BMJ Open

Using serum urate as a validated surrogate endpoint for flares in patients with gout: protocol for a systematic review and meta-regression analysis

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-012026
Article Type:	Protocol
Date Submitted by the Author:	23-Mar-2016
Complete List of Authors:	Morillon, Melanie; Frederiksberg and Bispebjerg Hospital, Parker institute; Department of Rheumatology, Odense University Hospital, Stamp, Lisa; University of Otago, Christchurch, Department of Medicine, Taylor, William; University of Otago, Rehabilitation Teaching & Research Unit, Department of Medicine Fransen, Jaap; Radboud Uviversity Medical Centre, Rheumatology Dalbeth, Nicola; University of Auckland, Department of Immunology Singh, Jasvinder; University of Alabama at Birmingham & Birmingham Veterans Affairs, Department of Medicine Christensen, Robin; Frederiksberg and Bispebjerg Hospital, Parker Institute Lassere, Marissa; St George Hospital, Rheumatology
Primary Subject Heading:	Rheumatology
Secondary Subject Heading:	Rheumatology
Keywords:	GOUT, BIOMARKER, RHEUMATOLOGY, SURROGATE, OUTCOME

Using serum urate as a validated surrogate endpoint for flares in patients with gout: protocol for a systematic review and meta-regression analysis

Collaborators/Authors

[1,2]Melanie B. Morillon MD; [3]Lisa Stamp PhD, FRACP; [4]William Taylor PhD, FRACP; [5]Jaap Fransen PhD; [6]Nicola Dalbeth MD, FRACP; [7]Jasvinder A. Singh, MD, MPH; [1]Robin Christensen MSc, PhD; [8]Marissa Lassere MBBS (Hons) Grad Dip Epi, PhD, FRACP, FAFPHM;

Affiliations

[1]Musculoskeletal Statistics Unit, The Parker Institute, Department of Rheumatology, Copenhagen University Hospitals, Bispebjerg and Frederiksberg, Denmark

[2]Department of Rheumatology, Odense University Hospital, Denmark

[3]Department of Medicine, University of Otago, Christchurch, New Zealand

[4]Department of Medicine, University of Otago, Wellington, New Zealand

[5]JF Department of Rheumatology, Radboud University Medical Centre, Nijmegen, The Netherlands.

[6]Department of Medicine, University of Auckland, New Zealand.

[7]Department of Medicine, University of Alabama at Birmingham & Birmingham Veterans Affairs Medical Center, Birmingham, Alabama, USA.

[8]Department of Rheumatology, St George Hospital, University of NSW, Sydney, Australia

Correspondence to:

Melanie Birger Morillon
Department of Rheumatology,
Odense University Hospital,
5300 Svendborg,
Danmark, DK
Email: melanie.birger.morillon@rsyd.dk

ABSTRACT

Introduction: Gout is the most common inflammatory arthritis in men over 40 years. Long term urate-lowering therapy is considered a key strategy for effective gout management. The primary outcome measure for efficacy in clinical trials of urate lowering therapy is serum urate levels, effectively acting as a surrogate for patient-centred outcomes such as frequency of gout attacks or pain, yet it is not clearly demonstrated that the strength of the relationship between serum urate and clinically relevant outcomes is sufficiently strong for serum urate to be considered an adequate surrogate. Our objective is to investigate the strength of the relationship between changes in serum urate in randomised controlled trials and changes in clinically relevant outcomes according to the 'Biomarker-Surrogacy Evaluation Schema version 3' (BSES-3), documenting the validity of selected instruments by applying the 'OMERACT Filter 2.0'.

Methods and analysis: A systematic review described in terms of the PRISMA reporting guidelines will identify all relevant studies. Standardised data elements will be extracted from each study by 2 independent reviewers and disagreements resolved by discussion. The data will be analysed by meta-regression of the between-arm differences in the change in serum urate level (independent variable) from baseline to 3 months (or 6 and 12 months if 3 month values not available) against flare rate, tophus size and number and pain at the final study visit (dependent variables).

Ethics and dissemination: This study will not require specific ethics approval since it is based on analysis of published (aggregated) data. The intended audience will include health care researchers, policymakers and

clinicians. Results of the study will be disseminated by peer-review publication.

Protocol registration: In Prospero, CRD42016026991. Accepted January 2016.

STRENGHTS AND LIMITATIONS

- Our study's strengths include clinical expertise in rheumatology
- The content experts in the group have extensive knowledge of the literature and ▫experience with gout treatment
- The methodologists in the group are members of the OMERACT Gout Working Group, and have experience with conducting and reporting randomised clinical trials, systematic reviews and meta-analyses.
- A possible and anticipated weakness may be the quantity and quality of the trials we identify

INTRODUCTION

Clinicians making treatment decisions should refer to methodologically strong clinical trials examining the impact of therapy on clinically important outcomes (i.e. outcomes that are important to patients). However, clinically important outcomes can be difficult to study, as the required trials need very large sample sizes, or long-term patient follow-up. Thus researchers or drug developers look for alternatives. Substituting surrogate end points for the target event allows conduct of shorter and smaller trials, thus offering a solution to the dilemma, if the endpoints are convincing as surrogate endpoints.

There are obvious advantages to using biomarkers and surrogate endpoints, but concerns about clinical applicability and statistical validity to evaluate these aspects hinder their efficient application. A surrogate end point may be defined as an "objective" laboratory measurement or a physical sign used as a substitute for a clinically meaningful end point that measures directly how a patient feels, functions or survives (1). This definition was recommended and further explored at a National Institute of Health (NIH) sponsored workshop in 1998 which agreed on definitions for biomarker, surrogate endpoint and clinical endpoint. The agreed definition of a biomarker states "*a biological marker (biomarker) is a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes or pharmacologic responses to a therapeutic intervention*" (2) (3).

In gout, monosodium crystal formation occurs when super-saturation levels are reached ~6.8mg/dL (0.41mmol/l) at 37^0C. Reduction in serum

urate (SU) to <6mg/dL (0.36mmol/L) is a key goal in the long-term management of gout. As such SU measurement has become an integral part of the management of gout and a critical outcome measure in clinical studies of gout therapies. The OMERACT Delphi exercise identified SU as a mandatory outcome measure in chronic gout studies with the highest median rating (4). SU as a biomarker makes inherent sense given the strong relationship between the risk of gout and SU. However, is SU a surrogate endpoint of relevant clinical outcomes such as gout attacks, tophus regression and radiological damage?

Background

At OMERACT (Outcome Measures in Rheumatology Clinical Trials) 8 (Malta, 2006) Lassere et al proposed a schema for evaluation of biomarkers as surrogate endpoints (5). The schema was operationalized as a score obtained from four domains: target outcome, study design, statistical strength and penalties (5). This schema was based on the NIH definitions of biomarker, surrogate endpoint and clinical endpoint published in 2001 (2). The distinction between a surrogate and a biomarker was determined by the strength of association between the biomarker and the clinical endpoint of interest. To be called a surrogate, it was proposed that a biomarker must meet the rank (score) of at least 3 within the Target Outcome, Study Design, and Statistical Strength domains, and there must not be evidence from a Randomised Controlled Trial (RCT) that use of the biomarker caused patient harm (5).

At OMERACT 9 (Kananaskis, 2008) the soluble biomarker group revised the requirements for the specific situation of a soluble biomarker being predictive of structural radiographic damage in ankylosing spondylitis, psoriatic arthritis and rheumatoid arthritis (6). There was an increased emphasis on the technical assay requirements of the biomarker but the strength of association domain, while discussed in the text, did not appear in the OMERACT 9 levels of evidence framework. There was no consensus on all aspects of the framework, and the criteria by which a soluble biomarker could be said to meet the levels of evidence framework were not defined.

At OMERACT 10 (Kota Kinabalu, 2010) evidence was presented that SU fulfilled the OMERACT 9 soluble biomarker requirements in terms of domain 4 (performance criteria) and limited evidence from observational studies and one RCT that changes in SU were associated with changes in patient-centred outcomes for the disease of gout (7). However, the meeting did not endorse SU as a biomarker for clinically relevant outcomes for gout. The reasons for the lack of endorsement might be that the strength of evidence was weak, the criteria for endorsement are unclear and the chosen patient-centred outcomes (particularly number of flares) were not universally held to be clinically meaningful.

