.L. Edwards, MD, Department of Medicine, University of Florida; F. Sivera, MD, Sección de Reumatología, Hospital General Universitario de Alicante; C. Diaz-Torne, MD, Hospital de la Santa Creu I Sant Pau; P.A. MacDonald, BSN, NP, TAP Pharmaceuticals Products Inc.; F. McQueen, MD, Department of Medicine, University of Auckland; H.R. Schumacher, MD, Veterans Affairs Medical Center.

Address correspondence to Associate Prof. W. Taylor, Department of Medicine, University of Otago, Wellington, PO Box 7343, Wellington 6242, New Zealand. E-mail: will.taylor@otago.ac.nz

Gout is a common disease characterized by acute self-limiting attacks of arthritis, sometimes progressing to a chronic arthropathy, that is due to deposition of monosodium urate crystals (MSU). Intraarticular MSU crystals initiate an acute inflammatory arthritis, but prolonged hyperuricemia can result in macroscopic subcutaneous and intraarticular deposits, commonly known as tophi. The drive for evidence-based medicine and the advent of new agents for treatment of gout has highlighted the absence of validated outcome measures for use in natural history or intervention studies. Since 2002 the OMERACT Gout Special Interest Group (SIG) has worked towards defining outcome measures for studies in acute and chronic gout^{1,2}. Consensus exercises have identified proposed domains for measurement, organized as mandatory, discretionary, and for further research³. The process by which domains were selected and the ratification of these domains at OMERACT 9 is discussed elsewhere by Schumacher, *et al*⁴.

Considerable progress has been made in identifying measurement instruments for many of these domains and validating identified instruments with respect to the OMERACT filter of truth, discrimination, and feasibility. At OMERACT 9 all studies that addressed any aspect of the OMERACT filter for each instrument were presented in the gout workshop and at an informal gout group meeting. The domains considered mandatory are outlined in Tables 1 and 2, with the potential instrument(s) identified and a summary

Table 1. Summary of the OMERACT filter applied to identified instruments for mandatory domains in acute gout studies.

Domain	Potential Instrument/s	Feasibility	Face Validity	Truth		Content Validity	Reliability	Discrimination		Ref
				Construct Validity	Internal Consistency			Change Sensitivity	Between Group Discrimination	
Pain	Patient assessment of study joint pain 5-point Likert scale (0 = none, 1 = mild, 2 = moderate, 3 = severe to 4 = extreme)	√	√	Yes	NA	√	NR	Large effect sizes	No between intervention-group differences* Differences in pain levels between groups defined by patient reported response to treatment	11
Joint swelling	100 m VAS Investigator assessment of tenderness of study joint 4-point scale (0 = no pain to 3 = patient states there is pain, winces, and withdrawals)	√	√	NR	NA	NR	NR	NR	NR	
Joint tenderness	Investigator assessment of swelling of study joint 4-point scale (0 = none to 3 = bulging beyond joint margins)						Not reported in gout			
PGART	5-point scale (0 = excellent to 4 = poor)	√	√	Yes	NA	√	NR	NR	NR	13
Physician global assessment of response	5-point scale (0 = excellent to 4 = poor)	√	√	No	NA	√	NR	NA	NR	13
Functional Disability	HAQ-DI	√					Not reported in acute gout			

* No placebo group. NR: Not reported in gout; NA: Not applicable; VAS: visual analog scale; HAQ-DI: Health Assessment Questionnaire Disability Index; PGART: Patient global assessment of response to treatment; SF-36: Medical Outcomes Study Short Form 36 Survey.

of the OMERACT filter applied to these outcome measures. A previous review summarizes all instruments reported in the literature⁵, and here we summarize data validating the instruments and highlight unresolved issues, which form the future research agenda.

Methods. During preparation for OMERACT 9 members of the gout group were assigned target domains for identifying studies and meeting abstracts that validated measurement instruments in acute or chronic gout. Summaries were disseminated by email for discussion within the group, and members were encouraged to identify any further relevant studies. Because limited numbers of studies were identified, all were presented at the OMERACT 9 gout workshop, regardless of quality. Further, literature searches were performed using the PubMed database (1968–2008) by one author (RG) using the keyword “gout” and the domains listed in Tables 1 and 2; this confirmed that no further studies required inclusion.

A summary of studies on instruments for measurement of identified domains was presented at the OMERACT 9 gout workshop. At the plenary session attendees voted for each domain they considered to be relevant and to have an instrument of measurement sufficiently assessed as meeting the OMERACT criteria of truth, discrimination, and feasibility.

ACUTE GOUT RESULTS

An acute gout attack is characterized by abrupt onset of inflammatory arthritis in one or more joints. The natural history of an attack is to resolve in 7 to 10 days. Outcome domains focus therefore on pain, joint inflammation (tenderness and swelling), and the impact on the individual (patient global and functional disability). Although potential instruments for each of these domains have been identified, assessment of their technical properties has been limited to single studies without placebo comparator. Although most OMERACT criteria have been fulfilled, further validation in

Table 2. Summary of the OMERACT filter applied to identified instruments for mandatory domains in chronic gout studies.

