FEBUXOSTAT FOR THE MANAGEMENT OF HYPERURICAEMIA IN PATIENTS WITH GOUT

Produced by	School of Health and Related Research (ScHARR), The University of Sheffield
Authors	Matt Stevenson, Senior Operational Researcher, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA. Abdullah Pandor, Research Fellow, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA.
Correspondence to	Matt Stevenson, Senior Operational Researcher, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA.
Date completed	31 March 2008

Source of funding: This report was commissioned by the NIHR HTA Programme as project number HTA/07/69.

Declared competing interests of the authors: None.

Acknowledgements

Names and job titles of individuals who provided advice.

Dr Kelsey M Jordan, Consultant Rheumatologist, Brighton and Sussex University Hospitals NHS Trust.

Dr Michael L Snaith, Consultant Rheumatologist, Derbyshire Primary Care trust and Honorary Senior Lecturer, University of Sheffield.

We would also like to thank Gill Rooney, Project Administrator, ScHARR, for her help in preparing and formatting the report.

This report should be referenced as follows:

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

List of abbreviations

BSR	British Society for Rheumatology and British Health Professionals in Rheumatology
CEAC	Cost-Effectiveness Acceptability Curve
CI	Confidence Interval
CIC	Commercial-In-Confidence
ERG	Evidence Review Group
EULAR	European League Against Rheumatism
FDA	Food and Drug Administration
HRQOL	Health-Related Quality of Life
ITT	Intention-To-Treat
mg/d	milligrams per day
mg/dL	milligrams per decilitre
MS	Manufacturer's Submission
OMERACT -SIG	Outcome Measures In Rheumatology Clinical Trials - Special Interest Group
PSA	Probabilistic Sensitivity Analyses
QALY	Quality Adjusted Life Year
RCT	Randomised Controlled Trial
SD	Standard Deviation
SE	Standard Error
SPC	Summary of Product Characteristics
sUA	Serum Uric Acid
µmol/L	micromoles per litre
WMD	Weighted Mean Difference

1 SUMMARY

1.1 Scope of the submission

The manufacturer's submission (MS) generally reflects the scope of the appraisal issued by NICE, and is appropriate to the NHS. The majority of the MS reflects the use of febuxostat in adults with hyperuricaemia in whom urate deposition has already occurred (including a history, or presence of, tophus and/or gouty arthritis); however, it does not reflect the broader population outlined in the decision problem (adults unresponsive to, or intolerant of, allopurinol or with renal impairment). The intervention is defined as febuxostat for the management of hyperuricaemia in patients with gout. The MS considered allopurinol as the most relevant comparator, as reflected in the scope; however, no comparisons with alternative standard care (including sulphinpyrazone, benzbromarone, probenecid, or a combination of these medications) for adults unresponsive to, or intolerant of, allopurinol or with renal impairment were undertaken. The outcome measures identified in the scope were all relevant and included surrogate (serum uric acid levels, sUA) and clinical outcomes (gout flares, reduction in tophi size), tolerance and health related quality of life. The results provided are presented in terms of cost per quality adjusted life year (QALY) with a time horizon of two years within the base case, which is extended to five years within the sensitivity analyses, with the perspective of costs taken from an NHS and Personal Social Services perspective.

1.2 Summary of submitted clinical effectiveness evidence

- The main evidence in the submission is derived from two head-to-head, phase III, multi-arm, randomised, double blind, controlled trials (52 week FACT study and 28 week APEX trial) comparing the efficacy and safety of febuxostat with fixed dose allopurinol in patients with hyperuricaemia (sUA levels ≥8 mg/dL) and gout. Supplementary data from an ongoing long-term, open label extension study of the two head-to-head trials (EXCEL) are provided to assess the clinical efficacy of febuxostat over a period of two years.

- The pooled (not meta-analysis) clinical efficacy analysis of the FACT and APEX trials (representing 1689 patients) showed that febuxostat (80 mg/d and 120 mg/d) was significantly more effective than fixed dose allopurinol (300/100 mg/d) at reducing sUA levels to <6 mg/dL. However, a large percentage of patients on febuxostat did not achieve the primary endpoint and the fixed dose regimen employed for allopurinol patients may have introduced bias.

- Despite the significantly greater effect on sUA levels with febuxostat (including mean percentage reduction from baseline) than allopurinol, there were generally no

differences between treatments in more clinically important outcomes such as gout flares and tophi resolution (secondary endpoints).

- Post hoc sub-group analysis showed that febuxostat was more effective than allopurinol in decreasing sUA levels to less than 6 mg/dL in patients with baseline sUA concentrations less than 9 mg/dL, between 9 and 10 mg/dL, and in patients greater than 10 mg/dL. In addition, significantly more febuxostat recipients achieved lower sUA levels to therapeutic targets (<5 mg/dL) than fixed dose allopurinol. No subgroup analyses were conducted for patients with renal impairment, non responders to allopurinol or patients with severe disease.

- Results from the long-term extension period of the EXCEL study showed that more patients on febuxostat (80 mg/d and 120 mg/d) remained on initial treatment than fixed dose allopurinol (300/100 mg/d) after more than 24 months of follow-up and the number of tophi and gout flares were reduced over time. However, these data need to be interpreted with caution as the MS does not provide any statistical analysis (including event rates over time) or data on withdrawals due to gout flares, adverse events, or non response.

- Although the adverse event profile was similar in those receiving febuxostat compared to those receiving allopurinol, more febuxostat recipients discontinued treatment prematurely (the statistical analysis comparing the rates of discontinuation between the treatment groups were not reported in the MS or in the requested supplementary data; however, the primary published peer reviewed clinical paper for the FACT study reports that the rates of discontinuation were significantly higher in febuxostat recipients [$p<0.04$] than those receiving allopurinol). Reasons for withdrawal included gout flares and adverse events such as liver function test abnormalities.

1.3 Summary of submitted cost-effectiveness evidence

The submitted cost-effectiveness evidence reports an incremental cost per QALY of £16,324 (95% confidence interval, CI: 6,281 to 239,928) after two years of treatment. The incremental cost per QALY was below £20,000 in 63% of the simulations undertaken.

1.4 Commentary on the robustness of submitted evidence

1.4.1 Strengths

- The MS conducted a systematic search for clinical- and cost-effectiveness studies of febuxostat for the treatment of gout. It appears unlikely that any additional trials would have met the inclusion criteria had the search been widened to include other databases.

- The two identified trials, which represent the main clinical efficacy evidence, were of reasonable methodological quality (with some limitations), and measured a range of outcomes that are as appropriate and clinically relevant as possible.

- For a given specified comparison, that of two years of febuxostat treatment or allopurinol therapy to all patients, who are assumed to continue on that treatment for the entire study period, the report provides a rough estimate of the incremental cost per QALY.

1.4.2 Weaknesses

- The processes undertaken by the manufacturer for screening studies, data extraction and applying quality criteria to included studies are not explicitly clear in the MS. These factors limit the robustness of the systematic review.

- A simple pooled analysis of the patient level data, from the two head-to-head trials, was undertaken by the manufacturer. Although the methods for this type of data pooling were not explicitly described, the statistical approach for combining the data appears to be inappropriate as it fails to preserve randomisation and introduces bias and confounding. The resulting pooled data should therefore be treated with caution.

- The decision problem facing a funding agency is not that which was modelled in the MS. In the real-world, patients would be allowed to switch between treatments and some patients will discontinue therapy. The ERG recommended that a sequential treatment algorithm be modelled but this was not undertaken by the manufacturer.

- Despite acknowledging that some parameters in the model were incorrect, for example the price of allopurinol, or were not included in the probabilistic sensitivity analyses (PSA) when they should be, or that some parameters included in the PSA should have been excluded, such as the price of febuxostat, no further PSA runs were undertaken. As such, the incremental cost per QALY ratios provided in the MS will be subject to inaccuracy. In addition, as previously stated, the cost per QALY ratios

are those for a model structure that is not optimal, and incremental cost per QALYs for treatment algorithms are not presented.

1.4.3 Areas of uncertainty

- There is uncertainty around the clinical- and cost-effectiveness of febuxostat in comparison to other relevant treatments (including sulphinpyrazone, benzbromarone, probenecid, or a combination of these medications) for adults unresponsive to, or intolerant of, allopurinol or with renal impairment

- Long-term efficacy and safety data are limited on febuxostat

- There is uncertainty around the relationship between sUA levels and the expected number of gout flares.

- The incremental cost per QALY of sequential approaches of treatment is uncertain as these have not been modelled. The inclusion of sequential treatments is likely to produce a more cost-effective solution than allowing only one treatment for the duration of the model.