In parallel to OMERACT, Lassere et al systematically reviewed the biomarker-surrogate literature and modified the levels of evidence schema built on the OMERACT 8 proposal (5) which over time went through 3 iterations ('Biomarker-Surrogacy Evaluation Schema version' (BSES), BSES1 which was the OM 8 proposal (5), BSES2 which specified the statistical criteria more precisely (8) and BSES3 which replaced the penalties domain

with a combined clinical and pharmacological generalizability domain). BSES3 contains 4 domains: study design, target outcome, statistical evaluation, and generalizability. It also specified the kind of statistical association required to justify the link between biomarker and clinical endpoint being sufficiently strong to consider the biomarker as a surrogate endpoint (9) (10).

In 2012 blood pressure was evaluated using the BSES3 and online material described its application and interpretation (10). The BSES3 framework represents the currently best available approach to validating a biomarker as a surrogate endpoint. We propose that this framework be endorsed by OMERACT as the framework for validation of biomarker-surrogates for rheumatology clinical trials. It represents the logical extension of work developed at OMERACT 8 and provides a clear pathway by which a putative biomarker, soluble or otherwise, can be evaluated, in contrast to the OMERACT 9 framework. For example, Lassere et al has used trial-level data and the BSES3 framework to convincingly show that diastolic and systolic blood pressure are valid surrogate endpoints for stroke risk reduction (10). In a recent meta-regression the approach has also been used to evaluate progression-free survival (PFS) in metastatic renal cell carcinoma (11).

Rationale

We wish to use the example of SU as a soluble biomarker for the major clinical endpoint of acute gout attacks, in the disease of gout. A minor clinical endpoint would be tophus size change from baseline to final visit,

the change in the number of tophi, and pain. Other patient relevant endpoints included in the OMERACT core-set of outcomes for clinical trials in patients with chronic gout will also be evaluated in exploratory analyses: health related quality of life (HRQOL), patient global assessment of disease activity, and physical disability (activities limitation).

The justification for choosing this biomarker and the clinical endpoint of flares as the major endpoint is described as follows:

- Firstly, SU is recommended as a treatment target by several guidelines for the management of gout (12-14). This strongly implies (although it is not stated explicitly) that changes in SU or achievement of a target level of SU will be strongly associated with clinically relevant outcomes.
- Secondly, some regulatory bodies (e.g. Food and Drug Administration and European Medicines Agency) have tended to assume that beneficial drug effects on SU will likely have beneficial effects on clinical outcomes in gout. NICE recommended that febuxostat be available for people who are intolerant of allopurinol or who have contraindications to allopurinol (15). In other words, although NICE did not see persuasive evidence for improved clinical outcomes with the use of febuxostat, it was sufficient that the drug effectively lowered SU to below 6 mg/dL.
- Thirdly, we have previously shown that SU fulfils the technical performance criteria for a valid soluble biomarker proposed at OMERACT 9 (7). Flare (acute attack) of gout is a key clinical

manifestation of gout. It constitutes the primary or only manifestation for several years until persistent, tophaceous disease develops. In the expectation that effective management strategies aim to prevent chronic tophaceous disease from occurring it is justifiable to focus on attacks as the clinically relevant endpoint for the majority of people for gout. Although gout attacks can vary in severity (often modified by acute gout treatment), it is clear that every attack is associated with some level of symptoms and disability. Gout attacks therefore align with how a patient 'feels or functions' and can be reasonably be identified as a clinically relevant endpoint (1).

However, we recognize that other clinical outcomes are relevant and will evaluate these within the same framework. This proposal fits in the Filter 2.0 framework by making explicit and quantifying the link between Core Area domains of Pathophysiology Manifestations (biomarker) and domains of Life Impact (flare, pain, HRQOL, tophus). This framework links disease-centred variables of biological and pathological processes with patient-centred variables of how a patient feels, functions and survives as proposed at OMERACT 6 (5).

Objectives

There are two objectives:

1) To determine the strength of the relationship between SU and patient relevant outcomes, including flares, tophi, HRQOL, pain and function using meta-regression of randomised controlled trials.
2) To evaluate whether SU is a surrogate endpoint for clinically relevant outcomes in patients with gout as defined by the BSES3 framework.

Hypothesis

A reduction in SU will be associated with improvement in clinically relevant patient reported outcomes including gout flares and tophus size/number.

METHODS AND ANALYSIS

Protocol and registration

The protocol for the systematic review and meta-regression analysis was prepared while planning and documenting the review methods, guarding the project team against arbitrary decision making during review conduct, and to prompt global collaboration (16). Our protocol was prepared according to the recommendations given in PRISMA-P (16) and registered on PROSPERO (CRD42016026991); this protocol and coming manuscripts will conform to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for reporting systematic reviews and meta-analyses (17).

Eligibility criteria

The eligibility criteria for objective 1 is any randomised controlled trial comparing an active drug (alone or in combination) in patients with gout

with any control or placebo, with a minimum duration of three months. The eligibility criteria for objective 2 are any randomised controlled trial, controlled clinical trial, or open label trial (OLT) comparing an (apparently) active drug (alone or in combination) in patients with gout with any control or placebo, with a minimum duration of three months and longitudinal observational studies of gout with a minimum duration of 3 months.

For both criteria, patients will be at least 18 years of age and meeting the preliminary American College of Rheumatology (ACR) criteria for acute arthritis of primary gout (18) or given a diagnosis of gout as described by the authors.

Search and selection of trials

The following electronic databases will be searched: PubMed, EMBASE, the Cochrane Library including the Cochrane Central Register of Controlled Trials (CENTRAL) and Cochrane Database of Systematic Reviews (CDSR). The search will be limited to English language studies in humans, but not limited by year of publication. The reference lists from comprehensive reviews and identified clinical trials are also manually searched.

Results of the various searches will be reviewed independently by two authors (LS and MM). Titles and abstracts will be reviewed and if further information is required (to assess eligibility criteria), the full text will be obtained. A record of reasons for excluding studies will be kept enabling generation of a figure illustrating the flow of information through the different phases of the systematic review continuing to meta-regression

analysis. Disagreements will be resolved by an independent third mediator (WT).

Data extraction

EndNote X7 software will be used to manage the records retrieved from searches of electronic databases. Results from hand searches will be tracked on a Microsoft Excel spreadsheet. A customised data extraction form will be created in Microsoft Excel to capture all the information available for each individual trial.

The biomarker is defined as the change in SU from baseline to 3 months, or where 3 month values are not available, the value at 6 months or 12 months (in order of preference). This can be estimated if only baseline and change is reported.

The clinical endpoints (dependent variables) are defined as follows:

- Major outcome: gout-flares
- Minor outcomes: size of sentinel tophus (if size was not measured, we will use number, or presence/absence in order of preference) and pain at final study visit'

Exploratory analyses: health related quality of life (HRQOL) (SF36), patient global assessment of disease activity, and physical disability (activities limitation; e.g., HAQ).

Effect sizes for continuous endpoints will be recorded as the standardised mean difference. If there is more than one active treatment arm, analysis will treat this as a separate study i.e. sub-study (see Meta-regression

Analysis). All variable values will be based upon the intention-to-treat population from each study whenever possible.