Domain	Potential Instrument/s	Feasibility	Face Validity	Construct Validity	Truth Internal Consistency	Content Validity	Reliability	Discrimination Change in Sensitivity	Between Group Discrimination	Ref	
Serum urate	Serum urate (Trinder Assay)	√	√	NA	NA	NA	√	ES 1.21 to 4.02	56% to 94% of patients taking febuxostat achieved SUA < 6.0 mg/dl at 28 days vs 0% treated with placebo (p < 0.001)	19, 39	
Flare Tophus regression	Under development Physical measurement (calipers or tape)	√	√	Calipers: High correlation with CT measure of tophus	NA	NA	√	Calipers: ICC intra-observer 0.996, inter-observer 0.985 Tape: Mean average percentage difference 32% (inter-rater). p < 0.05 Site dependent	To be assessed when instrument developed. No between intervention-group differences† Tape: After one year, greater reduction in size if achieved SUA < 6 mg/dl (75% vs 50%). Sensitive to change over 2 yrs.	20, 23–25	
	US	Requires trained operator and equipment	√	80% of tophi detected with MRI detected with US	NA	NA	√	ICC intra-observer > 0.9, inter-observer 0.71–0.83 > 0.8 SDD 4mm to 9 mm	Guyatt ES 2.8 Change in tophus size different for those whose SUA controlled vs those whose SUA not controlled (p = 0.003)	20, 21	
	CT	Requires trained operator and equipment	√	High correlation with caliper measure of tophus. Detected 89% of tophi detected by physical examination	NA	NA	√	ICC intra-observer 1.00, inter-observer 0.989	ND	ND	25
	MRI	Requires trained operator and expensive equipment	√	ND	NA	NA	√	Mean tophus volume difference: Intra-reader difference 0.05+/- 0.97cm ³ Inter-reader 0.89+/-2.05 cm ³	ND	ND	26
HRQOL and functional disability	SF-36	√	√	√	√	√	√	Stable in subjects with no flare	In subjects with frequent flares	Discriminates between groups on basis of disease severity, physicians global and higher vs lower flare frequency	28
	HAQ-DI	√	√	Strong correlation with SF-36, clinical severity	√	√	√	√	In individuals with change in pain score	Discriminates between groups based on clinical severity score	30
	GAQv2.0-GI	√	√	√	√	√	√	√	Not reported	Discriminated between groups based on disease severity and flare frequency	
Pain	Patient assessment of pain due to gout 5-point Likert scale (0 = none to 4 = extreme) 100 mm VAS									ND in chronic gout	

continued next page

Table 2. Continued.

Domain	Potential Instrument/s	Feasibility	Face Validity	Truth		Content Validity	Reliability	Discrimination Change in Sensitivity	Between Group Discrimination	Ref
				Construct Validity	Internal Consistency					
Patient global	Patient global assessment of disease						ND in chronic gout			
Work disability	NI									
Joint inflammation	NI ? 66/68 joint count or 44/46 Joint count (Ritchie Index)									

† No placebo group. NA: Not applicable; US: ultrasound; ND: not assessed; VAS: visual analog scale; SUA: serum uric acid; HRQOL: health related quality of life; CT: computerized tomography; MRI: magnetic resonance imaging; ES: effect size; NI: not identified; GAQ: Gait Assessment Questionnaire; HAQ-DI: Health Assessment Questionnaire Disability Index; SDD: smallest detectable difference; SF-36: Medical Outcomes Study Short Form 36 Survey.

other data sets and through assessment of between-group differences would be desirable.

Pain. Intense pain is a key feature of acute gout attack; hence pain has been a primary outcome measure in all published randomized controlled trials in acute gout⁵⁻⁸. A majority of studies have used a 5-point Likert scale^{5,7}, with others reporting pain on 100 mm visual analog scale (VAS)^{6,8}. Only one study assessed technical properties of the 5-point Likert pain scale against the OMERACT filter using data from 2 parallel-group, 7-day randomized controlled trials of etoricoxib and indomethacin^{9,10} for treatment of acute gout¹¹. Daily pain assessments were recorded by the subjects. In these 339 individuals pain assessment demonstrated construct validity (expected correlation with other clinical indicators of disease activity), sensitivity to change (with large effect sizes), and discrimination between groups defined by patient global assessment of response to treatment (PGART), investigator's global assessment of response to treatment (IGARTJ), and discontinuation due to lack of efficacy. Discrimination between treatment groups has not been shown since these trials included active comparators. Nevertheless, at plenary voting 96% of participants agreed that pain was a relevant domain and that this instrument be accepted.

Further discussion may be required regarding the reporting of this instrument; an outcome metric for pain could include absolute or percentage pain reduction at set times; however, the minimum detectable and clinically significant difference has not been established for acute gout. This could be expressed as categorical data (e.g., proportion of subjects reaching level of pain reduction at given time) or continuous data (e.g., mean percentage pain reduction at given time). Given that acute gout studies are generally of short duration, typically 1 to 8 days, it is not unwieldy to present data in both formats. Other pain outcomes could include time to first evidence of any relief, meaningful

relief, and complete relief; although these have not been validated. The clinimetric performance of Likert versus VAS pain scale has not been compared in gout trials, but data from the osteoarthritis literature suggest these instruments perform similarly¹².

Joint swelling and joint tenderness. Recent studies assessing joint swelling and tenderness have typically used a 5-point scale in an index joint^{7,9,10}. In subjects with more than one joint involved, a "study joint" must be selected. The method of index joint selection is not defined, but patient nomination of most severe joint is a practical solution. It may be necessary for investigators to exclude certain joints where other factors, like overlying tophi, ulceration, or concomitant degenerative joint disease, may preclude assessment or influence response. The technical performance of these 5-point scales for joint swelling and tenderness in gout has not been assessed. At plenary voting, domains joint swelling and joint tenderness had 79% and 72% agreement that these domains are relevant, and it was recognized that a future research agenda must address validation of 5-point scales for tenderness and swelling, choice of index joint, and definition of a response metric.