- There is uncertainty between the relationship between sUA levels and underlying patient utility.

1.5 Key issues

- The head-to-head trials presented in the MS directly compared febuxostat with fixed dose allopurinol. However, gout management guidelines and the allopurinol summary of product characteristics (SPC) generally recommend dose titration of allopurinol according to therapeutic targets (usual maintenance dose in mild conditions 100–200 mg/d, in moderately severe conditions 300–600 mg/d, in severe conditions 700–900 mg/d). Nevertheless, the MS and our clinical advisors suggest that dose escalation is rarely used by most clinicians in clinical practice.

- Whilst measures such as gout flares and tophi resolution were secondary outcomes, these are more clinically important. Randomised controlled trial (RCT) evidence shows that, even though more febuxostat recipients achieved the recommended biochemical goal (<6 mg/dL), this did not translate into an advantage over allopurinol in clinically important outcomes.

- The ERG has serious concerns regarding the cost-effectiveness analyses undertaken, primarily due to the model structure and the strategies compared. The analyses presented only compare two strategies, the provision of allopurinol for the entire time horizon and the provision of febuxostat for the entire time horizon. For both interventions, treatment was assumed to persist regardless of the patient response. In the clarification letter sent to the manufacturer the ERG suggested that the following strategies, where patients progress to the next intervention following lack of response, should be evaluated as a minimum. Allopurinol – Febuxostat – No Treatment; Febuxostat – Allopurinol – No-Treatment; Allopurinol – No Treatment and Febuxostat – No Treatment. Assumptions regarding the efficacies of second-line interventions would be necessary, with the base-case assuming that the effect on sUA levels and gout flares remains at the level seen in first-line treatments. These strategies allow the option for patients to be prescribed the cheaper, generic, intervention (allopurinol) and to only progress to the more expensive, non-generic treatment (febuxostat) if a lack of response was identified. Often such a strategy is estimated to be the most cost-effective approach. The inclusion of a no treatment arm is to determine whether any of the two active interventions are cost-effective, without this the more cost-effective treatment of the two treatments, where neither are cost-effective, could be selected.

- The ERG has serious concerns regarding the data selected to estimate the relationship between sUA levels and the number of gout flares expected. A large portion of the data that was collected to develop this linkage was excluded (accounting for 51% of all patients and 77% for UK patients) with less than robust reasoning.

- The ERG has additional serious concerns about the interpretation of the multivariate analyses. It is indicated that, there is no significant association between sUA levels and the number of gout flares reported within the data set used. This analysis has apparently been overlooked in favour of a bivariate analysis that does not include other confounders. Note that whilst no statistically significant relationship was found within this data set, this does not mean that such a relationship does not exist as indicated in clinical guidelines.

2. BACKGROUND

2.1 Critique of manufacturer's description of underlying health problem

The manufacturer's description of the underlying health problem is brief and fairly accurate. It states that gout is a common condition with a prevalence rate of 1.4% in the UK.[1,2] However, the manufacturer's discussion of context (p17 to 31, MS) does not indicate how many individuals are eligible for febuxostat treatment in England and Wales. Details of patients eligible for febuxostat treatment (based on percentage market share) are discussed and/or presented in the MS in the cost-effectiveness section (p.134, 137 to 138, MS). Although there are discrepancies between the estimated number of patients on allopurinol (Table 7.1 and Table 7.5, MS), the current treatment of choice, the MS estimates that between [redacted] patients in England and Wales would be eligible for febuxostat treatment in the first year, rising to between [redacted] patients at five years.

2.2 Critique of manufacturer's overview of current service provision

The manufacturer's overview of current service provision is adequate although some discussion around specific points is required.

The MS states (p18, MS) that drugs that lower serum uric acid (sUA) are prescribed to patients with gout to prevent recurrent flares and to prevent the deposition of monosodium urate crystals. Furthermore, the MS suggest (p19, MS) that several trials of urate lowering therapy have shown that an effective reduction of sUA levels reduces and ultimately prevents the occurrence of gout flares and that tophi can be reduced and may even disappear. The studies cited to support this do not appear to be controlled trials. Although some discussion has been provided (as additional information) by the manufacturer on the validity of sUA levels as a surrogate outcome for clinical endpoints, it is unclear what the strength of evidence (and or relationship) is linking sUA levels and clinical outcomes (e.g. gout recurrence, reduction in tophi).

The MS states (p17-18, MS) that the disease severity of gout can be described as patients progressing through four recognisable stages – asymptomatic hyperuricaemia, acute gout attack, intercritical gout and chronic tophaceous gout. The treatment for chronic gout cases with recurrent flares and tophi focuses on the cause of gout (including hyperuricaemia) and include lifestyle changes and urate lowering therapy. However, as noted in a recent Cochrane

review,[3] the evidence on when to start urate-lowering drug therapy is controversial. Some advocate that only patients who experience more than four episodes per year should be given such treatment, whereas others suggests that patients should be considered only after tophi development has been observed. A cost-effective analysis demonstrated that urate-lowering treatment is cost-effective in patients with one recurrent gouty attack per year and cost-saving in patients who have two or more attacks a year.[4] The recent British Society for Rheumatology and British Health Professionals in Rheumatology (BSR) guidelines recommend pharmacological treatment in patients with gouty arthritis and/or tophi, and in patients with uncomplicated gout if a second attack or further attacks occur within a year, as 40% of people will not experience another attack within the first year and lifestyle modifications can be effective.[5]

The MS widely refers to the guidelines published by the BSR and the European League Against Rheumatism (EULAR) on the management of gout for achieving target sUA levels. However, the MS does not make it explicitly clear that the targets recommended by the BSR (that is, sUA below 300 μmol/L [5 mg/dL] by pharmacological treatment in patients with gouty arthritis and/or tophi, and in patients with uncomplicated gout if a second attack or further attacks of gout occur within one year)[5] and EULAR (that is, a target serum uric acid level of ≤ 360 μmol/L [6 mg/dL] to eliminate gout flares and to resolve tophi)[6] are based on expert consensus agreement and have not been tested *a priori* by clinical trials.

The MS states that allopurinol is the most commonly used urate lowering drug in the UK (89% of gout treatments), with most cases (98%) using doses of 300 mg or less per day. Although the MS reports that allopurinol (300 mg/d) has limited efficacy in lowering sUA, many patients may need higher doses for optimal control of sUA levels.[5] The BSR and EULAR guidelines generally recommend initial long-term treatment of recurrent uncomplicated gout, which normally begins with low dose allopurinol (50–100 mg/d) to reduce the risk of adverse reactions and increased (titrated in increments of 50 to 100 mg every few weeks, adjusted if necessary for renal function, to achieve the targeted sUA levels) only if the serum urate response is unsatisfactory (maximum dose 900 mg/d).[5,6] Furthermore, the ERG notes that the allopurinol SPC[7] and the British National Formulary[8] also recommends dose titration (usual maintenance dose in mild conditions 100–200 mg/d, in moderately severe conditions 300–600 mg/d, in severe conditions 700–900 mg/d) according

to therapeutic targets. Despite these guidelines, the MS suggests that dose-escalation is rarely used in clinical practice.

Finally, the MS states (p6, MS) that no particular additional test or investigations are required with febuxostat therapy; however, this contradicts the febuxostat SPC (Appendix 1, MS and revised Annex I) as its states that theophylline levels should be monitored in patients starting febuxostat therapy.

3. CRITIQUE OF MANUFACTURER'S DEFINITION OF DECISION PROBLEM

A summary of the decision problem addressed by the MS is shown in Table 1.