Risk of bias in individual studies and judging the quality of evidence

The RCTs will be assessed for methodological quality (i.e. internal validity) using the Cochrane Risk of Bias tool (19). If at least one of the domains is rated as inadequate, the trial will be considered at high risk of bias. If all domains are judged as low, the trial will be considered at low risk of bias. Otherwise, the trial is considered as having unclear risk of bias. Data extraction and risk-of-bias assessment will be performed independently by 2 reviewers; disagreements will be resolved by a third reviewer. While interpreting the overall findings after the meta-analysis etc., GRADE (Grading of Recommendations Assessment, Development and Evaluation) will be used to rate the overall quality of the evidence based on both the apparent risk of bias, publication bias, imprecision, inconsistency, indirectness, and magnitude of effect; i.e., the GRADE ratings of very low, low, moderate, or high-quality evidence per outcome will reflect the extent to which we are confident that the effect estimates are correct (20)

Meta-regression analysis

To combine the individual study results, we will perform meta-analyses using SAS software (PROC MIXED version 9. 3; SAS Institute Inc., Cary, NC, USA), applying a restricted maximum likelihood (REML) method to estimate the between-study variance (i.e. T^2) and the combined estimate of effect. We will estimate the anticipated heterogeneity between trials with a

standard (Cochran's) Q-test statistic, and we will evaluate this based on the I^2 value, which is interpreted as the percentage of variability in treatment effect estimates that is due to between study heterogeneity rather than chance.

The primary purpose of this project is to evaluate the surrogacy status of SU as a "predictor" of gout flare rate reduction using meta-regression of randomised controlled trials. Randomisation is essential for the causal surrogacy relationship, therefore, only randomised controlled trials will be included in the main meta-regression analysis. Non-randomised study designs will be summarised separately by meta-regression to confirm the consistency of association between the biomarker and clinical endpoints in other contexts. Cohort studies will be summarised as a narrative review. The analyses of both randomised and non-randomised studies contribute to the evaluation of serum urate within the BSES3 framework

Furthermore, in the meta-regression, the relationship between serum urate and clinically relevant outcomes can be undertaken using different outcome metrics. We will define these as primary and secondary analyses. In the primary analysis the dependent variable is a rate ratio (i.e. an incidence density ratio) comparing the ratio of incidence rates of gout flare events in active versus control arms occurring at any given point in time; incidence rate is the occurrence of an event over person-time (i.e., in this setting in person-months). The rate ratio allows trials of different duration to be included in the analysis. The independent variable is between arm difference of within-arm change (on-trial SU from baseline SU) of SU. Therefore, in a trial of 3 months duration, flare rate over 3 months is the

dependent variable and change in SU over 3 months is the independent variable.

In secondary analyses the dependent variable is risk ratio reduction (RRR) of within trial gout flare rate. The relative ratio reduction (also called the risk ratio reduction) is the flare risk in the control arm minus the flare risk in the active arm, divided by the flare risk in the control arm (this can also be calculated by 1- Relative Risk (RR), where relative risk is the flare risk in the active arm divided by the flare risk in the control arm). Therefore the relative risk reduction (RRR) is the difference in flare risk in two arms (control-active), expressed as a percentage of the risk of the control arm.

The independent variable is within trial, by-arm difference of proportion with SU less than 6mg/dL at the end of the trial.

In a randomised controlled trial, by-arm difference in SU change is likely to be causal and change in SU is easily interpretable as a surrogacy metric in gout by clinicians. Relative risk reduction is more familiar to clinicians than rate ratio but ignores trial duration. Although SU less than 6mg/dL is the most common primary endpoint of RCTs of gout interventions, a by-arm difference in proportion achieving a SU target may be more difficult to interpret than a serum urate change. In addition to gout flares, the SU as a surrogate endpoint for two other clinical outcomes, HRQoL and tophus size, will also be evaluated as secondary clinical outcomes. If the trial does not report these outcomes, the authors will be contacted and the by-arm outcomes requested.

A quantitative evaluation of trial-level statistical surrogacy using the BSES3 (10) includes determining the slope coefficient of the surrogacy

relationship, trial-level R^2 (coefficient of determination) (21)and the Surrogate Threshold Effect (STE) (22, 23, 24) and Surrogate Threshold Effect Proportion (STEP) (8, 10) of the surrogate and true-clinical-endpoint relationship using data from a meta-regression of randomized controlled trials.

The STE is informative as it captures both the slope and dispersion of the surrogate-true relationship in a single metric (24). The STE is the serum urate difference needed to predict the primary clinical endpoint, gout flare rate ratio, in a new trial, if only serum urate is measured in the new trial. The STE is determined by comparing the difference between control and active arms SU and flare rate respectively as follows: (i) calculate the SU change and gout flare rate ratio based on each arm in each trial, (ii) calculate the difference between control and active arms for SU change and gout flare rate ratio, (iii) regress SU and gout flare rate ratio difference values using weighted by trial size errors-in-variables (specifying a reliability coefficient of 0.9) regression and by a weighted by trial-size meta-regression (as a sensitivity analysis), (iv) calculate the 95% prediction limits of the regression, and (v) find the SU value where the 95% prediction line intersects with the horizontal flare rate x-axis of no flare rate ratio benefit (where the flare rate ratio y-axis is equal to 1.0). Similar analyses will be explored with flare rate relative risk reduction and proportion with SU less than 6mg/dL at the end of the trial. In this analysis the interest is the SU target <6mg/dL by-arm proportion where the 95% prediction line intersects with the horizontal flare rate x-axis of no flare relative risk reduction benefit (i.e. where the flare relative risk reduction y-axis is equal to zero).

Subsequent analyses will evaluate HR-QoL and topus size as clinically relevant outcomes.

Where more than two arms from a single trial are present, the by-arm comparisons are down-weighted following A'Hern (25) because all within trial comparisons are not independent. In all trial comparisons, this requires that a single 'control' comparator is determined. In trials with a true placebo, the placebo is the control comparator. In trials without placebo, then the control comparator is an intervention arm that best reflects usual care. For example, in a 5-arm trial with a true placebo there are 4 comparisons, and each comparison is down-weighted using analytic weights (10). This allows all arms from each trial to be evaluated in the meta-regression but adjusted for multiple comparisons with the control.

The primary and secondary analysis is prespecified as an all drug classes combined analysis. In addition to the STE, slope, $R^2_{\text{trial-level,}}$ and regression diagnostics, we will also evaluate the impact of effect modifiers; male sex, disease duration (less than 2 years, 2 to 10 years, more than 10 years), presence of clinical tophi (yes, no) on the SU and gout flare rate relationship. Furthermore, study design and other trial related methodological issues, including effect of differential cross-over, differential drop-out, whether trials included mandatory flare-prevention strategies such as mandatory colchicine and NSAIDs, GRADE ratings (20) , and risk of bias tool (19) ratings will also be explored.

The SIGN checklist (26) will be used to evaluate the methodology of longitudinal observational studies of gout.

Once these statistical results are available (i) serum urate reduction and (ii) serum urate target <6mg/dL will be evaluated as a surrogate endpoint gout using the BSES3 criteria.

DISCUSSION

It is important to emphasize that the evaluation of serum urate as a surrogate endpoint is for the context of using serum urate as an endpoint in clinical trials (surrogate biomarker). This is quite different to using SU to help guide clinical decision making, for example treating to a specific SU target, or to identify that treatment is working (monitoring biomarker). Although the meta-regression approach undertaken by the proposed study will help inform clinical decision making, the evidence needed for treatment targets requires a different research design.

Complete application of the BSES3 framework ideally also uses individual patient level data from multiple clinical trials. Although this analysis is planned, it is contingent upon agreement of relevant pharmaceutical companies to share their data and is therefore not a formal part of this protocol.

Observational studies will be included in the search strategy, but will be reported separately as a narrative review in light of the inherent risk of bias in non-randomised and uncontrolled observational study designs.

Contributors

LS, WT, ND, JF, JS, RC and ML, participated in the conception and design of this protocol.

RC, JF and ML provided statistical advice for the design and analysis.

LS, WT, ND, JF, RC and ML drafted the protocol.

LS, WT, ND, JF, MM, JS and RC and ML critically reviewed the manuscript for important intellectual content and approved the final version.

Funding

Musculoskeletal Statistics Unit, The Parker Institute, is supported by grants from the Oak Foundation.

Competing interests

None

References

1. Temple R. Clinical measurement in Drug Evaluation. In: Nimmo W, Tucker G, editors. A regulatory authority's opinion about surrogate endpoints: John Wiley & Sons Ltd; 1995. p. 3-22.