Patient global. In the gout group meeting, Naomi Schlesinger presented data validating PGART¹³ using the clinical data set described for pain assessment^{9,10}. PGART was reported on a 5-point Likert scale. At 2 and 8 days, the PGART showed moderate correlations with 5-point Likert Pain Assessment, joint tenderness, joint swelling and IGART and swelling. Since global assessments of change/response require a comparison with current state to a previous state, they are subject to recall bias and may be unduly influenced by current state, i.e., an individual with high disease activity at final assessment may respond differently to one with low disease activity, despite similar extent of change from baseline. Similarly individuals with less

severe disease at baseline may show greater improvements in global ratings. For example, a subgroup analysis of the data from etoricoxib-indomethacin equivalency studies showed participants with patients with monoarticular disease had significantly greater improvements in IGART and PGART compared to those with oligoarticular disease¹⁴. At plenary voting PGART was endorsed as a valid domain with 83% in agreement, however physician global assessment of response was not endorsed, with only 62% agreeing this domain was relevant. As only one study has validated the PGART instrument, and because this validation exercise used data sets from active comparator trials, further validation using treatment groups that show wider variation in response and that include patients with larger spectrum of disease activity will be worthwhile.

Functional disability. There are no data on functional disability measurement during acute gout. This clearly remains an area requiring further work. The Health Assessment Questionnaire (HAQ) is a potential instrument for this domain but has not been validated. The preliminary International Classification of Functioning, Disability and Health (ICF) core set for acute arthritis¹⁵ may be a useful guide to comprehensively consider all the ways in which acute gout may affect function. Although no instrument for this domain has been formally identified or validated, 75% of plenary voters agreed that this domain is relevant.

CHRONIC GOUT RESULTS

The term chronic gout encompasses individuals with recurrent attacks of acute gout (intercritical gout) or who have clinically evident tophi. Tophi form in subcutaneous tissue and intraarticularly in the context of persistent hyperuricemia. The arthritis in chronic gout includes acute gout attacks (also known as “flares”) due to acute inflammation and chronic granulomatous synovitis, both in response to MSU crystals. Morbidity in chronic gout is due to the arthritis and the consequences of tophi.

Serum urate. In chronic gout the aim of gout therapy is “cure,” i.e., complete dissolution of MSU crystals within the joint. Supersaturation of urate with crystal formation occurs at 37°C at 6.8 mg/dl. The incidence of gout also increases with increasing serum urate. Serum urate testing is an integral component of the clinical management of gout. In the recent Delphi exercise regarding outcome measures for gout studies, serum urate was considered mandatory for studies of chronic gout, with the highest median rating. It is important to recognize that as an outcome measure, serum urate is a surrogate for outcomes relevant to patients, such as tophus regression and attacks of gout. Thus work is planned to collate the published data that demonstrate legitimacy of serum urate as a surrogate according to the OMERACT surrogate marker criteria. This work will also formally consider if serum urate lowering as an outcome measure fulfills the OMERACT filter.

Although there is clear agreement that serum urate lowering is important in studies of chronic gout, there are a number of issues regarding reporting of serum urate. These issues were discussed by the gout group during OMERACT:

Should there be a “cutoff” target serum urate and if so, what should the “cutoff” be? There is consistent evidence demonstrating that serum urate lowering to a target of 6 mg/dl is associated with clinical improvement in gout. This has been demonstrated in relation to clearance of MSU crystals from asymptomatic joints, reduction in gout flares, and regression of subcutaneous and intraarticular tophi¹⁶⁻²¹. However, some studies have suggested that for prevention of flares and regression of tophi, achieving a target lower than 6 mg/dl has further benefits^{18,20,21}, especially for the most severe chronic gout disease. This suggests that use of a “cutoff” target of 6 mg/dl as the sole measure of urate lowering may not be sufficient to discriminate between those with a “good” and “excellent” response to therapy, and clinical subgrouping of patients depending on severity and expected outcomes would be of interest for further studies.

How should serum urate lowering be represented? Within clinical trials and longitudinal studies of chronic gout, serum urate lowering has been reported in a number of different ways. This includes as a continuous variable (e.g., mean serum urate during study period, mean post-baseline serum urate during study period, percentage reduction in serum urate, area under the curve), and as a dichotomous variable (e.g., presence of serum urate < 6 mg/dl at the last study visit or at the last 3 study visits). The group agreed that given the documented clinical benefits of serum urate lowering below 6 mg/dl, the dichotomous target of 6 mg/dl should be included within the reporting of serum urate lowering for all clinical trials. However, this dichotomy is associated with loss of information (e.g., average value and distribution) in comparison to representation of serum urate as a continuous variable. Reporting serum urate as a continuous variable may therefore also be useful.

What timepoints should be used? Assessment of more than one serum urate over time is optimal, as this measure can vary within individuals depending on the presence of an acute gout flare, and other factors such as diet, hydration state, renal function, and alcohol consumption.

Serum urate lowering has different clinical impacts depending on the timepoint measured. The pharmacokinetic profile (such as half-life or distribution volume) for each urate-lowering drug tested may influence the result of an analysis. Time from intake to serum urate determination may range from 10 to 24 hours. When therapy is first initiated, intensive urate lowering is associated with frequent gout flares (i.e., worse outcomes). The benefits of serum urate lowering are seen later in the course of treatment, typically after one year of therapy. These findings suggest that outcomes in chronic gout studies should include serum urate lowering over longer timepoints (at least one year).

