BMJ Open

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

Melanie B Morillon,[1,2] Lisa Stamp,[3] William Taylor,[4] Jaap Fransen,[5] Nicola Dalbeth,[6] Jasvinder A Singh,[7] Robin Christensen,[1] Marissa Lassere[8]

CrossMark

For numbered affiliations see end of article.

Correspondence to
Dr Melanie Birger Morillon;
melanie.birger.morillon@rsyd.dk

ABSTRACT

Introduction: Gout is the most common inflammatory arthritis in men over 40 years of age. 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' (BSES3), documenting the validity of selected instruments by applying the 'OMERACT Filter 2.0'.

Methods and analysis: A systematic review described in terms of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) reporting guidelines will identify all relevant studies. Standardised data elements will be extracted from each study by 2 independent reviewers and disagreements are 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 are 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 healthcare researchers, policymakers and clinicians. Results of the study will be disseminated by peer-reviewed publications.

Trial registration number: CRD42016026991.

Strengths and limitations of this study

- 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 Outcome Measures in Rheumatology Clinical Trials (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.

on clinically important outcomes (ie, 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 end points are convincing as surrogate end points.

There are obvious advantages to using biomarkers and surrogate end points, 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,

INTRODUCTION

Clinicians making treatment decisions should refer to methodologically strong clinical trials examining the impact of therapy

BMJ

surrogate end point and clinical end point. 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 supersaturation levels of ~6.8 mg/dL (0.41 mmol/L) are reached at 37°C. Reduction in serum urate (SU) to <6 mg/dL (0.36 mmol/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 Outcome Measures in Rheumatology Clinical Trials (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 end point of relevant clinical outcomes such as gout attacks, tophus regression and radiological damage?

Background

At OMERACT 8 (Malta, 2006) Lassere et al[5] proposed a schema for the evaluation of biomarkers as surrogate end points. The schema was operationalised 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 end point and clinical end point 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 end point of interest. To be called a surrogate, it was proposed that a biomarker must meet the rank (score) of at least three within the target outcome, study design and statistical strength domains, and there must not be evidence from a randomised controlled trial (RCT) that the 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 there is limited evidence from observational studies and one RCT that changes in SU

were associated with changes in patient-centred outcomes for 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 the number of flares) were not universally held to be clinically meaningful.

In parallel to OMERACT, Lassere et al[5] systematically reviewed the biomarker–surrogate literature and modified the levels of evidence schema built on the OMERACT 8 proposal which over time went through three iterations ('Biomarker-Surrogacy Evaluation Schema version' (BSES), BSES1 which was the OMERACT 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 generalisability domain). BSES3 contains four domains: study design, target outcome, statistical evaluation and generalisability. It also specified the kind of statistical association required to justify the link between the biomarker and the clinical end point being sufficiently strong to consider the biomarker as a surrogate end point.[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 end point. 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[10] have used trial-level data and the BSES3 framework to convincingly show that diastolic and systolic blood pressures are valid surrogate end points for stroke risk reduction. 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 end point of acute gout attacks, in the disease of gout. A minor clinical end point would be tophus size change from baseline to final visit, the change in the number of tophi, and pain. Other patient relevant end points 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 end point of flares as the major end point is described as follows:

- First, 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.

- Second, some regulatory bodies (eg, 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. National Institute for Health and Care Excellence (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.

- Third, 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 end point 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 end point.[1]

However, we recognise 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 RCTs.
2. To evaluate whether SU is a surrogate end point 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 Preferred Reporting Items for Systematic Reviews and Meta-Analysis Protocols (PRISMA-P)[16] and registered on PROSPERO (CRD42016026991); this protocol and coming manuscripts will conform to the PRISMA guidelines for reporting systematic reviews and meta-analyses.[17]

Eligibility criteria
The eligibility criteria for objective 1 is any RCT comparing an active drug (alone or in combination) in patients with gout with any control or placebo, with a minimum duration of 3 months. The eligibility criteria for objective 2 are any RCT, 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 3 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 MBM). 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 end points (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: HRQOL (36-Item Short Form Health Survey, SF-36), patient global assessment of disease activity, and physical disability (activities limitation; eg, Health Assesment Questionnaire (HAQ)).

Effect sizes for continuous end points 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 that is, substudy (see Meta-regression analysis section). All variable values will be based on 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 (ie, 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 two reviewers; disagreements will be resolved by a third reviewer. While interpreting the overall findings after the meta-analysis, etc, Grading of Recommendations Assessment, Development and Evaluation (GRADE) 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; that is, the GRADE ratings of very low-quality, low-quality, moderate-quality 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 V.9. 3; SAS Institute, Cary, North Carolina, USA), applying a restricted maximum likelihood (REML) method to estimate the between-study variance (ie, 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 RCTs. Randomisation is essential for the causal surrogacy relationship; therefore, only RCTs 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 end points 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 SU within the BSES3 framework.

Furthermore, in the meta-regression, the relationship between SU 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 (ie, 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 (ie, 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 RR 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<6 mg/dL at the end of the trial.

In a RCT, 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<6 mg/dL is the most common primary end point of RCTs of gout interventions, a by-arm difference in proportion achieving an SU target may be more difficult to interpret than a SU change. In addition to gout flares, the SU as a surrogate end point 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 STE proportion (STEP)[8][10] of the surrogate and true-clinical-end point relationship using data from a meta-regression of RCTs.