2. De Gruttola V, Clax P, DeMets D, Downing G, Ellenberg S, Friedman L, et al. Considerations in the evaluation of surrogate endpoints in clinical trials. Summary of a National Institutes of Health workshop. Control Clin Trials. 2001;22(5):485-502.

3. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharm Ther. 2001;69(3):89-95.

4. Schumacher H, Taylor W, Edwards L, Grainger R, Schlesinger N, Dalbeth N, et al. Outcome domains for studies of acute and chronic gout. Journal of Rheumatology. 2009;36(10):2342-5.

5. Lassere M, Johnson K, Boers M, Tugwell P, Brooks P, Simon L, et al. Definitions and validation criteria for biomarkers and surrogate endpoints: Development and testing of a quantitative hierarchical levels of evidence schema. J Rheumatol. 2007;34(3):607-15.

6. Maksymowych W, Landewé R, Tak P-P, Ritchlin C, Ostergaard M, Mease P, et al. Reappraisal of OMERACT 8 Draft Validation Criteria for a Soluble Biomarker Reflecting Structural Damage Endpoints in Rheumatoid Arthritis, Psoriatic Arthritis, and Spondyloarthritis: The OMERACT 9 v2 Criteria J Rheumatol. 2009;36(8):1785-91.

7. Stamp L, Zhu X, Dalbeth N, Jordan S, Edwards N, Taylor W. Serum urate as a soluble biomarker in chronic gout – evidence that serum urate fulfils the OMERACT validation criteria for soluble biomarkers. Semin Arthritis Rheum. 2011;40:483-500.

8. Lassere M. The biomarker-surrogacy evaluation schema: A review of the biomarker-surrogate literature and a proposal for a criterion-based, quantitative, multidimensional hierarchical levels of evidence schema for evaluating the status of biomarkers as surrogate endpoints. Statistical Methods in Medical Research. 2008;17(3):303-40.

9. Lassere M. The Innovative Evaluation Schema of Validating Putative Surrogate Endpoints. Joint Statistical Meeting; Vancouver, Canada 2010.

10. Lassere M, Johnson K, Schiff M, Rees D. Is blood pressure reduction a valid surrogate endpoint for stroke prevention? an analysis incorporating a systematic review of randomised controlled trials, a by-trial weighted errors-in-variables regression, the surrogate threshold effect (STE) and the biomarker-surrogacy (BioSurrogate) evaluation schema (BSES. BMC Medical Research Methodology. 2012;12:27.

11. Johnson K, Liauw W, Lassere M. Evaluating surrogacy metrics and investigating approval decisions of progression-free survival (PFS) in metastatic renal cell cancer: a systematic review. Ann Oncology. 2015;26:485-96.

12. Jordan KM, Cameron JS, Snaith M, Zhang W, Doherty M, Seckl J, et al. British Society for Rheumatology and British Health Professionals in Rheumatology guideline for the management of gout. Rheumatology. 2007;46(8):1372-4.

13. Khanna D, Fitzgerald J, Khanna P, Sangmee B, Singh M, Neogi T, et al. 2012 American College of Rheumatology Guidelines for the Management of Gout. Part 1: Systematic Nonpharmacologic and Pharmacologic Therapeutic Approaches to Hyperuricaemia. Arthritis Care Res. 2012;64(10):1431-46.

14. Zhang W, Doherty M, Bardin T, Pascual E, Barskova V, Conaghan P, et al. EULAR evidence based recommendations for gout. Part II: Management. Report of a task force of the EULAR Standing Committee for International

Clinical Studies Including Therapeutics (ESCISIT). Ann Rheum Dis. 2006;65(10):1312-24.

15. Stevenson M, Pandor A. Febuxostat for the management of hyperuricaemia in patients with gout: a NICE single technology appraisal. Pharmacoeconomics. 2011;29(2):133-40.

16. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015;349:g7647 doi: 10.1136/bmj.g7647.

17. Liberati A, Altman D, Tetzlaff J, Mulrow C, Gøtzsche P, Ioannidis J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol. 2009;62(10):e1-34. doi: 10.1016/j.jclinepi.2009.06.006.

18. Wallace S, Robinson H, Masi A, Decker J, McCarty D, Yu T. Preliminary criteria for the classification of the acute arthritis of primary gout. Arthritis Rheum. 1977;20(3):895-900.

19. Higgins J, Altman D, Gøtzsche P, Jüni P, Moher D, Oxman A, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011(;343:d5928. doi: 10.1136/bmj.d5928.).

20. Guyatt, Gordon et al. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. Journal of Clinical Epidemiology , Volume 66 , Issue 2 , 151 - 157

21. Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. Biostatistics. 2000;1(1):49-67.

22. Burzykowski T, Buyse M. Surrogate threshold effect: An alternative measure for meta-analytic surrogate endpoint validation. Pharmaceut Statist. 2006;5:173-86.

23. Johnson K, Ringland C, Stokes B. Response rate or time to progression as predictors of survival in trials of metastatic colorectal cancer or non-small-cell lung cancer: a meta-analysis. Lancet Oncology. 206;7:741-6.

24. Johnson, K.R., Freemantle, N., Anthony, D.M., Lassere, M.N.D. LDL-cholesterol differences predicted survival benefit in statin trials by the surrogate threshold effect (STE) (2009) *Journal of Clinical Epidemiology*, 62 (3), pp. 328-336.

25. A'Hern R, Ebbs S, Baum M. Does chemotherapy improve survival in advanced breast cancer? A statistical overview. Br J Cancer. 1988;57(6):615-8.

26. http://www.sign.ac.uk/methodology/checklists.html

BMJ Open

Using serum urate as a validated surrogate endpoint for flares in patients with gout: protocol for a systematic review and meta-regression analysis

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-012026.R1
Article Type:	Protocol
Date Submitted by the Author:	16-Aug-2016
Complete List of Authors:	Morillon, Melanie; Frederiksberg and Bispebjerg Hospital, Parker institute; Department of Rheumatology, Odense University Hospital, Stamp, Lisa; University of Otago, Christchurch, Department of Medicine, Taylor, William; University of Otago, Rehabilitation Teaching & Research Unit, Department of Medicine Fransen, Jaap; Radboud Uviversity Medical Centre, Rheumatology Dalbeth, Nicola; University of Auckland, Department of Immunology Singh, Jasvinder; University of Alabama at Birmingham & Birmingham Veterans Affairs, Department of Medicine Christensen, Robin; Frederiksberg and Bispebjerg Hospital, Parker Institute Lassere, Marissa; St George Hospital, Rheumatology
Primary Subject Heading:	Rheumatology
Secondary Subject Heading:	Rheumatology
Keywords:	GOUT, BIOMARKER, RHEUMATOLOGY, SURROGATE, OUTCOME

Using serum urate as a validated surrogate endpoint for flares in patients with gout: protocol for a systematic review and meta-regression analysis

Collaborators/Authors

[1,2]Melanie B. Morillon MD; [3]Lisa Stamp PhD, FRACP; [4]William Taylor PhD, FRACP; [5]Jaap Fransen PhD; [6]Nicola Dalbeth MD, FRACP; [7]Jasvinder A. Singh, MD, MPH; [1]Robin Christensen MSc, PhD; [8]Marissa Lassere MBBS (Hons) Grad Dip Epi, PhD, FRACP, FAFPHM;

Affiliations

[1]Musculoskeletal Statistics Unit, The Parker Institute, Department of Rheumatology, Copenhagen University Hospitals, Bispebjerg and Frederiksberg, Denmark

[2]Department of Rheumatology, Odense University Hospital, Denmark

[3]Department of Medicine, University of Otago, Christchurch, New Zealand

[4]Department of Medicine, University of Otago, Wellington, New Zealand

[5]JF Department of Rheumatology, Radboud University Medical Centre, Nijmegen, The Netherlands.

[6]Department of Medicine, University of Auckland, New Zealand.

[7]Department of Medicine, University of Alabama at Birmingham & Birmingham Veterans Affairs Medical Center, Birmingham, Alabama, USA.