How reliable is the measure? Serum urate measurement using the Trinder assay with uricase is reliable with typical between-laboratory and between-method coefficients of variation of < 5%. Antioxidants such as vitamin C may affect the results, but tube additives such as heparin or EDTA do not. Following uricase therapy, *ex vivo* degradation of urate can occur within the collection tube, leading to spuriously low serum urate measurements. This degradation can be prevented by strict sample handling, with rapid processing of blood for serum under refrigerated conditions immediately followed either by assay or frozen storage until assay.

In summary, serum urate appears to meet the OMERACT filter of truth (face validity), discrimination (reliability, sensitivity to change, between group sensitivity), and feasibility for studies in chronic gout. The dichotomous target of 6 mg/dl should be included within the reporting of serum urate lowering for all clinical trials. Further work is needed to clarify whether supplementary reporting of serum urate (such as lower targets or additional continuous measures) improves the sensitivity of the measure in clinical trials. Plenary voting endorsed inclusion of serum urate as an outcome measure in chronic gout, with 83% endorsement.

Flare. Gout flare, or an acute attack of gout, is a significant concern for individuals with gout and thus a key outcome measure. Potential items for inclusion in an operational definition of gout flare have been identified by consensus²². An American College of Rheumatology-European League against Rheumatism initiative is currently undertaking an observational study to test the accuracy and reliability of these items, or combination of items, for determining the presence of a gout flare. An operational definition can then be tested in randomized controlled trials. Two unresolved issues surrounding flare as an outcome include: (a) Is flare frequency sufficiently discriminating in longterm trials of urate lowering therapy? (b) What is the minimal clinically important change or reduction in flare frequency that patients and physicians would consider important? Plenary voting had 84% of voters in agreement with flare being included as an outcome domain, pending development of an instrument of measurement.

Tophus measurement. Tophus formation is a frequent manifestation of chronic gout. These lesions represent collections of MSU crystals, surrounded by inflammatory cells and connective tissue. Tophi may cause pain, cosmetic problems, mechanical obstruction of joint movement, and joint destruction. Given the clinical relevance of these lesions, change in tophus size is likely to be an important outcome measure in clinical trials of chronic gout. In the recent Delphi exercise regarding outcome measures for clinical trials, tophus regression was considered mandatory for studies of chronic gout. However, the optimal method for measuring tophus size remains uncertain at present. A number of potential methods of assessment have been studied:

a. Physical measurement of subcutaneous tophi using a tape

measure: A validation exercise of tophus area assessment using a tape measure has been reported. In this study, the mean (\pm SD) difference in tophus areas between visits was -0.2 ± 835 mm² (95% CI -162 to 162 mm²) and the mean (\pm SD) average percentage difference (APD) was $29\% \pm 33\%$. The mean (\pm SD) APD between raters was $32\% \pm 27\%$ ²³. Large variations in measurements were noted for elbow tophi. This method has been used in a randomized clinical trial of patients on urate lowering therapy¹⁹. After one year of urate lowering therapy, there was no significant difference between febuxostat and allopurinol treated groups. However, posthoc analysis did show a trend to greater tophus regression in those with mean postbaseline serum urate < 6 mg/dl at week 52 (75% vs 50%, $p = 0.06$). In more recent studies, this method has demonstrated sensitivity to change in response to effective urate lowering therapy over 2 years²⁴.

b. Physical measurement of tophus size using calipers: An alternative method of physical measurement is assessment of the longest diameter of the tophus using Vernier calipers. This method of tophus size assessment was recently validated in a study comparing physical and computed tomography (CT) assessment of tophus size²⁵. In this study, the intraclass correlation coefficient (ICC) for intraobserver reproducibility was 0.996 and for interobserver reproducibility was 0.985. There was strong correlation between CT and physical tophus measurement ($r = 0.91$, $p < 0.0001$), and physical measurement had similar reliability to CT measurement. This method was also used in a 5 year longitudinal observational study of patients treated with urate lowering therapy²⁰: velocity of reduction was measured by analyzing the time to complete resolution of the sentinel tophus; this measure correlated with intensity of urate lowering ($rr = -0.62$, $p < 0.05$).

c. Magnetic resonance imaging (MRI) measurement of tophus size: A multicenter study assessed the intra- and inter-reader reproducibility of quantitative tophus volume measurements from MRI scans in subjects with palpable gouty tophi²⁶. After optimization of the protocol, subjects underwent pre- and post-gadolinium-enhanced MRI scans of a selected tophus on 2 occasions separated by 5–10 days. Unenhanced spin-echo images provided satisfactory tophi images and were less subject to interfering artifacts than gadolinium-enhanced gradient-echo images. Intrareader reproducibility was excellent, with no statistically significant difference in mean tophus volume between visits (mean difference -0.05 ± 0.97 cm³). A small, but statistically significant difference in inter-reader mean tophus volume was detected (mean difference 0.89 ± 2.05 cm³; $p < 0.05$). This study demonstrates that MRI scanning can quantify tophus size in gout, and accurate measurement does not require contrast media. The sensitivity to change of this method, in response to therapy or over time, has not yet been assessed.