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 SU difference needed to predict the primary clinical end point, gout flare rate ratio, in a new trial, if only SU 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: (1) calculate the SU change and gout flare rate ratio based on each arm in each trial, (2) calculate the difference between control and active arms for SU change and gout flare rate ratio, (3) 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), (4) calculate the 95% prediction limits of the regression and (5) 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<6 mg/dL at the end of the trial. In this analysis the interest is the SU target <6 mg/dL by-arm proportion where the 95% prediction line intersects with the horizontal flare rate x-axis of no flare relative risk reduction benefit (ie, where the flare relative risk reduction y-axis is equal to zero). Subsequent analyses will evaluate HRQoL and tophus 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 et al[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 five-arm trial with a true placebo there are four 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 the secondary analyses are prespecified as an all drug classes combined analysis. In addition to the STE, slope, $R^2_{trial-level}$, and regression diagnostics, we will also evaluate the impact of effect modifiers; male sex, disease duration (<2 years, 2–10 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 the effect of differential cross-over, differential drop-out, whether trials included mandatory flare-prevention strategies such as mandatory colchicine and non-steroidal anti-inflammatory drug (NSAID), GRADE ratings[20] and risk of bias tool[19] ratings will also be explored.

The Scottish Intercollegiate Guidelines Network (SIGN) checklist[27] will be used to evaluate the methodology of longitudinal observational studies of gout.

Once these statistical results are available (1) SU reduction and (2) SU target <6 mg/dL will be evaluated as a surrogate end point gout using the BSES3 criteria.

DISCUSSION

It is important to emphasise that the evaluation of SU as a surrogate end point is for the context of using SU as an end point 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 the 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 on 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.

Author affiliations
[1]Musculoskeletal Statistics Unit, Department of Rheumatology, The Parker Institute, Copenhagen University Hospitals, Bispebjerg and Frederiksberg, Frederiksberg, Denmark
[2]Department of Rheumatology, Odense University Hospital, Svendborg, 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, 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, New South Wales, Australia

REFERENCES

1. Temple R. Clinical measurement in drug evaluation. In: Nimmo W, Tucker G, eds. *A regulatory authority's opinion about surrogate endpoints*. John Wiley & Sons Ltd, 1995:3–22.
2. De Gruttola VG, Clax P, DeMets D, *et al.* Considerations in the evaluation of surrogate endpoints in clinical trials. Summary of a National Institutes of Health workshop. *Control Clin Trials* 2001;22:485–502.
3. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharm Ther* 2001;69:89–95.
4. Schumacher HR, Taylor W, Edwards L, *et al.* Outcome domains for studies of acute and chronic gout. *J Rheumatol* 2009;36: 2342–5.
5. Lassere MN, Johnson KR, Boers M, *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;34:607–15.
6. Maksymowych WP, Landewé R, Tak PP, *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;36:1785–91.
7. Stamp LK, Zhu X, Dalbeth N, *et al.* 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 MN. 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. *Stat Methods Med Res* 2008;17:303–40.
9. Lassere M. The innovative evaluation schema of validating putative surrogate endpoints. *Joint Statistical Meeting*, Vancouver, Canada, 2010.
10. Lassere M, Johnson KR, Schiff M, *et al.* 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 Med Res Methodol* 2012;12:27.
11. Johnson KR, Liauw W, Lassere MN. Evaluating surrogacy metrics and investigating approval decisions of progression-free survival (PFS) in metastatic renal cell cancer: a systematic review. *Ann Oncol* 2015;26:485–96.
12. Jordan KM, Cameron JS, Snaith M, *et al*, British Society for Rheumatology and British Health Professionals in Rheumatology Standards, Guidelines and Audit Working Group (SGAWG). British Society for Rheumatology and British Health Professionals in Rheumatology guideline for the management of gout. *Rheumatology (Oxford)* 2007;46:1372–4.
13. Khanna D, Fitzgerald JD, Khanna PP, *et al.*, American College of Rheumatology. 2012 American College of Rheumatology Guidelines for the management of gout. Part 1: systematic nonpharmacologic and pharmacologic therapeutic approaches to hyperuricaemia. *Arthritis Care Res (Hoboken)* 2012;64:1431–46.
14. Zhang W, Doherty M, Bardin T, *et al.*, EULAR Standing Committee for International Clinical Studies Including Therapeutics. 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;65:1312–24.
15. Stevenson M, Pandor A. Febuxostat for the management of hyperuricaemia in patients with gout: a NICE single technology appraisal. *Pharmacoeconomics* 2011;29:133–40.
16. Shamseer L, Moher D, Clarke M, *et al.* Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015;349:g7647.
17. Liberati A, Altman DG, Tetzlaff 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:e1–34.
18. Wallace SL, Robinson H, Masi AT, *et al.* Preliminary criteria for the classification of the acute arthritis of primary gout. *Arthritis Rheum* 1977;20:895–900.
19. Higgins JP, Altman DG, Gøtzsche PC, *et al.* The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.
20. Guyatt G, Oxman AD, Sultan S, *et al.* GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. *J Clin Epidemiol* 2013;66:151–7.
21. Thompson SG, Higgins JP. Treating individuals 4: can meta-analysis help target interventions at individuals most likely to benefit? *Lancet* 2005;365:341–6.
22. Buyse M, Molenberghs G, Burzykowski T, *et al.* The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics* 2000;1:49–67.
23. Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. *Pharm Stat* 2006;5:173–86.
24. Johnson KR, Ringland C, Stokes BJ. 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 Oncol* 2006;7:741–6.
25. Johnson KR, Freemantle N, Anthony DM, *et al.* LDL-cholesterol differences predicted survival benefit in statin trials by the surrogate threshold effect (STE). *J Clin Epidemiol* 2009;62:328–36.
26. A'Hern RP, Ebbs SR, Baum MB. Does chemotherapy improve survival in advanced breast cancer? A statistical overview. *Br J Cancer* 1988;57:615–18.
27. http://www.sign.ac.uk/methodology/checklists.html