[8]Department of Rheumatology, St George Hospital, University of NSW, Sydney, Australia

Correspondence to:

Melanie Birger Morillon
Department of Rheumatology,
Odense University Hospital,
5300 Svendborg,
Danmark, DK
Email: melanie.birger.morillon@rsyd.dk

ABSTRACT

Introduction: Gout is the most common inflammatory arthritis in men over 40 years. Long term urate-lowering therapy is considered a key strategy for effective gout management. The primary outcome measure for efficacy in clinical trials of urate lowering therapy is serum urate levels, effectively acting as a surrogate for patient-centred outcomes such as frequency of gout attacks or pain, yet it is not clearly demonstrated that the strength of the relationship between serum urate and clinically relevant outcomes is sufficiently strong for serum urate to be considered an adequate surrogate. Our objective is to investigate the strength of the relationship between changes in serum urate in randomised controlled trials and changes in clinically relevant outcomes according to the 'Biomarker-Surrogacy Evaluation Schema version 3' (BSES-3), documenting the validity of selected instruments by applying the 'OMERACT Filter 2.0'.

Methods and analysis: A systematic review described in terms of the PRISMA reporting guidelines will identify all relevant studies. Standardised data elements will be extracted from each study by 2 independent reviewers and disagreements resolved by discussion. The data will be analysed by meta-regression of the between-arm differences in the change in serum urate level (independent variable) from baseline to 3 months (or 6 and 12 months if 3 month values not available) against flare rate, tophus size and number and pain at the final study visit (dependent variables).

Ethics and dissemination: This study will not require specific ethics approval since it is based on analysis of published (aggregated) data. The intended audience will include health care researchers, policymakers and

clinicians. Results of the study will be disseminated by peer-review publication.

Protocol registration: In Prospero, CRD42016026991. Accepted January 2016.

STRENGHTS AND LIMITATIONS

- Our study's strengths include clinical expertise in rheumatology
- The content experts in the group have extensive knowledge of the literature and ▫experience with gout treatment
- The methodologists in the group are members of the OMERACT Gout Working Group, and have experience with conducting and reporting randomised clinical trials, systematic reviews and meta-analyses.
- A possible and anticipated weakness may be the quantity and quality of the trials we identify

INTRODUCTION

Clinicians making treatment decisions should refer to methodologically strong clinical trials examining the impact of therapy on clinically important outcomes (i.e. outcomes that are important to patients). However, clinically important outcomes can be difficult to study, as the required trials need very large sample sizes, or long-term patient follow-up. Thus researchers or drug developers look for alternatives. Substituting surrogate end points for the target event allows conduct of shorter and smaller trials, thus offering a solution to the dilemma, if the endpoints are convincing as surrogate endpoints.

There are obvious advantages to using biomarkers and surrogate endpoints, but concerns about clinical applicability and statistical validity to evaluate these aspects hinder their efficient application. A surrogate end point may be defined as an "objective" laboratory measurement or a physical sign used as a substitute for a clinically meaningful end point that measures directly how a patient feels, functions or survives (1). This definition was recommended and further explored at a National Institute of Health (NIH) sponsored workshop in 1998 which agreed on definitions for biomarker, surrogate endpoint and clinical endpoint. The agreed definition of a biomarker states "*a biological marker (biomarker) is a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes or pharmacologic responses to a therapeutic intervention*" (2) (3).

In gout, monosodium crystal formation occurs when super-saturation levels are reached ~6.8mg/dL (0.41mmol/l) at 37^0C. Reduction in serum

urate (SU) to <6mg/dL (0.36mmol/L) is a key goal in the long-term management of gout. As such SU measurement has become an integral part of the management of gout and a critical outcome measure in clinical studies of gout therapies. The OMERACT Delphi exercise identified SU as a mandatory outcome measure in chronic gout studies with the highest median rating (4). SU as a biomarker makes inherent sense given the strong relationship between the risk of gout and SU. However, is SU a surrogate endpoint of relevant clinical outcomes such as gout attacks, tophus regression and radiological damage?

Background

At OMERACT (Outcome Measures in Rheumatology Clinical Trials) 8 (Malta, 2006) Lassere et al proposed a schema for evaluation of biomarkers as surrogate endpoints (5). The schema was operationalized as a score obtained from four domains: target outcome, study design, statistical strength and penalties (5). This schema was based on the NIH definitions of biomarker, surrogate endpoint and clinical endpoint published in 2001 (2). The distinction between a surrogate and a biomarker was determined by the strength of association between the biomarker and the clinical endpoint of interest. To be called a surrogate, it was proposed that a biomarker must meet the rank (score) of at least 3 within the Target Outcome, Study Design, and Statistical Strength domains, and there must not be evidence from a Randomised Controlled Trial (RCT) that use of the biomarker caused patient harm (5).

At OMERACT 9 (Kananaskis, 2008) the soluble biomarker group revised the requirements for the specific situation of a soluble biomarker being predictive of structural radiographic damage in ankylosing spondylitis, psoriatic arthritis and rheumatoid arthritis (6). There was an increased emphasis on the technical assay requirements of the biomarker but the strength of association domain, while discussed in the text, did not appear in the OMERACT 9 levels of evidence framework. There was no consensus on all aspects of the framework, and the criteria by which a soluble biomarker could be said to meet the levels of evidence framework were not defined.

At OMERACT 10 (Kota Kinabalu, 2010) evidence was presented that SU fulfilled the OMERACT 9 soluble biomarker requirements in terms of domain 4 (performance criteria) and limited evidence from observational studies and one RCT that changes in SU were associated with changes in patient-centred outcomes for the disease of gout (7). However, the meeting did not endorse SU as a biomarker for clinically relevant outcomes for gout. The reasons for the lack of endorsement might be that the strength of evidence was weak, the criteria for endorsement are unclear and the chosen patient-centred outcomes (particularly number of flares) were not universally held to be clinically meaningful.

In parallel to OMERACT, Lassere et al systematically reviewed the biomarker-surrogate literature and modified the levels of evidence schema built on the OMERACT 8 proposal (5) which over time went through 3 iterations ('Biomarker-Surrogacy Evaluation Schema version' (BSES), BSES1 which was the OM 8 proposal (5), BSES2 which specified the statistical criteria more precisely (8) and BSES3 which replaced the penalties domain

with a combined clinical and pharmacological generalizability domain). BSES3 contains 4 domains: study design, target outcome, statistical evaluation, and generalizability. It also specified the kind of statistical association required to justify the link between biomarker and clinical endpoint being sufficiently strong to consider the biomarker as a surrogate endpoint (9) (10).

In 2012 blood pressure was evaluated using the BSES3 and online material described its application and interpretation (10). The BSES3 framework represents the currently best available approach to validating a biomarker as a surrogate endpoint. We propose that this framework be endorsed by OMERACT as the framework for validation of biomarker-surrogates for rheumatology clinical trials. It represents the logical extension of work developed at OMERACT 8 and provides a clear pathway by which a putative biomarker, soluble or otherwise, can be evaluated, in contrast to the OMERACT 9 framework. For example, Lassere et al has used trial-level data and the BSES3 framework to convincingly show that diastolic and systolic blood pressure are valid surrogate endpoints for stroke risk reduction (10). In a recent meta-regression the approach has also been used to evaluate progression-free survival (PFS) in metastatic renal cell carcinoma (11).

Rationale

We wish to use the example of SU as a soluble biomarker for the major clinical endpoint of acute gout attacks, in the disease of gout. A minor clinical endpoint would be tophus size change from baseline to final visit,

the change in the number of tophi, and pain. Other patient relevant endpoints included in the OMERACT core-set of outcomes for clinical trials in patients with chronic gout will also be evaluated in exploratory analyses: health related quality of life (HRQOL), patient global assessment of disease activity, and physical disability (activities limitation).