d. CT measurement of tophus size: CT scanning has high

sensitivity and specificity for detection of tophus size. A recent study has analyzed the reliability of CT scanning for measurement of the volume of subcutaneous tophus in hands²⁵. In this study, the ICC for intraobserver reproducibility was 1.000 and for interobserver reproducibility was 0.989. CT also identified most, but not all, subcutaneous nodules identified as tophi by physical examination. The sensitivity to change of this method, in response to therapy or over time, has not yet been assessed.

e. Ultrasonographic (US) measurement of tophus size: US measurement of deep, intraarticular or periarticular tophi not accessible to physical measurement (tendon or ligament) was tested for validity using aspiration of nodules and MRI²¹. Over 80% of nodules aspirated yielded MSU crystals, and all nodules greater than 1 cm in diameter were aspirate positive. Although correlation between MRI and US measurement of tophus diameters was good, variability was high. For US, the ICC for intraobserver reproducibility were > 0.9, and for interobserver reproducibility were 0.71-0.83. In a 12-month, prospective, observational, blinded for the observer, urate-lowering therapy intervention study, US measurement showed a good effect size, and a strong correlation was reported between average serum urate levels during urate-lowering therapy and change in both maximal diameter and volume of tophi²¹. This study indicates that US measurement of tophus size is reliable, valid, and sensitive to change in the short term. This method has not yet been tested in randomized studies. The use of new volumetric probes (that may eliminate acquisition bias), and paired-reading of acquired images may improve accuracy.

In summary, all methods of tophus measurement assessed to date have face validity. To fulfil criterion validity, methods would have to compare favorably to a “gold standard;” however, in this context an accepted gold standard does not exist. Good agreement between instruments provides evidence that these measures have construct validity. Reliability of the methods also appears to be acceptable. A further discrimination issue related to all methods of tophus measurement is sensitivity to change, either in response to effective therapeutic agents or over time. The few studies published to date have involved long time periods (1 to 5 yrs), and the time required to reliably observe changes in tophus size is currently unknown. It should be noted that even when using a highly effective urate lowering drug such as febuxostat, statistically significant differences in tophus size by physical measurement were not observed at one year¹⁹.

Physical measurement techniques have advantages over advanced imaging techniques with respect to feasibility; these techniques are simple to perform, cost effective and acceptable to patients. However, physical measurement does not allow for storage of images, or later cross checking of data. And physical measurement allows for measurement of only superficial subcutaneous tophi and not data on the

size of intraarticular tophi. For measurement of intraarticular tophi, only US has documented reliability and sensitivity to change. However, this method is operator dependent, requires relatively expensive equipment, and may not be feasible for all clinical trials in chronic gout.

These data and issues were presented during the gout workshop and were discussed in a tophus measurement breakout group. A number of additional points were raised in the breakout group: The relevance of tophus regression as an outcome measure was considered, both with respect to clinical relevance and as a “surrogate” for the total urate pool. Selection of tophi for assessment may be important, and it is currently unknown whether measurement of a single sentinel tophus is sufficient, or whether a predetermined higher number of tophi should be measured. The degree of change in size may depend on size at baseline and site of deposition, raising the question of whether all tophi respond in the same way to urate lowering therapy. Reporting of change in tophus size is not currently standardized; a number of options have been used, including percentage change from baseline, time to resolution of tophus, or velocity of change over time. The importance of US for assessment of intraarticular tophi was noted by the breakout group, although standardization of an US scoring system for gout is needed. US assessment of intraarticular tophi in combination with physical measurement of subcutaneous tophi may contribute to a definition of remission in patients with gout.

Other methods of advanced imaging of tophi were also discussed, including the potential role of new techniques such as 3-dimensional US and dual energy CT. The breakout group also discussed whether aspiration of the tophus should be undertaken prior to selection of a sentinel tophus for ongoing monitoring; this was not considered necessary, and it was noted that such an approach would restrict the feasibility and patient acceptability of tophus measurement in clinical trials.

These uncertainties were reflected in voting at the conclusion of the workshop. When participants were asked which of the tophus measurement methods passed the OMERACT filter for use in chronic gout trials, no instrument was endorsed (tape measure 39%, calipers 52%, US 49%, MRI 3%, and CT 10%). In the discussion following voting participants indicated that there were insufficient data available at present.

The data presented in the gout workshop and the additional issues discussed in the tophus measurement breakout group form the basis of an ongoing research agenda to address the following questions:

- a. How do the number and size of tophi relate to the total urate pool?
- b. Do all tophi respond to urate lowering therapy in a similar way? For measurement of tophus size in clinical trials, which site and what number of tophi should be selected? Is

- a “single joint” approach reliable?
- c. What time period is needed for reliable assessment of change in tophus size?
- d. How should change in tophus size be reported?
- e. What is an appropriate definition of remission for use in clinical trials in gout?
- f. What is the role of advanced imaging for tophus assessment in clinical trials, particularly through standardization of an US scoring system?