Table 1: Decision problem as issued by NICE and addressed by the MS

	Final scope issued by NICE	Decision problem(s) addressed in the submission
Population	Adults with hyperuricaemia in whom urate deposition has already occurred (including a history or presence of tophus and gouty arthritis and /or nephrolithiasis)	Adults with hyperuricaemia in whom urate deposition has already occurred (including a history or presence of tophus, gouty arthritis).
Intervention	Febuxostat	Febuxostat
Comparator(s)	The standard comparators to be considered include: • Allopurinol • Alternative standard care (including sulphinpyrazone, benzbromarone, probenecid, or a combination of those) for adults unresponsive or intolerant to allopurinol • Allopurinol (dose adjusted according to glomerular filtration rate [GFR]), benzbromarone, or a combination of those for adults with renal impairment	The standard comparators considered included: • Allopurinol • Allopurinol for adults with renal impairment
Outcomes	The outcome measures to be considered include: • Serum uric acid levels • Gout flares • Reduction in tophi size • Tolerance • Health-related quality of life	The outcome measures considered included: • Serum uric acid levels • Gout flares • Reduction in tophi size • Tolerance • Health-related quality of life •

Economic Analysis	The reference case stipulates that the cost-effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life-year. The time horizon for the economic evaluation should be sufficiently long so as to incorporate all the important costs and benefits related to long-term therapy in this chronic condition. Costs will be considered from an NHS and Personal Social Services perspective.	Cost-effectiveness of treatments expressed in cost per quality – adjusted life-year over 2 years (based on 1-year trial data and a further 1-year extrapolation). Costs considered from NHS and Personal Social Service perspective
Special Considerations and Other Issues	If the evidence allows, the appraisal will consider: • Subgroups of patients for whom the technology is particularly appropriate due to greater clinical effectiveness or higher baseline risk (for example, subgroups related to risk factors, comorbidities or clinical features. Patients with sUA levels above 540 μmol/L (9 mg/dL), patients with tophi and patients with mild and moderate renal impairment.) • Patients intolerant of, or contraindicated to, allopurinol • Patients whose gout is unresponsive to allopurinol Guidance will be issued in accordance with the marketing authorisation.	The subgroup analysis showed the response rate of the primary efficacy endpoint increased with age, in female, in Caucasians (versus non-Caucasians), in improved renal function, in lower baseline sUA and presence of tophus. This increase of response rate by subgroup occurred in all treatment groups and does not change statistically significantly superiority of febuxostat 80 mg and 120 mg to allopurinol 300/100 mg. Therefore, no subgroup analyses were conducted as the size of subgroups did not allow for any power to detect any differences of treatments between the subgroups that were clinical relevant relative to differences of treatments provided by the full treatment groups.

3.1 Population

The manufacturer's statement of the decision problem appropriately defines the population as adults with hyperuricaemia in whom urate deposition has already occurred (including a history or presence of, tophus and/or gouty arthritis)

3.2 Intervention

Febuxostat is a novel, orally administered, nonpurine selective inhibitor of xanthine oxidase, which reduces sUA levels. Although febuxostat does not have a UK marketing authorisation (at the time of writing), the anticipated target indication is for the treatment of chronic

hyperuricaemia in conditions where urate deposition has already occurred (including a history, or presence of, tophus and/or gouty arthritis). The proposed course of treatment is continuous at a recommended dose of 80 mg/d without regard to food; if an additional urate lowering effect is required (sUA >360 μmol/L [6 mg/dL] after 2 to 4 weeks), the dose can be increased to 120 mg/d.

3.3 Comparators

The decision problem addressed in the MS states that the standard comparator considered included allopurinol, only. However, the NICE remit and final scope document (Table 1) states that comparisons should be made with (1) allopurinol; or (2) with alternative standard care (including sulphinpyrazone, benzbromarone, probenecid, or a combination of these medications) for adults unresponsive or intolerant to allopurinol; or (3) with allopurinol (dose adjusted according to glomerular filtration rate), benzbromarone, or a combination of these medications for adults with renal impairment.

The manufacturer's decision (p9 and p100, MS) to include allopurinol as the comparator was based on evidence from market research data, clinical opinion (not clear in the MS if expert clinical opinion was sought) and treatment recommendations by the BSR. This evidence suggests that allopurinol is considered to be the urate-lowering drug of choice and is the most commonly used regimen in the UK. The MS states that the alternative standard care of sulphinpyrazone, benzbromarone, probenecid, or a combination of these medications were not considered as comparators because these are rarely used in clinical practice (<3%) due to limitations in efficacy, safety profile and contraindications.

Although the ERG acknowledges allopurinol as the most potentially relevant comparator for all patients with hyperuricaemia and gout, it also considers sulphinpyrazone, benzbromarone (not licensed in the UK but used off-licence) and probenecid as potentailly relevant comparators, particularly in adults who are unresponsive or intolerant to allopurinol or with renal impairment. The use of these treatments are also advocated by the BSR guidelines, which recommend uricosuric agents (preferably sulphinpyrazone and benzbromarone) as second-line treatment options for chronic gout, in those producing and under-excreting a normal or reduced amount of urate, and in those resistant to or intolerant of allopurinol.[5]

3.4 Outcomes

The decision problem outlines five relevant outcomes to be assessed and all of these are stated to have been addressed in the MS. A summary of the measures used in the principal trials supporting the MS are shown in Table 2.

Table 2: Outcome measures in trials

Outcome in decision problem	Outcome measures used in principal trials providing supportive evidence in MS	
	FACT	APEX
Serum uric acid levels (sUA)	a) Proportion of patients in each treatment group whose last three sUA levels were < 360 µmol/L (6.0 mg/dL) (primary outcome) b) Proportion of patients with sUA levels < 360 µmol/L (6.0 mg/dL) at the final visit c) Percent reduction in sUA levels from baseline	a) Proportion of patients in each treatment group whose last three sUA levels were < 360 µmol/L (6.0 mg/dL) (primary outcome) b) Proportion of patients with sUA levels < 360 µmol/L (6.0 mg/dL) at the final visit c) Percent reduction in sUA levels from baseline
Gout flares	Proportion of patients requiring treatment for gout flare between weeks 8 and 52 (end study)	Proportion of patients requiring treatment for gout flare between weeks 8 and 28 (end of study)
Reduction in tophi size	In patients with primary palpable tophi at screening visit • Percent reduction in primary tophus size • The reduction in the total number of tophi	In patients with primary palpable tophi at screening visit • Percent reduction in primary tophus size • The reduction in the total number of tophi
Tolerance	Adverse events	Adverse events
Health related quality of life	Not reported	Not reported

Although the MS notes that all the clinical and laboratory procedures used are standard and generally accepted (p55, MS), the submission does not provide a detailed discussion of the most appropriate instruments for measuring each outcome and what their implications would be for affected individuals.

The ERG acknowledges that sUA levels are an appropriate surrogate (biochemical) outcome marker that should be measured using a specific enzymatic assay. The therapeutic target levels for lowering serum urate appear to be contentious. The MS states (p55-56, MS) that the BSR guidelines advocate sUA levels below 300 μmol/L (5 mg/dL)[5] whereas the EULAR guidelines recommend a target sUA of ≤ 360 μmol/L (6 mg/dL).[6] The rationale and strength of evidence supporting these targets is not appropriately discussed in the MS, nor is the relationship between sUA levels and clinical outcomes.

Gout flares are important clinical outcome measures as they are the most easily recognised concerns of patients. Measurement of gout flares in the FACT and APEX trials (proportion of patients requiring treatment) were recorded by individual investigators on a gout flare collection form (p56, MS). No details are provided in the MS on the criteria and features used to define a gout flare (including severity), criteria required for treatment, effect of response to treatments and details on monoarticular versus polyarticular flares.

The ERG notes that there are currently no standardised or validated tools available to measure tophus size as an outcome;[9] however, measurements of tophi may be made in several ways, including simple physical measurements of the nodules with a caliper or tape, magnetic resonance imaging, ultrasound or serial photographs. The MS suggests (p56, MS) that the FACT and APEX trials used simple physical measurements to identify any primary tophus. Although this manual method may be appropriate (good intra- and inter-observer reproducibility), it is unclear if this technique is reliable and accurate for measuring tophi, which commonly develops at the elbow and the joints of the hands and feet, at all locations. Data from the APEX trial (p65, MS) suggest that there is larger variability in measurements for elbow tophi, possibly due to olecranon bursal fluid.

The metric used for the evaluation of the interventions is cost per quality-adjusted life-year (QALY) gained, which is in accordance with the NICE reference case.

3.5 Time frame

The manufacturer's time horizon in the health economic model does not extend to a patients lifetime. This is defended in the submission (p101, MS) by stating that time horizons of greater duration would extrapolate beyond the length of the clinical trials; however, sensitivity analyses have extended the time horizon to five years, with little impact on cost per QALY (p128-129, MS). This minimal impact on cost-effectiveness with the change in time horizon is expected as each patient is assumed to have constant sUA levels (p123-124, MS) between months 4 to 24, which is assumed to persist to 60 months within the sensitivity analysis.

3.5 Other relevant factors

The statement of the decision problem proposes that, if evidence allows, the submission will consider subgroup of patients for whom febuxostat is particularly appropriate. The MS states (p7 and p99, MS) that no subgroup analyses were undertaken as the size of the subgroups lacked statistical power. However, the primary surrogate outcomes were analysed according to baseline sUA levels (<9 mg/dL, 9 to <10 mg/dL, ≥10 mg/dL) but these subgroups (p70 and p79, MS) were not powered to test for significant differences between groups. The MS states that subgroup analyses considered in the phase III trials and the long-term extension studies included patients with renal impairment, non responders to allopurinol 300 mg, patients with severe disease and patients defined by age and gender. A multivariate logistic regression showed that the treatment population was fairly homogeneous when defined by age, gender and disease severity but no comment was made with regards to people with renal impairment or people intolerant or unresponsive to allopurinol. As individual patient-level data is available to the manufacturer, the ERG proposes that subgroup analyses in people with renal impairment or people intolerant or unresponsive to allopurinol may have been useful.