The justification for choosing this biomarker and the clinical endpoint of flares as the major endpoint is described as follows:

- Firstly, SU is recommended as a treatment target by several guidelines for the management of gout (12-14). This strongly implies (although it is not stated explicitly) that changes in SU or achievement of a target level of SU will be strongly associated with clinically relevant outcomes.
- Secondly, some regulatory bodies (e.g. Food and Drug Administration and European Medicines Agency) have tended to assume that beneficial drug effects on SU will likely have beneficial effects on clinical outcomes in gout. NICE recommended that febuxostat be available for people who are intolerant of allopurinol or who have contraindications to allopurinol (15). In other words, although NICE did not see persuasive evidence for improved clinical outcomes with the use of febuxostat, it was sufficient that the drug effectively lowered SU to below 6 mg/dL.
- Thirdly, we have previously shown that SU fulfils the technical performance criteria for a valid soluble biomarker proposed at OMERACT 9 (7). Flare (acute attack) of gout is a key clinical

manifestation of gout. It constitutes the primary or only manifestation for several years until persistent, tophaceous disease develops. In the expectation that effective management strategies aim to prevent chronic tophaceous disease from occurring it is justifiable to focus on attacks as the clinically relevant endpoint for the majority of people for gout. Although gout attacks can vary in severity (often modified by acute gout treatment), it is clear that every attack is associated with some level of symptoms and disability. Gout attacks therefore align with how a patient 'feels or functions' and can be reasonably be identified as a clinically relevant endpoint (1).

However, we recognize that other clinical outcomes are relevant and will evaluate these within the same framework. This proposal fits in the Filter 2.0 framework by making explicit and quantifying the link between Core Area domains of Pathophysiology Manifestations (biomarker) and domains of Life Impact (flare, pain, HRQOL, tophus). This framework links disease-centred variables of biological and pathological processes with patient-centred variables of how a patient feels, functions and survives as proposed at OMERACT 6 (5).

Objectives

There are two objectives:

1) To determine the strength of the relationship between SU and patient relevant outcomes, including flares, tophi, HRQOL, pain and function using meta-regression of randomised controlled trials.
2) To evaluate whether SU is a surrogate endpoint for clinically relevant outcomes in patients with gout as defined by the BSES3 framework.

Hypothesis

A reduction in SU will be associated with improvement in clinically relevant patient reported outcomes including gout flares and tophus size/number.

METHODS AND ANALYSIS

Protocol and registration

The protocol for the systematic review and meta-regression analysis was prepared while planning and documenting the review methods, guarding the project team against arbitrary decision making during review conduct, and to prompt global collaboration (16). Our protocol was prepared according to the recommendations given in PRISMA-P (16) and registered on PROSPERO (CRD42016026991); this protocol and coming manuscripts will conform to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for reporting systematic reviews and meta-analyses (17).

Eligibility criteria

The eligibility criteria for objective 1 is any randomised controlled trial comparing an active drug (alone or in combination) in patients with gout

with any control or placebo, with a minimum duration of three months. The eligibility criteria for objective 2 are any randomised controlled trial, controlled clinical trial, or open label trial (OLT) comparing an (apparently) active drug (alone or in combination) in patients with gout with any control or placebo, with a minimum duration of three months and longitudinal observational studies of gout with a minimum duration of 3 months.

For both criteria, patients will be at least 18 years of age and meeting the preliminary American College of Rheumatology (ACR) criteria for acute arthritis of primary gout (18) or given a diagnosis of gout as described by the authors.

Search and selection of trials

The following electronic databases will be searched: PubMed, EMBASE, the Cochrane Library including the Cochrane Central Register of Controlled Trials (CENTRAL) and Cochrane Database of Systematic Reviews (CDSR). The search will be limited to English language studies in humans, but not limited by year of publication. The reference lists from comprehensive reviews and identified clinical trials are also manually searched.

Results of the various searches will be reviewed independently by two authors (LS and MM). Titles and abstracts will be reviewed and if further information is required (to assess eligibility criteria), the full text will be obtained. A record of reasons for excluding studies will be kept enabling generation of a figure illustrating the flow of information through the different phases of the systematic review continuing to meta-regression

analysis. Disagreements will be resolved by an independent third mediator (WT).

Data extraction

EndNote X7 software will be used to manage the records retrieved from searches of electronic databases. Results from hand searches will be tracked on a Microsoft Excel spreadsheet. A customised data extraction form will be created in Microsoft Excel to capture all the information available for each individual trial.

The biomarker is defined as the change in SU from baseline to 3 months, or where 3 month values are not available, the value at 6 months or 12 months (in order of preference). This can be estimated if only baseline and change is reported.

The clinical endpoints (dependent variables) are defined as follows:

- Major outcome: gout-flares
- Minor outcomes: size of sentinel tophus (if size was not measured, we will use number, or presence/absence in order of preference) and pain at final study visit'

Exploratory analyses: health related quality of life (HRQOL) (SF36), patient global assessment of disease activity, and physical disability (activities limitation; e.g., HAQ).

Effect sizes for continuous endpoints will be recorded as the standardised mean difference. If there is more than one active treatment arm, analysis will treat this as a separate study i.e. sub-study (see Meta-regression

Analysis). All variable values will be based upon the intention-to-treat population from each study whenever possible.

Risk of bias in individual studies and judging the quality of evidence

The RCTs will be assessed for methodological quality (i.e. internal validity) using the Cochrane Risk of Bias tool (19). If at least one of the domains is rated as inadequate, the trial will be considered at high risk of bias. If all domains are judged as low, the trial will be considered at low risk of bias. Otherwise, the trial is considered as having unclear risk of bias. Data extraction and risk-of-bias assessment will be performed independently by 2 reviewers; disagreements will be resolved by a third reviewer. While interpreting the overall findings after the meta-analysis etc., GRADE (Grading of Recommendations Assessment, Development and Evaluation) will be used to rate the overall quality of the evidence based on both the apparent risk of bias, publication bias, imprecision, inconsistency, indirectness, and magnitude of effect; i.e., the GRADE ratings of very low, low, moderate, or high-quality evidence per outcome will reflect the extent to which we are confident that the effect estimates are correct (20)

Meta-regression analysis

To combine the individual study results, we will perform meta-analyses using SAS software (PROC MIXED version 9. 3; SAS Institute Inc., Cary, NC, USA), applying a restricted maximum likelihood (REML) method to estimate the between-study variance (i.e. T^2) and the combined estimate of effect. We will estimate the anticipated heterogeneity between trials with a

standard (Cochran's) Q-test statistic, and we will evaluate this based on the I^2 value, which is interpreted as the percentage of variability in treatment effect estimates that is due to between study heterogeneity rather than chance. Although our meta-regression analysis is undertaken correctly from a technical point of view, relations with averages of patients' characteristics can be potentially misleading. Thus, following our systematic review, we will attempt to get access to individual participant datasets investigating patients' characteristics; this will to some extent move us away from looking at relations across trials, to inspection of relations within trials (21)

The primary purpose of this project is to evaluate the surrogacy status of SU as a "predictor" of gout flare rate reduction using meta-regression of randomised controlled trials. Randomisation is essential for the causal surrogacy relationship, therefore, only randomised controlled trials will be included in the main meta-regression analysis. Non-randomised study designs will be summarised separately by meta-regression to confirm the consistency of association between the biomarker and clinical endpoints in other contexts. Cohort studies will be summarised as a narrative review. The analyses of both randomised and non-randomised studies contribute to the evaluation of serum urate within the BSES3 framework

Furthermore, in the meta-regression, the relationship between serum urate and clinically relevant outcomes can be undertaken using different outcome metrics. We will define these as primary and secondary analyses. In the primary analysis the dependent variable is a rate ratio (i.e. an incidence density ratio) comparing the ratio of incidence rates of gout flare

events in active versus control arms occurring at any given point in time; incidence rate is the occurrence of an event over person-time (i.e., in this setting in person-months). The rate ratio allows trials of different duration to be included in the analysis. The independent variable is between arm difference of within-arm change (on-trial SU from baseline SU) of SU. Therefore, in a trial of 3 months duration, flare rate over 3 months is the dependent variable and change in SU over 3 months is the independent variable.

In secondary analyses the dependent variable is risk ratio reduction (RRR) of within trial gout flare rate. The relative ratio reduction (also called the risk ratio reduction) is the flare risk in the control arm minus the flare risk in the active arm, divided by the flare risk in the control arm (this can also be calculated by 1- Relative Risk (RR), where relative risk is the flare risk in the active arm divided by the flare risk in the control arm). Therefore the relative risk reduction (RRR) is the difference in flare risk in two arms (control-active), expressed as a percentage of the risk of the control arm.