Health-related quality of life. Although health-related quality of life (HRQOL) is a proposed mandatory domain in studies of chronic gout, it has not been reported in published interventional studies to date⁵. The performance of the widely used Medical Outcomes Study Short Form-36 Health Survey (SF-36) and Disability index of the HAQ (HAQ-DI), have recently been assessed in chronic gout. Both these instruments are generally accepted to pass the OMERACT filter criteria²⁷. During the OMERACT 9 workshop Waltrip, *et al* presented data from a 12-month, prospective, observational study of urate-lowering therapy in individuals with chronic gout²⁸. This demonstrated the SF-36 is reliable (stable in subjects with no flares), sensitive to change (in subjects with frequent flares), and discriminates between groups on the basis of disease severity, physician’s global assessment, and between groups with higher versus lower flare frequency. The HAQ-DI also performed well in this study. Although there are some concerns with floor and ceiling effects of this instrument, this concern applies across studies of many rheumatic diseases for which this instrument is used²⁹. In an observational study of over 200 patients with chronic gout the HAQ-DI has been shown to show construct validity (strong correlation with SF-36 and clinical severity), test-retest reliability, sensitivity to change (in individuals with a change in pain score), and between group discrimination (based on clinical severity score)³⁰.

A novel instrument for assessment of gout in clinical trials, the Gout Assessment Questionnaire (GAQ1.0), has been developed³¹. After patient interviews and assessment by rheumatologists and experts in patient reported outcomes, this instrument has been revised to include a section to describe the impact of gout on HRQOL (GAQv2.0 - Gout Impact)³². During the workshop a validation study in over 300 gout patients in 3 North American centers was presented, showing the instrument is feasible, has face and content validity, is reliable, and discriminates between patients with severe and mild gout and on the basis of flare frequency. A longitudinal study is under way to address issues of responsiveness to change and minimal important differences.

During the gout workshop, 2 breakout groups discussed these data and the use of these instruments. It was generally agreed that HRQOL was an important domain, but concerns remain regarding use of generic instruments. In particular, it is uncertain whether generic HRQOL instruments can discriminate between impact on HRQOL from gout versus

common comorbidities in subjects with chronic gout. This is a valid concern as a mailout survey has shown individuals with gout have poorer HRQOL when measured with the SF-36 but this was largely attributable to comorbidities and sociodemographic characteristics³³. However, in a similar community-based study using the WHO-QOL Bref, a reduction in physical health-related quality of life domain remained impaired after adjustment for comorbidities³⁴. In longitudinal gout treatment studies of medium duration it is reasonable to assume that comorbidities will be relatively stable in individuals over time, thus any improvement in HRQOL is likely to be attributable to treatment. Much interest was expressed in the GAQ v2.0-Gout Impact scale; however, further data assessing the instruments in other populations and over time are required. These studies are underway.

At plenary voting, participants endorsed the SF-36 (77%) as a measure of HRQOL that passes the OMERACT filter for use in clinical trials, but not the HAQ-DI (58%), or the GAQv2.0-Gout Impact (30%).

Activity limitations. The HAQ-DI is considered to measure functional disability. As already discussed this instrument shows construct and internal validity, but other technical properties are yet to be assessed. It is also not known if the HAQ-DI encompasses all the ways in which chronic gout may impact an individual. Jasvinder Singh presented unpublished data during the gout workshop assessing the performance of the Katz Index of Independence in activities of daily living (Katz 6 ADL)³⁵ in a mail out survey to veterans with gout. Although it passed truth and feasibility filters, the discriminative ability remains unclear and is the focus of future work. At voting the Katz 6 ADL was not supported to pass the OMERACT filter for use (21%). Development of an ICF Core set for chronic gout may be useful to ensure any established or new instruments comprehensively cover all ways in which chronic gout can impact function and disability.

Pain. It is likely pain can be measured with instruments similar to those proposed for acute gout, but these remain to be validated in chronic gout studies. In this context pain may relate to subcutaneous tophi or joint damage from intraarticular tophi and from acute flares. A number of issues will require consideration during assessment of such tools:

- a. Whether pain is assessed in a single index location or is reported as all pain related to gout. It will be important to consider the face validity of this measurement as it is unknown if individuals with gout are able to accurately discriminate pain due to gout from that due to other common causes of musculoskeletal pain over extended periods of time.
- b. Whether reported pain predominantly occurs in the context of flares; thus pain may be a redundant measure if flare frequency is measured.
- c. When there is likely to be considerable baseline variation in pain reported, depending on duration since last flare and burden of tophaceous deposits, whether pain assessment will show sufficient sensitivity or have floor effects.

Patient global. The instrument for these domains can be a 5-point Likert scale or 100 mm visual analog scale. These instruments have not previously been used as endpoints in randomized controlled trials in chronic gout⁵. Assessment in future studies is required. Given that the outcomes of concern to patients probably include frequency of flares, pain, and tophus regression, it will be important to assess if a global assessment of response to therapy adds meaningful information and if both measures are required.

Work disability. No instrument has been identified for measurement of work disability in gout. The concept of work disability includes work loss and work limitations. Conceptually this equates to the loss of worker productivity, which is a measure of absenteeism (time away from work) and presenteeism (person is still at work but not performing to full capacity/expectations)³⁶. Measurement of worker productivity was the focus of a workshop at OMERACT 9 and considerable progress has been made towards defining instruments of measurement and development of a metric, which may incorporate absenteeism and presenteeism. It is anticipated that the instrument developed may be applicable across a variety of musculoskeletal conditions, although assessment in gout will be required.

Joint inflammation. Although consensus exercises have found a high median rating for the potential domain of joint inflammation, the components of inflammation to be measured have not been defined. An instrument is likely to be an investigator joint count, including number of tender or swollen joints, with tenderness or swelling being a dichotomous measure for each joint. Reporting of joint swelling in gout may be problematic, as both tophaceous joint disease and synovitis may cause joint swelling. The joint count would need to include the feet, as gout predominantly affects the lower limb. The 66/68 joint count or 44 joint count (Ritchie Index) used in rheumatoid arthritis may be appropriate, but these will need to be assessed in the arthritis of chronic gout. Since these instruments may be unfeasible for repeated measurement in large studies of months to years in duration, the technical properties of a patient reported joint count may also warrant investigation. It is also possible that joint inflammation could be measured with other

techniques, e.g., ultrasound, which may be better suited to tracking inflammation in the context of monoarticular or oligoarticular involvement.