4. CLINICAL EFFECTIVENESS

4.1 Critique of manufacturer's approach

4.1.1 Description of manufacturers search strategy and comment on whether the search strategy was appropriate.

The searches undertaken by the manufacturer to identify all relevant randomised controlled trials (RCTs) were conducted in December 2007. The search strategy utilises terms to identify the condition (gout), the intervention (febuxostat) and the type of evidence (RCTs, economic analyses). No language restrictions appear to have been applied. The strategy is simple but effective and the methodological filters used to identify types of evidence are representative of some of the best ones available. Only three databases were searched however (Pubmed, Embase and The Cochrane Library) so key data may have been missed, particularly regarding unpublished data (no research registers, such as the National Research Register or Current Controlled Trials, were searched). Other key databases overlooked include the Science Citation Index (Web of Science) and BIOSIS.

It is noted that the term allopurinol was omitted from the search strategy. Whilst this can be defended on the basis that the manufacturers were aware of head-to-head trials between febuxostat and allopurinol that are likely to be the most appropriate comparison, reference to previous allopurinol trials could provide re-assurance that the head-to-head trials did not, by chance, favour or disfavour allopurinol. Our clinical advisors have commented that the results for allopurinol appear to be lower than would have been expected from previous clinical trials.

4.2 Statement of the inclusion/exclusion criteria used in the study selection and comment on whether they were appropriate.

Details of the inclusion and exclusion criteria, as reported in the MS, is reproduced in Table 3 (p35, MS).

Table 3: Inclusion/exclusion criteria in the MS study selection

	Clinical effectiveness
Inclusion criteria	• Randomised phase II and phase III studies including the clinical effect of febuxostat on gout, compared to placebo or an active control
Exclusion criteria	• Non-randomised clinical studies, e.g. phase I studies on healthy volunteers • Preclinical studies

Although the inclusion/exclusion criteria appear to be (mostly) appropriate there appears to be some irregularities in the MS.

The statement of the decision problem proposes that the standard comparators to consider include alternative standard care (including sulphinpyrazone, benzbromarone, probenecid, or a combination of those) for adults unresponsive or intolerant to allopurinol. Although inclusion of studies that assess the clinical effect of febuxostat on gout compared to active controls are most appropriate, the MS has also considers no treatment (i.e. placebo) as an option for standard care. The ERG acknowledges that no treatment (i.e. placebo) may be a viable option for some adults, particularly for patients unresponsive or intolerant to allopurinol, and is an appropriate comparator.

The manufacturer's inclusion/exclusion criteria for the clinical evidence does not specify restrictions by length of follow up; however, the pooled analyses (not meta-analysis) conducted by the manufacturer (p73, MS) excluded a four week, phase II, randomised placebo controlled trial (TMX-00-004). In addition, the cost-effectiveness section (p94, MS) only included studies of at least 12 weeks duration to assess the clinical effect of febuxostat

on gout. The MS does not provide a reason for the different inclusion/exclusion criteria between the two sections nor does it provide an appropriate rationale for limiting studies by duration.

The ERG notes that a plausible rationale for excluding studies of less than 12 weeks duration may be because it will minimise tachyphylaxis effects, and as noted in the MS (p19), gout flares may be triggered during the first month after initiation of urate-lowering therapy (before the patients has adjusted to longer-term profile of lower uric acid levels) or after prophylaxis withdrawal (the FACT and APEX trials had eight-week prophylaxis). Due to these effects, the exclusion of short-term studies (less than 12 weeks duration), in both the clinical effectiveness and cost-effectiveness section, would seem reasonable. Based on these criteria, one four-week, phase II, double blind, randomised, placebo-controlled, dose-response study (TMX-00-004) has been inappropriately included in the clinical evidence section.

4.2.1 Table of identified studies. What studies were included in the submission and what were excluded.

The MS identified three pivotal RCTs. Of these, the FACT and APEX trials were direct head-to-head, phase III, randomised comparisons of febuxostat versus allopurinol, whereas the TMX-00-004 study was a phase II, dose-response, placebo controlled trial. Details of the study design and patient characteristics are summarised in Table 4. The MS also identified two ongoing open-label extension studies – EXCEL (extension of the FACT and APEX trials) and FOCUS (extension of TMX-00-004).

The manufacturer's initial submission did not contain a QUORUM flow diagram relating to any of the literature searches. Those provided subsequently are not complete and do not conform to the QUORUM statement flow diagram (www.consort-statement.org/QUOROM.pdf).

It is noteworthy that the MS has included a four week placebo controlled study (TMX-00-004) as a relevant RCT (p35-37, MS); however, it has subsequently been excluded (no specific reasons given in the MS, except that it was a four week trial) from the meta-analysis

section (section 5.5, p72-86, MS). As discussed in section 4.1.2, it may seem reasonable to exclude short-term studies of less than 12 weeks duration in the systematic review of the clinical evidence.

Table 4: Characteristics of studies

Study	Design	Participants	Interventions [a]	Outcomes	Duration (planned)
Randomised controlled trials (completed)					
FACT (C02-010)[10]	Phase III, multi-arm, multi-centre, randomised, double-blind, allopurinol-controlled, parallel-group trial (n=762) in USA and Canada	Adults (aged 18 to 85 years) with hyperuricaemia (>480 µmol/L [8.0 mg/dL]) and a history or presence of gout, defined as having 1 or more of the following: • Presence of characteristic urate crystals in the joint fluid and/or • Tophus proven to contain urate crystals and/or • Presence of at least six of the American Rheumatism Association criteria and/or • Renal function, defined as serum creatinine level ≤1.5 mg/dL and creatinine clearance ≥50 mL per minute	T1: Febuxostat, 80mg/d (n=257) T2: Febuxostat, 120mg/d (n=251) T3: Allopurinol, 300 mg/d (n=254)	Primary efficacy endpoint: • Proportion of patients whose last three sUA levels were <360 µmol/L (6.0 mg/dL) Secondary endpoints: • Proportion of patients with sUA levels <360 µmol/L (6.0 mg/dL) • Percent reduction in sUA levels from baseline • Proportion of patients requiring treatment for gout flare between weeks 8 and 52 (end of study) • In patients with primary palpable tophi at screening visit o Percent reduction in primary tophus size o The reduction in the total number of tophi	52 weeks
APEX (C02-009)[11]	Phase III, multi-arm, multi-centre, randomised, double-blind, allopurinol- and placebo	As above, but renal function defined as serum creatinine level <2.0 mg/dL and calculated creatinine clearance ≥20 mL per minute	T1: Febuxostat, 80 mg/d (n=267) T2: Febuxostat, 120 mg/d (n=269) T3: Febuxostat, 240 mg/d (n=134) T4: Allopurinol, 300/100 mg/d (n=268)[b] T5 : Placebo (n=134)	As above but proportion of patients requiring treatment for gout flare between weeks 8 and 28 (end of study)	28 weeks

	controlled, parallel-group trial (n=1072) in the USA				
TMX-00-004[12]	Phase II (dose response), multi-centre randomised double-blind, placebo controlled, parallel-group trial (n=153) in the USA	Adults (aged 18 to 85 years) with hyperuricaemia (>480 µmol/L [8.0 mg/dL]) who met the American College of Rheumatology preliminary criteria for classification of acute arthritis of primary gout (no further details provided in MS)	T1: Febuxostat, 40 mg/d (n=37) T2: Febuxostat, 80 mg/d (n=40) T3: Febuxostat, 120 mg/d (n=38) T4: Placebo (n=38)	Primary efficacy endpoint: • Proportion of patients with sUA levels <360 µmol/L (6.0 mg/dL) on day 28 Secondary endpoints: • Proportion of patients with sUA levels that had decreased to <360 µmol/L (6.0 mg/dL) on day 7, 14 and 21 • Percent reduction in sUA levels from baseline at each visit • Percent reduction in daily urinary uric acid excretion from baseline to day 28	4 weeks
Open label extension studies (ongoing)					
EXCEL [c] (C02-021)	Open label, randomised, active-controlled extension of FACT and APEX trials (n=735)	Subset of patients from the FACT and APEX trials (final visit in these trials was regarded as the first visit in the EXCEL trial)	T1: Febuxostat 80mg/d (n = 299) T2: Febuxostat 120 mg/d (n = 291) T3: Allopurinol 300/100 mg/d (n = 145) (Dose titration and switching was permitted after month 1 and before month 6. However, after 6 months, patients were regarded as	As FACT trial but secondary outcomes include: • Proportion of subjects whose sUA levels decreased to <6.0 mg/dL across treatment changes • Changes in Quality of Life based on the SF-36, the Gout Questionnaire, Health Assessment Questionnaire and the Minnesota	24 months