The independent variable is within trial, by-arm difference of proportion with SU less than 6mg/dL at the end of the trial.

In a randomised controlled trial, by-arm difference in SU change is likely to be causal and change in SU is easily interpretable as a surrogacy metric in gout by clinicians. Relative risk reduction is more familiar to clinicians than rate ratio but ignores trial duration. Although SU less than 6mg/dL is the most common primary endpoint of RCTs of gout interventions, a by-arm difference in proportion achieving a SU target may be more difficult to interpret than a serum urate change. In addition to gout flares,

the SU as a surrogate endpoint for two other clinical outcomes, HRQoL and tophus size, will also be evaluated as secondary clinical outcomes. If the trial does not report these outcomes, the authors will be contacted and the by-arm outcomes requested.

A quantitative evaluation of trial-level statistical surrogacy using the BSES3 (10) includes determining the slope coefficient of the surrogacy relationship, trial-level R^2 (coefficient of determination) (22)and the Surrogate Threshold Effect (STE) (23, 24, 25) and Surrogate Threshold Effect Proportion (STEP) (8, 10) of the surrogate and true-clinical-endpoint relationship using data from a meta-regression of randomized controlled trials.

The STE is informative as it captures both the slope and dispersion of the surrogate-true relationship in a single metric (25). The STE is the serum urate difference needed to predict the primary clinical endpoint, gout flare rate ratio, in a new trial, if only serum urate is measured in the new trial. The STE is determined by comparing the difference between control and active arms SU and flare rate respectively as follows: (i) calculate the SU change and gout flare rate ratio based on each arm in each trial, (ii) calculate the difference between control and active arms for SU change and gout flare rate ratio, (iii) regress SU and gout flare rate ratio difference values using weighted by trial size errors-in-variables (specifying a reliability coefficient of 0.9) regression and by a weighted by trial-size meta-regression (as a sensitivity analysis), (iv) calculate the 95% prediction limits of the regression, and (v) find the SU value where the 95% prediction line intersects with the horizontal flare rate x-axis of no flare rate ratio benefit

(where the flare rate ratio y-axis is equal to 1.0). Similar analyses will be explored with flare rate relative risk reduction and proportion with SU less than 6mg/dL at the end of the trial. In this analysis the interest is the SU target <6mg/dL by-arm proportion where the 95% prediction line intersects with the horizontal flare rate x-axis of no flare relative risk reduction benefit (i.e. where the flare relative risk reduction y-axis is equal to zero). Subsequent analyses will evaluate HR-QoL and topus size as clinically relevant outcomes.

Where more than two arms from a single trial are present, the by-arm comparisons are down-weighted following A'Hern (26) because all within trial comparisons are not independent. In all trial comparisons, this requires that a single 'control' comparator is determined. In trials with a true placebo, the placebo is the control comparator. In trials without placebo, then the control comparator is an intervention arm that best reflects usual care. For example, in a 5-arm trial with a true placebo there are 4 comparisons, and each comparison is down-weighted using analytic weights (10). This allows all arms from each trial to be evaluated in the meta-regression but adjusted for multiple comparisons with the control.

The primary and secondary analysis is prespecified as an all drug classes combined analysis. In addition to the STE, slope, $R^2_{\text{trial-level,}}$ and regression diagnostics, we will also evaluate the impact of effect modifiers; male sex, disease duration (less than 2 years, 2 to 10 years, more than 10 years), presence of clinical tophi (yes, no) on the SU and gout flare rate relationship. Furthermore, study design and other trial related methodological issues, including effect of differential cross-over,

differential drop-out, whether trials included mandatory flare-prevention strategies such as mandatory colchicine and NSAIDs, GRADE ratings (20) , and risk of bias tool (19) ratings will also be explored.

The SIGN checklist (27) will be used to evaluate the methodology of longitudinal observational studies of gout.

Once these statistical results are available (i) serum urate reduction and (ii) serum urate target <6mg/dL will be evaluated as a surrogate endpoint gout using the BSES3 criteria.

DISCUSSION

It is important to emphasize that the evaluation of serum urate as a surrogate endpoint is for the context of using serum urate as an endpoint in clinical trials (surrogate biomarker). This is quite different to using SU to help guide clinical decision making, for example treating to a specific SU target, or to identify that treatment is working (monitoring biomarker). Although the meta-regression approach undertaken by the proposed study will help inform clinical decision making, the evidence needed for treatment targets requires a different research design.

Complete application of the BSES3 framework ideally also uses individual patient level data from multiple clinical trials. Although this analysis is planned, it is contingent upon agreement of relevant pharmaceutical companies to share their data and is therefore not a formal part of this protocol.

Observational studies will be included in the search strategy, but will be reported separately as a narrative review in light of the inherent risk of bias in non-randomised and uncontrolled observational study designs.

Contributors

LS, WT, ND, JF, JS, RC and ML, participated in the conception and design of this protocol.

RC, JF and ML provided statistical advice for the design and analysis.

LS, WT, ND, JF, RC, MM, JS and ML drafted the protocol.

LS, WT, ND, JF, MM, JS and RC and ML critically reviewed the manuscript for important intellectual content and approved the final version.

Funding

Musculoskeletal Statistics Unit, The Parker Institute, is supported by grants from the Oak Foundation.

Competing interests

None

References

1. Temple R. Clinical measurement in Drug Evaluation. In: Nimmo W, Tucker G, editors. A regulatory authority's opinion about surrogate endpoints: John Wiley & Sons Ltd; 1995. p. 3-22.

2. De Gruttola V, Clax P, DeMets D, Downing G, Ellenberg S, Friedman L, et al. Considerations in the evaluation of surrogate endpoints in clinical trials. Summary of a National Institutes of Health workshop. Control Clin Trials. 2001;22(5):485-502.

3. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharm Ther. 2001;69(3):89-95.

4. Schumacher H, Taylor W, Edwards L, Grainger R, Schlesinger N, Dalbeth N, et al. Outcome domains for studies of acute and chronic gout. Journal of Rheumatology. 2009;36(10):2342-5.

5. Lassere M, Johnson K, Boers M, Tugwell P, Brooks P, Simon L, et al. Definitions and validation criteria for biomarkers and surrogate endpoints: Development and testing of a quantitative hierarchical levels of evidence schema. J Rheumatol. 2007;34(3):607-15.

6. Maksymowych W, Landewé R, Tak P-P, Ritchlin C, Ostergaard M, Mease P, et al. Reappraisal of OMERACT 8 Draft Validation Criteria for a Soluble Biomarker Reflecting Structural Damage Endpoints in Rheumatoid Arthritis, Psoriatic Arthritis, and Spondyloarthritis: The OMERACT 9 v2 Criteria J Rheumatol. 2009;36(8):1785-91.

7. Stamp L, Zhu X, Dalbeth N, Jordan S, Edwards N, Taylor W. Serum urate as a soluble biomarker in chronic gout – evidence that serum urate fulfils the OMERACT validation criteria for soluble biomarkers. Semin Arthritis Rheum. 2011;40:483-500.

8. Lassere M. The biomarker-surrogacy evaluation schema: A review of the biomarker-surrogate literature and a proposal for a criterion-based, quantitative, multidimensional hierarchical levels of evidence schema for evaluating the status of biomarkers as surrogate endpoints. Statistical Methods in Medical Research. 2008;17(3):303-40.

9. Lassere M. The Innovative Evaluation Schema of Validating Putative Surrogate Endpoints. Joint Statistical Meeting; Vancouver, Canada 2010.

10. Lassere M, Johnson K, Schiff M, Rees D. Is blood pressure reduction a valid surrogate endpoint for stroke prevention? an analysis incorporating a systematic review of randomised controlled trials, a by-trial weighted errors-in-variables regression, the surrogate threshold effect (STE) and the biomarker-surrogacy (BioSurrogate) evaluation schema (BSES. BMC Medical Research Methodology. 2012;12:27.