Other outcome measures. Radiographic damage index. A radiographic damage index has recently been validated for use in studies of chronic gout³⁷. This is a modified Sharp/van der Heijde scoring method, incorporating the hand distal interphalangeal joints. This system accurately represents joint damage in gout, is reproducible and reliable, is able to discriminate between early and late disease, and is feasible. Radiographic damage using this scoring system is a strong predictor of hand function in patients with gout³⁸. The sensitivity to change of this scoring system has not yet been reported, either over time or in response to therapy. Analysis of existing paired sets of radiographs, longitudinal observational study data, and clinical trial data is planned to further address this issue.

Response criteria for acute and chronic gout. In studies measuring multiple outcomes it may be useful to combine measures into a single composite metric, which then becomes a dichotomous outcome. A dichotomous responder index has the benefits of simple interpretation and conversion into “number needed to treat.” Establishment of a responder index will require: a. Consensus exercises to decide if a composite index of response adds value or meaning or whether a single measure is sufficiently comprehensive. b. Empiric data analysis to determine redundancy. c. Testing of proposed definition of response in clinical trials of interventions of effective treatments compared to less effective treatments so ability to discriminate between effective and less effective treatments.

Conclusion. The last 5 years have seen considerable progress in establishing domains for measurement in gout studies. As domains have been clarified, the focus of research has shifted to defining the technical properties of potential instruments for measuring these domains. The instruments that can currently be recommended and have passed the OMERACT filter are shown in Table 3. Efforts now move to identifying or further validating instruments in other domains.

Table 3. Instruments for measurement in gout studies considered to have passed the OMERACT filter (not necessarily endorsed by plenary voting at OMERACT 9).

Domain	Acute Gout	Plenary Voting, %	Chronic Gout	Plenary Voting, %
Pain	5-point Likert scale	96		
Patient global	PGART	83		
HRQOL			SF-36	77
			HAQ-DI	58
Tophus burden			Calipers	52
			US	49

PGART: Patient global assessment of response to treatment; HRQOL: health related quality of life; SF-36: Medical Outcomes Study Short Form 36 Survey; HAQ-DI: Health Assessment Questionnaire Disability Index; US: ultrasound.