			therapeutic failure and were to discontinue treatment if their sUA levels remained at ≥6 mg/dL)[d]	Living with Heart Failure Questionnaire (Source: http://clinicaltrials.gov/ct2/show/NCT00175019)	
FOCUS (TMX-01-005)[13]	Open label, non-randomised, uncontrolled extension of TMX-00-004	Subset of patients from the TMX-00-004 trial	Febuxostat 80 mg/d (n=116) (Dose titration to lower [40mg/d] or higher doses [120 mg/d] was permitted between 4 and 28 weeks)	Primary efficacy endpoint: • Proportion of subjects whose sUA levels decreases to or is maintained at <6.0 mg/dL Secondary endpoints: • Percent reduction in sUA levels from baseline (Source: http://clinicaltrials.gov/ct2/show/NCT00174941)	5 years

sUA, serum uric acid

[a] In both the FACT and APEX trials, flare prophylactic treatment with naproxen 250 mg twice daily or colchicine 0.6 mg/d was given from day –14 or day 1 to the day before the week 8 visit. In the TMX-00-004 trials, flare prophylactic treatment with colchicine 0.6 mg/d was given from day –14 to the day before the day 14 visit. In the EXCEL study, flare prophylactic treatment with naproxen 250 mg twice daily or colchicine 0.6mg/d was given for 8 weeks

[b] Reduced dose of 100mg was used for 10 patients with renal impairment, defined as serum creatinine >1.5 but ≤2.0 mg/dL

[c] The EXCEL study was initially designed as an single-arm extension study of the FACT and APEX trials, where patients were only assigned to febuxostat 80 mg. However, the protocol was amended, as recommended by the US Food and Drug Administration, to be randomised and actively controlled to evaluate long-term efficacy

[d] In the febuxostat 80 mg/d group, patients could be titrated to 120 mg/d or switch therapy after month 1 and before month 6. In the febuxostat 120mg/d group, patients could be titrated to 80 mg/d or switch therapy after month 1 and before month 6. In the allopurinol 300/100 mg/d group, patients could switch therapy after month 1 and before month 6 (dose dependent on serum creatinine). After month 6 and end of treatment subject should be on stable dose; however, dosage or treatment changes were allowed with sponsor approval

4.2.2 Details of any relevant studies that were not included in the submission ?

We have not been able, within the time available, to undertake full searches to identify all potentially relevant studies not identified by the manufacturer's search strategy. However, the ERG is confident that all relevant studies were included in the MS and details of ongoing trials that are likely to be reporting additional evidence within 12 months were reported.

4.2.3 Description and critique of manufacturers approach to validity assessment

The validity assessment tool used in the MS is not referenced but the questions are adequate. The completed validity assessment tool for the three pivotal trials, as reported in the MS, is reproduced in Table 5. The ERG acknowledges the validity assessment tool used in the MS was appropriate; however, some further discussion around specific points is required.

Table 5. Validity assessment of completed trials included by the manufacturer

Validity assessment	Trials		
	FACT	APEX	TMX-00-004
How was allocation concealed?	By interactive voice response system		
What randomisation technique was used?	Computer generated		
Was a justification of the sample size provided?	Yes		
Was follow-up adequate?	Yes		
Were the individuals undertaking the outcomes assessment aware of allocation?	No		
Was the design parallel-group or crossover? Indicate for each crossover trial whether a carry-over effect is likely.	Parallel-group		
Was the RCT conducted in the UK (or were one or more centres of the multinational RCT located in the UK)? If not, where was the RCT conducted, and is clinical practice likely to differ from UK practice?	No. The trials were conducted in the USA (FACT in Canada as well). The clinical practice for gout treatment, e.g., using allopurinol and the target sUA level, was similar as that in the UK. However, in the UK, 2007 BSR guidelines recommend a more stringent target sUA level of 300 μmol/L (5 mg/dL).		
How do the included in the RCT participants compare with patients who are likely to receive the intervention in the UK? Consider factors known to affect outcomes in the main indication, such as demographics, epidemiology, disease severity, and setting.	The RCT participants are representative overall of UK patients with gout and sUA ≥480 μmol/L (8 mg/dL) regarding clinical setting in primary care, demographics, co-morbidities and disease severity, except that the RCT participants had a higher rate of obesity and renal impairment than UK patients in general.		
For pharmaceuticals, what dosage regimens were used in the RCT? Are they within those detailed in the Summary of Product Characteristics?	Allopurinol 300 mg (100 mg for renally impaired) was used as an active comparator. The dose used is the recommended dose for moderate gout in the SPC and the dose of ≤ 300 mg per day is used in the majority (97.9%) of the GP patients in the UK treated with allopurinol.		
Were the study groups comparable?	Yes		
Were the statistical analyses used appropriate?	Yes		
Was an intention-to-treat analysis undertaken?	Yes		
Were there any confounding factors that may attenuate the interpretation of the results of the RCT(s)?	No. There were no clinically relevant differences in baseline characteristics between the febuxostat-treatment group and the allopurinol-treatment group.		

RCT, randomised controlled trial; SPC, Summary of Product Characteristics

The MS states that the sample size calculations in the FACT trial assumed a true response rate of 60% for the allopurinol treatment group to reach the primary endpoint; however, only 21% achieved this in the trial. No respective data were reported for the APEX trial. In addition, it appears that patient's that may be loss to follow up were not considered in the sample size calculations. As a result, the sample size calculations in the FACT trial (and probably the APEX trial) may have been underestimated.

In the FACT and APEX trials, allopurinol 300 mg/d (100 mg/d for renal impaired patients) was used as an active comparator. Although the MS states that the dose used was the recommended dose in the SPC for moderate gout, the ERG notes that this appears to be incorrect. According to the allopurinol SPC[7] dose titration is recommended (usual maintenance dose in mild conditions 100–200 mg/d, in moderately severe conditions 300–600 mg/d and in severe conditions 700–900 mg/d) according to therapeutic targets.

Although the MS states that all studies (FACT, APEX and TMX-00-004) were double blind, it is unclear from the evidence provided in the MS, whether investigators who administered the intervention were blinded to the treatment or if outcome assessors were blinded to the treatment allocation. In addition, the MS does not report if any of these studies assessed the success of blinding.

The MS states that the no clinically relevant differences were observed across treatment groups between the FACT and APEX trials in any demographic or baseline characteristics (no data reported for the TMX-00-004 trial). However, in the FACT trial, 35% had mild to moderate renal insufficiency and 44% had previous urate lowering therapy. In contrast no data were reported for these parameters in the APEX trial and the definition for renal impairment was different between the two trials. As individual patient data were available to the manufacturer, the provision of this information would have been useful.

The MS states that the participants in the included trials were similar to the UK population (except higher rate of obesity and renal impairment in RCT participants than UK patients in

general). The ERG notes that the mean age of the participants in the FACT and APEX trials was 52 years (data not reported for the TMX-00-004 trials); however, according to current UK data (from 2000 to 2005), the mean age of patients with gout is 65.59 years.[1]

The MS states that the mean compliance rate (determined by pill count) ranged from 95.0% to 97.8% across the treatment groups in the FACT and APEX trials (data not reported for TMX-00-004 study). In general, the validity of a study may be threatened if attrition is more than 20%. In the FACT and APEX trials, 33% and 28% of patients prematurely discontinued treatment, respectively. However, all withdrawals were accounted for and an intention to treat (ITT) analysis was undertaken.

Ideally in an ITT analysis participants should be analysed in the groups to which they were randomised regardless of which (or how much) treatment they actually received, and regardless of other protocol irregularities, such as ineligibility. The MS states that all primary and secondary efficacy analyses for the FACT and APEX trials (no information provided on the TMX-00-004 trial) were performed on the ITT population, except for the secondary efficacy analyses for the percent reduction in primary tophus size and the reduction in the total number of tophi. The ITT population was defined as all randomised subjects who received at least one dose of study drug and who had sUA levels ≥480 μmol/L (8.0 mg/dL) or greater at day –2 as determined by the central laboratory. Although the post-randomisation inclusions were pre-specified, the ERG acknowledges that the removal of ineligible patients (with sUA levels <480 μmol/L at day –2) from both study arms who received treatment after randomisation may be acceptable and will lead to an unbiased assessment of treatment effect in patients who do meet the inclusion criteria.[14,15]

4.2.4 Description and critique of manufacturers outcome selection

The main outcome measures selected by the manufacturer are summarised in Table 6.