11. Johnson K, Liauw W, Lassere M. Evaluating surrogacy metrics and investigating approval decisions of progression-free survival (PFS) in metastatic renal cell cancer: a systematic review. Ann Oncology. 2015;26:485-96.

12. Jordan KM, Cameron JS, Snaith M, Zhang W, Doherty M, Seckl J, et al. British Society for Rheumatology and British Health Professionals in Rheumatology guideline for the management of gout. Rheumatology. 2007;46(8):1372-4.

13. Khanna D, Fitzgerald J, Khanna P, Sangmee B, Singh M, Neogi T, et al. 2012 American College of Rheumatology Guidelines for the Management of Gout. Part 1: Systematic Nonpharmacologic and Pharmacologic Therapeutic Approaches to Hyperuricaemia. Arthritis Care Res. 2012;64(10):1431-46.

14. Zhang W, Doherty M, Bardin T, Pascual E, Barskova V, Conaghan P, et al. EULAR evidence based recommendations for gout. Part II: Management. Report of a task force of the EULAR Standing Committee for

International Clinical Studies Including Therapeutics (ESCISIT). Ann Rheum Dis. 2006;65(10):1312-24.

15. Stevenson M, Pandor A. Febuxostat for the management of hyperuricaemia in patients with gout: a NICE single technology appraisal. Pharmacoeconomics. 2011;29(2):133-40.

16. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015;349:g7647 doi: 10.1136/bmj.g7647.

17. Liberati A, Altman D, Tetzlaff J, Mulrow C, Gøtzsche P, Ioannidis J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol. 2009;62(10):e1-34. doi: 10.1016/j.jclinepi.2009.06.006.

18. Wallace S, Robinson H, Masi A, Decker J, McCarty D, Yu T. Preliminary criteria for the classification of the acute arthritis of primary gout. Arthritis Rheum. 1977;20(3):895-900.

19. Higgins J, Altman D, Gøtzsche P, Jüni P, Moher D, Oxman A, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011(;343:d5928. doi: 10.1136/bmj.d5928.).

20. Guyatt, Gordon et al. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. Journal of Clinical Epidemiology , Volume 66 , Issue 2 , 151 - 157

21. Thompson, S. G. and J. P. Higgins. "Treating individuals 4: can meta-analysis help target interventions at individuals most likely to

benefit?" Lancet. 2005; 365: 341–46.

22. Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. Biostatistics. 2000;1(1):49-67.

23. Burzykowski T, Buyse M. Surrogate threshold effect: An alternative measure for meta-analytic surrogate endpoint validation. Pharmaceut Statist. 2006;5:173-86.

24. Johnson K, Ringland C, Stokes B. Response rate or time to progression as predictors of survival in trials of metastatic colorectal cancer or non-small-cell lung cancer: a meta-analysis. Lancet Oncology. 206;7:741-6.

25. Johnson, K.R., Freemantle, N., Anthony, D.M., Lassere, M.N.D. LDL-cholesterol differences predicted survival benefit in statin trials by the surrogate threshold effect (STE) (2009) *Journal of Clinical Epidemiology*, 62 (3), pp. 328-336.

26. A'Hern R, Ebbs S, Baum M. Does chemotherapy improve survival in advanced breast cancer? A statistical overview. Br J Cancer. 1988;57(6):615-8.

27. http://www.sign.ac.uk/methodology/checklists.html

PRISMA-P 2015 Checklist

This checklist has been adapted for use with protocol submissions to *Systematic Reviews* from Table 3 in Moher D et al: Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews* 2015 **4**:1

Section/topic	#	Checklist item	Information reported		Line number(s)
			Yes	No	
ADMINISTRATIVE INFORMATION					
Title					
Identification	1a	Identify the report as a protocol of a systematic review	x☐	☐	Page 1:3-6
Update	1b	If the protocol is for an update of a previous systematic review, identify as such	☐	☐	
Registration	2	If registered, provide the name of the registry (e.g., PROSPERO) and registration number in the Abstract	x☐	☐	Page 3:8-9
Authors					
Contact	3a	Provide name, institutional affiliation, and e-mail address of all protocol authors; provide physical mailing address of corresponding author	x☐	☐	Page1: 10-54
Contributions	3b	Describe contributions of protocol authors and identify the guarantor of the review	x☐	☐	Page 20: 3-16
Amendments	4	If the protocol represents an amendment of a previously completed or published protocol, identify as such and list changes; otherwise, state plan for documenting important protocol amendments	x☐	☐	Page 20: 3-16
Support					
Sources	5a	Indicate sources of financial or other support for the review	x☐	☐	Page 20: 23-26
Sponsor	5b	Provide name for the review funder and/or sponsor	x☐	☐	Page 20: 23-26
Role of sponsor/funder	5c	Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol	x☐	☐	Page 20: 23-26
INTRODUCTION					
Rationale	6	Describe the rationale for the review in the context of what is already known	x☐	☐	Page 7: 50-57 Page 8:1-57

Section/topic	#	Checklist item	Information reported		Line number(s)
			Yes	No	
					page 9: 1-46
Objectives	7	Provide an explicit statement of the question(s) the review will address with reference to participants, interventions, comparators, and outcomes (PICO)	x☐	☐	Page9: 50-53 page 10: 1-12
METHODS					
Eligibility criteria	8	Specify the study characteristics (e.g., PICO, study design, setting, time frame) and report characteristics (e.g., years considered, language, publication status) to be used as criteria for eligibility for the review	x☐	☐	page10:53-57 page11:1-23
Information sources	9	Describe all intended information sources (e.g., electronic databases, contact with study authors, trial registers, or other grey literature sources) with planned dates of coverage	x☐	☐	Page11:28-41
Search strategy	10	Present draft of search strategy to be used for at least one electronic database, including planned limits, such that it could be repeated	x☐	☐	Will be uploades seperately
STUDY RECORDS					
Data management	11a	Describe the mechanism(s) that will be used to manage records and data throughout the review	x☐	☐	Page 12:10-21
Selection process	11b	State the process that will be used for selecting studies (e.g., two independent reviewers) through each phase of the review (i.e., screening, eligibility, and inclusion in meta-analysis)	x☐	☐	Page 11:44-57 page 12: 1-5
Data collection process	11c	Describe planned method of extracting data from reports (e.g., piloting forms, done independently, in duplicate), any processes for obtaining and confirming data from investigators	x☐	☐	Page 12: 10-21
Data items	12	List and define all variables for which data will be sought (e.g., PICO items, funding sources), any pre-planned data assumptions and simplifications	x☐	☐	Page 12: 23-57 Page13: 3-5
Outcomes and prioritization	13	List and define all outcomes for which data will be sought, including prioritization of main and additional outcomes, with rationale	x☐	☐	Page 12: 23-48
Risk of bias in individual studies	14	Describe anticipated methods for assessing risk of bias of individual studies, including whether this will be done at the outcome or study level, or both; state how this information will be used in data synthesis	x☐	☐	Page:13: 10-26
DATA					

Section/topic	#	Checklist item	Information reported		Line number(s)
			Yes	No	
Synthesis	15a	Describe criteria under which study data will be quantitatively synthesized	x☐	☐	Page 12: 23-48
	15b	If data are appropriate for quantitative synthesis, describe planned summary measures, methods of handling data, and methods of combining data from studies, including any planned exploration of consistency (e.g., I^2, Kendall's tau)	x☐	☐	Page 13:48-57 page14-17 page 18:3-19
	15c	Describe any proposed additional analyses (e.g., sensitivity or subgroup analyses, meta-regression)	x☐	☐	13: 46-57 page14:3-24
	15d	If quantitative synthesis is not appropriate, describe the type of summary planned	☐	x☐	
Meta-bias(es)	16	Specify any planned assessment of meta-bias(es) (e.g., publication bias across studies, selective reporting within studies)	x☐	☐	13: 10-41
Confidence in cumulative evidence	17	Describe how the strength of the body of evidence will be assessed (e.g., GRADE)	x☐	☐	page16: 26-41