REFERENCES

- Schumacher HR, Jr., Edwards LN, Perez-Ruiz F, et al. Outcome measures for acute and chronic gout. *J Rheumatol* 2005;32:2452-5.
- Schumacher HR, Taylor W, Joseph-Ridge N, et al. Outcome evaluations in gout. *J Rheumatol* 2007;34:1381-5.
- Taylor WJ, Schumacher HR, Jr., Baraf HS, 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.
- Schumacher HR, Taylor W, Edwards L, et al. Outcome domains for studies of acute and chronic gout. *J Rheumatol* 2009;36:2342-5.
- Taylor WJ, Schumacher HR, Jr., Singh JA, Grainger R, Dalbeth N. Assessment of outcome in clinical trials of gout — a review of current measures. *Rheumatol* 2007;46:1751-6.
- Man CY, Cheung IT, Cameron PA, Rainer TH. Comparison of oral prednisolone/paracetamol and oral indomethacin/paracetamol combination therapy in the treatment of acute gout-like arthritis: a double-blind, randomized, controlled trial. *Ann Emerg Med* 2007;49:670-7.
- Willburger RE, Mysler E, Derbot J, et al. Lumiracoxib 400 mg once daily is comparable to indomethacin 50 mg three times daily for the treatment of acute flares of gout. *Rheumatol* 2007;46:1126-32.
- Janssens HJ, Janssen M, van de Lisdonk EH, van Riel PL, van Weel C. Use of oral prednisolone or naproxen for the treatment of gout arthritis: a double-blind, randomised equivalence trial. *Lancet* 2008;371:1854-60.
- Schumacher HR, Jr, Boice JA, Daikh DI, et al. Randomised double blind trial of etoricoxib and indomethacin in treatment of acute gouty arthritis. *BMJ* 2002;324:1488-92.
- Rubin BR, Burton R, Navarra S, et al. Efficacy and safety profile of treatment with etoricoxib 120 mg once daily compared with indomethacin 50 mg three times daily in acute gout: a randomized controlled trial. *Arthritis Rheum* 2004;50:598-606.
- Schlesinger N, Norquist JM, Holmes R, Boice J, Watson DJ. Validation of a patient-reported assessment of pain in acute gouty arthritis [abstract]. *Ann Rheum Dis* 2007;66:234.
- Bolognese JA, Schnitzer TJ, Ehrich EW. Response relationship of VAS and Likert scales in osteoarthritis efficacy measurement. *Osteoarthritis Cartilage* 2003;11:499-507.
- Norquist JM, Levine K, Schlesinger N, Boice JA, Watson DJ. Critical evaluation of recommended outcome assessments in acute gouty arthritis [abstract]. *Ann Rheum Dis* 2008;67:247.
- Navarra S, Rubin BR, Yu Q, Smugar SS, Tershakovec AM. Association of baseline disease and patient characteristics with response to etoricoxib and indomethacin for acute gout. *Curr Med Res Opin* 2007;23:1685-91.
- Grill E, Zochling J, Stucki G, et al. International Classification of Functioning, Disability and Health (ICF) Core Set for patients with acute arthritis. *Clin Exp Rheumatol* 2007;25:252-8.
- 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.
- 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.
- Shoji A, Yamanaka H, Kamatani N. A retrospective study of the relationship between serum urate level and recurrent attacks of gouty arthritis: evidence for reduction of recurrent gouty arthritis with antihyperuricemic therapy. *Arthritis Rheum* 2004;51:321-5.
- Becker MA, Schumacher HR, Jr, Wortmann RL, MacDonald PA, Eustace D, Palo WA, et al. Febuxostat compared with allopurinol in patients with hyperuricemia and gout. *N Engl J Med* 2005;353:2450-61.
- Perez-Ruiz F, Calabozo M, Pijoan JI, Herrero-Beites AM, Ruibal A. Effect of urate-lowering therapy on the velocity of size reduction of tophi in chronic gout. *Arthritis Rheum* 2002;47:356-60.
- Perez-Ruiz F, Martin I, Canteli B. Ultrasonographic measurement of tophi as an outcome measure for chronic gout. *J Rheumatol* 2007;34:1888-93.
- Taylor WJ, Shewchuk R, Saag K, Schumacher Jr HR, Singh JA, Grainger R, et al. Towards a valid definition of gout flare - results of consensus exercises using Delphi methodology and cognitive mapping. *Arthritis Care Res* 2009;61:535-43.
- Schumacher HR, Jr, Becker MA, Palo WA, Streit J, MacDonald PA, Joseph-Ridge N. Tophaceous gout: quantitative evaluation by direct physical measurement. *J Rheumatol* 2005;32:2368-72.
- Wortmann RL, Becker MA, Schumacher HR, MacDonald PA, Hunt BJ, Joseph-Ridge N. Effect of febuxostat or allopurinol on the clinical manifestations of gout: reduction in gout flares and tophus size over time in the EXCEL trial [abstract]. *Arthritis Rheum* 2006;54:S642.
- Dalbeth N, Clark B, Gregory K, Gamble GD, Doyle A, McQueen FM. Computed tomography measurement of tophus volume: comparison with physical measurement. *Arthritis Rheum* 2007;57:461-5.
- Schumacher HR, Jr, Becker MA, Edwards NL, et al. Magnetic resonance imaging in the quantitative assessment of gouty tophi. *Int J Clin Pract* 2006;60:408-14.
- Strand V. Design of and outcome measures for use in clinical trials in patients with osteoarthritis. In: Moskowitz RW, Buckwalter JA, Goldberg VM, Hochberg MC, editors. *Osteoarthritis: diagnosis and medical/surgical management*. Philadelphia: Lippincott Williams and Wilkins; 2006:313-25.
- Becker MA, Schumacher HR, Benjamin KL, Gorevic P, Greenwald M, Fessel J, et al. Quality of life and disability in patients with treatment-failure gout. *J Rheumatol* 2009;36:1041-8.
- Ruta DA, Hurst NP, Kind P, Hunter M, Stubbings A. Measuring health status in British patients with rheumatoid arthritis: reliability, validity and responsiveness of the short form 36-item health survey (SF-36). *Br J Rheumatol* 1998;37:425-36.
- Alvarez-Hernandez E, Pelaez-Ballesteros I, Vazquez-Mellado J, et al. Validation of the Health Assessment Questionnaire disability index in patients with gout. *Arthritis Rheum* 2008;59:665-9.
- Colwell HH, Hunt BJ, Pasta DJ, Palo WA, Mathias SD, Joseph-Ridge N. Gout Assessment Questionnaire: Initial results of reliability, validity and responsiveness. *Int J Clin Pract* 2006;60:1210-7.
- Hirsch JD, Lee SJ, Terkeltaub R, Kanna D, Singh J, Sarkin A, et al. Evaluation of an instrument assessing influence of Gout on health-related quality of life. *J Rheumatol* 2008;35:2406-14.
- Singh JA, Strand V. Gout is associated with more comorbidities, poorer health related quality of life and higher health care utilization in US veterans. *Ann Rheum Dis* 2008;67:1310-6.
- Roddy E, Zhang W, Doherty M. Is gout associated with reduced quality of life? A case-control study. *Rheumatology* 2007;46:1441-4.
- Katz S, Down TD, Cash HR, Grotz RC. Progress in the development of the index of ADL. *Gerontologist* 1970;10:20-30.
- Escorpizo R, Bombardier C, Boonen A, et al. Worker productivity outcome measures in arthritis. *J Rheumatol* 2007;34:1372-80.
- Dalbeth N, Clark B, McQueen F, Doyle A, Taylor W. Validation of a radiographic damage index in chronic gout. *Arthritis Rheum* 2007;57:1067-73.
- Dalbeth N, Collis J, Gregory K, Clark B, Robinson E, McQueen FM. Tophaceous joint disease strongly predicts hand function in patients with gout. *Rheumatology* 2007;46:1804-7.
- Becker MA, Schumacher HR, Jr, Wortmann RL, et al. Febuxostat, a novel nonpurine selective inhibitor of xanthine oxidase: a twenty-eight-day, multicenter, phase II, randomized, double-blind, placebo-controlled, dose-response clinical trial examining safety and efficacy in patients with gout. *Arthritis Rheum* 2005;52:916-23.