Table 6: Manufacturer's main outcome selection

Primary efficacy endpoint • The proportion of subjects whose last 3 sUA levels were < 360 μmol/L (6.0 mg/dL)
Secondary efficacy endpoints • The proportion of subjects whose sUA levels were < 360 μmol/L (6.0 mg/dL) at final visit (defined as last visit of a specific subject) • The percent reduction in sUA levels from baseline • The percent reduction in primary tophus size, as determined by physical measurement in the subset of subjects with a primary palpable tophus at the screening visit • The reduction in the total number of tophi in the subset of subjects with palpable tophi at the screening visit • The proportion of subjects requiring treatment for a gout flare • Adverse events

sUA, serum uric acid

While no validated outcome measures have been defined for chronic gout,[3] a consensus has been reached among experts through the OMERACT (Outcome Measures In Rheumatology Clinical Trials) process. The Special Interest Group (SIG) for gout recommend the following domains as mandatory measures in chronic gout studies: serum urate, gout flares, tophus regression, health-related quality of life (HRQOL), patient global assessment, physical function (functional disability and joint impairment), joint damage imaging, participation (life-role) and safety and tolerability.[16] Although the MS does not consider all the outcomes recommended by OMERACT SIG; the ERG considers the manufacturer's outcome selection (a further critique on the appropriateness of these outcomes is discussed in section 3.4.) to be relevant and appropriate.

4.2.5 Describe and critique the statistical approach used

The manufacturer did not undertake a meta-analysis. The MS states that no meta-analysis was considered necessary as patient level data from pooled head-to-head RCTs was available which provided high level evidence of efficacy and safety (p73, MS). Despite the notable differences (such as length of study, definition of renal function, intervention sites (country), inclusion of a placebo and febuxostat 240mg/d group, and the use of lower doses of allopurinol based on renal function - see table 4 for further details) between the studies, the rationale for presenting and pooling individual patient data from the FACT and APEX trials (provided as additional information when requested) was primarily based on the similarity of

the design and patient selection criteria of the two head-to-head trials; however, the limitations and validity of this methodology was not discussed.

The statistical methods of pooling data were not explicit in the MS; however, it appears that pooling consisted of adding the number of events observed in a given treatment group across the trials and dividing the results by the total number of patients included in that group. For binary measures data were presented as difference in proportions (whilst a valid measure, the absolute or risk difference is naturally constrained as it is least likely to be consistent across a set of trials as it does not account for the underlying risk. Its use is problematic when it is applied to other patient groups with widely ranging expected risks and settings, as treatment benefit often relates to baseline risk) and for continuous measures data were presented as mean percentage change from baseline (statistical comparisons were undertaken between groups; however, the absolute mean differences between treatment groups were not reported in the MS). Although the MS states (p7 and p99, MS) that no subgroup analyses were undertaken as the size of the subgroups lacked statistical power, the primary clinical outcomes were analysed according to baseline sUA levels (<9 mg/dL, 9 to <10 mg/dL, ≥10 mg/dL).

The ERG notes that the pooling of data is considered as inadequate for the assessment of efficacy.[17,18,19] A pooled analysis focuses on treatment groups rather than on studies and ignores validity of the comparisons and is subject to bias termed 'Simpsons paradox in probability'.[17,18,19] A more satisfactory statistical technique involves combining the results from two or more separate studies in a meta-analysis.[20,17,18,19] Although the ERG requested the manufacturer to provide a meta-analysis (using individual patient level or aggregate data for all outcomes of interest) of the safety and efficacy evidence in accordance to the NICE guidance,[21] it was not forthcoming.

4.2.6 Summary statement

Although the majority of the MS reflects the anticipated licensed indication for the use of febuxostat, in adults with hyperuricaemia in whom urate deposition has already occurred (including a history, or presence of, tophus and/or gouty arthritis), it does not reflect the broader population outlined in the decision problem (adults unresponsive to or intolerant of allopurinol or with renal impairment).

The manufacturer's search strategy was adequately reported and the submission appears to contain all the relevant head-to-head RCTs of febuxostat versus allopurinol. Although allopurinol is the most potentially relevant comparator, the fixed dose regimen employed in the trials may be an issue of concern. No considerations or comparisons with alternative standard care (including sulphinpyrazone, benzbromarone, probenecid, or a combination of these medications) for adults unresponsive to or intolerant of allopurinol or with renal impairment were undertaken. The validity assessment tool used was satisfactory and outcomes selections were relevant and appropriate. Statistical methods were not explicitly described and the statistical approach for combining data appears to be inappropriate.

The submission also draws on ongoing open label extension studies to assess the long-term efficacy of febuxostat in patients with gout; however, due to the design and reporting of these studies, they are difficult to interpret.

4.3 Summary of submitted evidence

Three pivotal RCTs were identified in the manufacturer's systematic review. The FACT study was a phase III, three-arm, randomised, double-blind, 52 week multi-centre trial which compared the efficacy and safety of febuxostat (80 mg/d or 120 mg/d) with fixed dose allopurinol (300 mg/d) in 762 patients with hyperuricaemia (sUA levels ≥8 mg/dL [480 μmol/L]) and gout. The APEX trial was a phase III, five-arm, randomised, double-blind, 28 week multi-centre trial that compared the efficacy and safety of febuxostat (80 mg/d or 120 mg/d or 240 [safety dose] mg/d) with fixed dose (based on serum creatinine levels) allopurinol 300/100 mg/d and placebo in 1072 patients with hyperuricaemia (sUA levels ≥8 mg/dL [480 μmol/L]) and gout. In both these trials, prophylactic colchicine (0.6 mg/d) or naproxen (250 mg twice daily) was administered to all patients for two weeks prior to randomisation and for the first eight weeks of double-blind treatment. The TMX-00-004 trials was a phase II, dose-response (four-arm), randomised, double-blind, four week multi-centre trial that compared the efficacy of febuxostat (40 mg/d or 80 mg/d or 120 mg/d) with placebo in 153 patients with hyperuricaemia (sUA levels ≥8 mg/dL [480 μmol/L]) and gout. Prophylactic treatment involved colchicine (0.6 mg/d) 14 days prior to and post randomisation.

As individual patient level data were available to the manufacturer, a pooled analysis (not a meta-analysis) of the FACT and APEX trials was undertaken. Data from the TMX-00-004 trial were not included in the pooled analysis because it only included data for four weeks of treatment (p73, MS). No other reasons were given by the manufacturer. The pooled clinical efficacy analysis was based on an ITT population (patients who were randomised and received at least one dose of the study drug, with sUA levels ≥8 mg/dL at baseline) which included 1689 subjects (six patients in the FACT study and five patients in the APEX trial who were allocated to interventions did not meet the eligibility criteria and data from the febuxostat 240 mg/d [n=134] group from the APEX trial were excluded because the dose was only included in the trial to evaluate the safety, at twice the highest dose proposed for licensed indication, at the request of the US Food and Drug Administration, FDA).

Supplementary data from two ongoing open label extension studies were also provided by the manufacturer. A subset of patients (n=735) completing the double blind phase of the FACT and APEX trials were allowed to enter an amended, randomised, open-label extension study (EXCEL) to evaluate long-term (24 months) efficacy. These patients initially received febuxostat 80 mg/d or 120 mg/d for four weeks, after which the dose could be up-titrated to 120 mg/d or down-titrated to 80 mg/d or switched to alternate therapy based on maintaining sUA levels between 180 µmol/L (3 mg/dL) and 360 µmol/L (6 mg/dL). For the allopurinol group, patients initially received fixed dose (based on serum creatinine levels) allopurinol (300/100 mg/d), after which the therapy could be switched to alternate treatment. Dose stabilisation was to be achieved by six months. After six months, patients were regarded as therapeutic failures and were to discontinue treatment if their sUA levels remained at ≥6 mg/dL. Thus, only patients who achieved sUA levels of less than 6 mg/dL continued to participate in the study long-term. Colchicine (0.6 mg/d) or naproxen (250 mg twice daily) was provided as prophylaxis for the first eight weeks of treatment.

Patients who completed the TMX-00-004 trial (n=116) were enrolled in a non-randomised open label extension study (FOCUS) to evaluate long-term (five years) efficacy. These patients initially received febuxostat 80 mg/d for four weeks, after which the dose could be titrated to 40 mg/d or 120 mg/d, if needed. Dose stabilisation was to be achieved by week 28. No further details were reported in the MS.

4.3.1 Summary of results

This section presents the main clinical efficacy evidence, as reported in the MS, from a pooled analysis of two head-to-head studies (FACT and APEX trials). Supplementary data from an ongoing long-term, open label extension study of the two head to head trials (EXCEL) are also presented. Although the manufacturer provided a list of all the references, the manufacturer did not provide the ERG group a full copy of original papers, as required by NICE guidance.[21]

Efficacy

In the MS, it is difficult to compare the results of the pooled analysis as these have not been tabulated (as requested) or well described for each outcome of interest (p72-86, MS). A tabulated summary of such data is presented in Table 7 and 8. Additional information, not reported in the MS, was provided by the manufacturer in the clarifications of questions raised by the ERG.

The pooled analysis suggest that febuxostat 80 mg/d and 120 mg/d is significantly more effective than fixed dose allopurinol (300/100 mg/d) at reducing surrogate endpoints (sUA levels) either to target levels or from baseline ($p \leq 0.05$ for all comparisons). No statistically significant changes were observed with febuxostat 80 mg/d compared to allopurinol (300/100 mg/d) in the proportion of patients requiring treatment for gout flares during (week 1 to 8) or after (week 9 to 52) prophylaxis; however, allopurinol (300/100 mg/d) significantly reduced gout flares compared with febuxostat 120 mg/d during and after prophylaxis. No statistically significant differences were found between groups in the percentage reduction in tophus area, except at week 28 (significantly greater median percentage reductions were observed in the primary tophus size from baseline with febuxostat 120 mg/d compared with allopurinol).

Table 7: Pooled analysis of febuxostat 80 mg/d versus fixed dose allopurinol (300/100 mg/d)

Outcomes	Febuxostat (80 mg/d)	Allopurinol (300/100 mg/d)	Absolute difference	97.5%CI	p-value
Primary efficacy endpoint					
sUA levels <6.0 mg/dL at last 3 visits	262/517 (51%)	113/519 (22%)	29%	22.5,35.3	<0.001
Secondary efficacy endpoints					
sUA levels <6.0 mg/dL at final visit [a,b]	368/502 (73%)	190/505 (38%)	NR	NR	≤0.05
Reduction in number of tophi	NR	NR	-	-	-
Recurrent gout flares needing treatment					
Day 1 to week 8 (prophylaxis)	128/517 (25%)	113/519 (22%)	NR	NR	NS
Week 9 to 52	268/451 (59%)	260/471 (55%)	NR	NR	NS
Day 1 to week 52	312/517 (60%)	299/519 (58%)	NR	NR	NS
Percentage change from baseline					
sUA levels at final visit [a] (Mean ±SD)	-44.98 ±18.61	-33.36 ±15.02	NR	-	≤0.05
Tophus size (median)					
Week 28	-34.7	-28.6	NR	-	NS
Week 52	-83.4	-49.7	NR	-	NS
Final visit	-43.9	-25.0	NR	-	NS
Subgroup analysis					
sUA levels <6 mg/dL at final visit [a,b] by baseline sUA					
<9 mg/dL	114/132 (86%)	79/140 (56%)	NR	NR	≤0.05
9 to <10 mg/dL	122/163 (75%)	70/173 (40%)	NR	NR	≤0.05
≥10 mg/dL	132/207 (64%)	41/192 (21%)	NR	NR	≤0.05
sUA levels at final visit [a,b]					
<5.0 mg/dL	234/502 (47%)	65/505 (13%)	NR	NR	≤0.05
<4.0 mg/dL	96/502 (19%)	10/505 (2%)	NR	NR	≤0.05

sUA, serum uric acid; NR, not reported; SD, standard deviation; CI, confidence interval

[a] Defined as last visit of a specific subject

[b] Available case analysis (not ITT analysis)

Table 8: Pooled analysis of febuxostat 120 mg/d versus fixed dose allopurinol (300/100 mg/d)

Outcomes	Febuxostat (120 mg/d)	Allopurinol (300/100 mg/d)	Absolute difference	97.5%CI	p-value
Primary efficacy endpoint					
sUA levels <6.0 mg/dL at last 3 visits	329/519 (63%)	113/519 (22%)	42%	35.4, 47.9	<0.001
Secondary efficacy endpoints					
sUA levels <6.0 mg/dL at final visit [a,b]	402/507 (79%)	190/505 (38%)	NR	NR	≤0.05
Reduction in number of tophi	NR	NR	-	-	-
Recurrent gout flares needing treatment					
Day 1 to week 8 (prophylaxis)	187/519 (36%)	113/519 (22%)	NR	NR	≤0.05
Week 9 to 52	279/455 (61%)	260/471 (55%)	NR	NR	≤0.05
Day 1 to week 52	348/519 (67%)	299/519 (58%)	NR	NR	≤0.05
Percentage change from baseline					
sUA levels at final visit [a] (Mean ±SD))	-51.71 ±18.91	-33.36 ±15.02	NR	-	≤0.05
Tophus size (median)					
Week 28	-52.7	-28.6	NR	-	≤0.05
Week 52	-65.5	-49.7	NR	-	NS
Final visit	-43.8	-25.0	NR	-	NS
Subgroup analysis					
sUA levels <6 mg/dL at final visit [a,b] by baseline sUA					
<9 mg/dL	124/141 (88%)	79/140 (56%)	NR	NR	≤0.05
9 to <10 mg/dL	139/157 (89%)	70/173 (40%)	NR	NR	≤0.05
≥10 mg/dL	139/209 (67%)	41/192 (21%)	NR	NR	≤0.05
sUA levels at final visit [a,b]					
<5.0 mg/dL	331/507 (47%)	65/505 (13%)	NR	NR	≤0.05
<4.0 mg/dL	200/507 (39%)	10/505 (2%)	NR	NR	≤0.05

sUA, serum uric acid; NR, not reported; SD, standard deviation; CI, confidence interval

[a] Defined as last visit of a specific subject

[b] Available case analysis (not ITT analysis)

Further evidence of the clinical effectiveness of febuxostat was provided in the MS from an open label extension study of the FACT and APEX trial (EXCEL). Data could not be tabulated due to the poor and restrictive reporting of the limited results, thus a narrative summary, as reported in the MS, is provided. However, data should be interpreted with caution.

The EXCEL study (p109, MS) found that a large number of patients on febuxostat remained on initial treatment after more than 24 months of treatments (76% [n=299] febuxostat 80 mg/d, 71% [n=291] febuxostat 120 mg/d, and 40% [n=145] allopurinol). However, contradictory data are reported on page 83 of the MS which suggest low persistence rates for all interventions (35% febuxostat 80 mg/d, 10% febuxostat 120 mg/d and 5% allopurinol). Nevertheless, the percentage of patients with 100% resolution of tophi with the initial treatment assignments were 38%, 36% and 17% for febuxostat 80 mg/d, febuxostat 120 mg/d and allopurinol 300/100mg/d, respectively. The percentages of patients with at least a 50% reduction in primary tophus size were 65%, 71% and 57% for febuxostat 80 mg/d, febuxostat 120 mg/d and allopurinol 300/100 mg/d, respectively. For each year on febuxostat treatment, the number of gout flares decreased over time. As no statistical comparisons were reported in the MS, it is not clear if the findings were statistically different between treatment groups.

Critique of efficacy data reported

The methodological limitations of the pooled analysis, undertaken by the manufacturer, are described and critiqued in section 4.1.7. Nevertheless, the main issue that limits the robustness of the data reported in the MS is the use of a fixed dose allopurinol comparator in the FACT and APEX trials. Gout management guidelines[5,6] and the allopurinol SPC[7] generally recommend dose titration of allopurinol according to therapeutic targets (usual maintenance dose in mild conditions 100–200 mg/d, in moderately severe conditions 300–600 mg/d, in severe conditions 700–900 mg/d). Therefore, it is possible that allopurinol would have been more effective at lowering sUA levels if the dose had been titrated, but as noted by the manufacturer and the clinical advisors, dose escalation is rarely used by most clinicians in clinical practice.

Although febuxostat was more effective than fixed dose allopurinol at reducing sUA levels, a large percentage of patients on febuxostat (80 mg/d and 120 mg/d) did not achieve the primary endpoint (49% and 37%, respectively). While efficacy are lacking, it needs to established if higher febuxostat doses can safely achieve this outcome.

Prophylaxis treatment in the two head-to-head trials included the use of colchicine or naproxen against acute gout for only eight weeks after initiation of treatment with febuxostat or allopurinol. This appears to be unusually short and not in line with current practice. The BSR management guidelines recommend flare-prophylactic treatment with colchicine for up to six months following initiation of long-term treatment with urate-lowering therapy.[5] A longer prophylaxis, rather than the eight week period adopted in the clinical studies, will probably decrease the incidence of gout flares in the first few months of therapy.

The MS states (p7 and p99, MS) that no subgroup analyses were undertaken as the size of the subgroups lacked statistical power. However, the primary outcomes were analysed according to baseline sUA levels (<9 mg/dL, 9 to <10 mg/dL, ≥10 mg/dL) and the proportion of patients achieving BSR guideline targets (sUA <5 mg/dL) was also reported.

The open label extension study (EXCEL) evaluated topus size and gout flares in patients maintaining sUA levels between 180 µmol/L (3 mg/dL) and 360 µmol/L (6 mg/dL). As patients were allowed to change dose and treatments, results for gout flares and topus size were analysed according to stable treatments. The available data are difficult to interpret as events rates (n/N) are not reported (over time) and no statistical analysis is provided. It is not known to what extent the results are affected by the withdrawal of patients to drug induced flares, adverse effects, or non response. In addition, the results for number of patients with complete tophi resolution (present in 19.7% of patients at baseline) are very limited to draw firm conclusions.

Data checking the MS highlighted that not all results in the MS were from the ITT population. For example, the data for the proportion of patients with sUA levels <6 mg/dL at final visit, sUA levels <6 mg/dL at final visit by baseline sUA and sUA levels <5 mg/dL at final visit appear to be based on an available case analysis. Other inconsistencies included the number

of events in the manufacturers pooled analysis (febuxostat 80 mg/d group, n=269 and febuxostat 120mg/d group, n=348) are different to the combined individual rates (n= 268 and n=347, respectively) for gout flares between week 9 to end of study (p12 and p14, manufacturer's response to ERG queries). The time intervals also appear to be incorrect for example, day 1 to week 8 and week 8 to week 52 etc. The ERG assumes this should read day 1 to week 8 and week 9 to week 52 etc. Moreover, no data for statistical comparisons were provided for the pooled results (p12, manufacturer's response to ERG queries) on gout flares.

Safety and tolerability

The MS reports safety and tolerability data from the FACT and APEX trials. No additional safety data were reported from the open label extension study.

A summary of the rates of discontinuation, including reasons for premature termination are presented in Table 9. Although the rates of discontinuation in the FACT and APEX trials were higher in both the febuxostat groups than in the allopurinol group, the statistical analysis comparing the rates of discontinuation between the treatment groups were not reported in the MS or in the requested supplementary data. The primary published peer reviewed clinical paper for the FACT study reports that the rates of discontinuation were significantly higher in both the 80 mg/d febuxostat group and the 120 mg/d febuxostat group than in the allopurinol group (p=0.04 and p=0.002, respectively). The most common adverse event leading to withdrawal was liver function test abnormalities, with more patients receiving febuxostat 120 mg/d (n=7), and 80 mg/d (n=5) discontinuing as compared to allopurinol (n=1). Four patients in each of the febuxostat groups discontinued treatment because of rashes (mostly localised and transient maculopapular) compared to allopurinol (n=1; p=0.04 vs. febuxostat 120 mg/d)[10] Although requested, no equivalent data were provided by the manufacturer for the APEX trial, except that a greater number of febuxostat recipients ([redacted] subjects in total for febuxostat 80 mg/d and 120 mg/d) than allopurinol ([redacted]) discontinued due to rash.

Table 9: Number (%) of patients discontinuing treatment in the FACT and APEX trials (Data derived from MS, p44, p47 and p73)

	Interventions and trial (n)								
	Febuxostat 80 mg/d			Febuxostat 120 mg/d			Allopurinol 300/100 mg/d		
	FACT	APEX	All	FACT	APEX	All	FACT	APEX	All
Subjects randomised (received ≥1 dose of study drug)	256	267	523	251	269	520	253	268	521
Subjects who completed the study	168 (66%)	174 (65%)	342 (65%)	153 (61%)	200 (74%)	353 (68%)	187 (74%)	211 (79%)	398 (76%)
Primary reason for premature termination:	88 (34%)	93 (35%)	181 (35%)	98 (39%)	69 (26%)	167 (32%)	66 (26%)	57 (21%)	123 (24%)
Lost to follow-up	25 (28%)	19 (20%)	44 (24%)	18 (18%)	17 (25%)	35 (21%)	21 (32%)	17 (30%)	38 (31%)
Adverse event	16 (18%)	18 (19%)	34 (19%)	23 (23%)	16 (23%)	39 (23%)	8 (12%)	18 (32%)	26 (21%)
Gout flare [a]	10 (11%)	16 (17%)	26 (14%)	28 (29%)	16 (23%)	44 (26%)	9 (14%)	9 (16%)	18 (15%)
Personal reason [a]	19 (22%)	15 (16%)	34 (19%)	13 (13%)	8 (12%)	21 (13%)	13 (20%)	5 (9%)	18 (15%)
Other [a]	11 (13%)	13 (14%)	24 (13%)	14 (14%)	6 (9%)	20 (12%)	14 (21%)	1 (2%)	15 (12%)
Protocol violation	7 (8%)	6 (6%)	13 (7%)	2 (2%)	3 (4%)	5 (3%)	1 (2%)	6 (11%)	7 (6%)
Therapeutic failures	NR	6 (6%)	6 (3%)	NR	3 (4%)	3 (2%)	NR	1 (2%)	1 (1%)

NR, not reported

[a] Individual data reported for the APEX trial (p44, MS) are not in agreement with the pooled data reported on p73, MS (when excluding data from FACT trial). Moreover, there are also discrepancies between the initial MS and the manufacturer's response to ERG queries (e.g. number of patients that discontinued treatment due to gout flare whilst on placebo, febuxostat 80 mg/d, febuxostat 120 mg/d and allopurinol 300/100 mg/d were 0 (0%), 13 (14.0%), 6 (9.7%); and 1 (1.8%), respectively; personal reason: 9 (27.3%), 16 (17.2%), 16 (23.2%); and 9 (15.8%), respectively; and other reasons: 3 (9.1%), 15 (16.1%), 8 (11.6%); and 5 (9.8%), respectively).[CIC]

In the phase III clinical trials evaluating febuxostat in doses of 80 mg/d, 120 mg/d and 240 mg/d (representing two times the maximum clinical dose), the most common adverse events reported were upper respiratory tract infection, nasopharyngitis, diarrhoea, arthralgia, headache, pain in an extremity, influenza, back pain, nausea, hypertension, and alanine aminotransferase and aspartate aminotransferase increases. A summary of the pooled (data not reported separately for each study) treatment related adverse events, as reported in the MS, is presented in Table 10. In general, the number of adverse events appears to be similar across treatment groups (no statistical analysis reported in the MS).

Table 10: Pooled (FACT and APEX trials) treatment-related adverse events occurring in 2% or more of patients in any treatment group

	Treatment group (n)				
	Febuxostat				Allopurinol
	80 mg/d (n = 523)	120 mg/d (n = 520)	240 mg/d (n = 134)	All Doses (n = 1177)	300/100 mg/d (n = 521)
Total patients with at least 1 adverse event	119 (23%)	109 (21%)	39 (29%)	267 (23%)	101 (19%)
Diarrhoea (excluding infective)					
Diarrhoea	16 (3%)	12 (2%)	9 (7%)	37 (3%)	12 (2%)
Headaches NEC					
Headache	7 (1%)	12 (2%)	6 (4%)	25 (2%)	12 (2%)
Nausea and vomiting symptoms					
Nausea	11 (2%)	7 (1%)	6 (4%)	24 (2%)	4 (< 1%)
Vomiting	3 (< 1%)	0	0	3 (< 1%)	2 (< 1%)
Neurological signs and symptoms NEC					
Dizziness	7 (1%)	3 (< 1%)	4 (3%)	14 (1%)	2 (< 1%)
Dysgeusia	1 (< 1%)	1 (< 1%)	2 (1%)	4 (< 1%)	0
Gastrointestinal and abdominal pains (excluding oral and throat)					
Abdominal pain	1 (< 1%)	1 (< 1%)	4 (< 1%)	6 (< 1%)	1 (< 1%)
Abdominal pain lower	1 (< 1%)	0	0	1 (< 1%)	1 (< 1%)
Abdominal pain upper	1 (< 1%)	3 (< 1%)	1 (< 1%)	5 (< 1%)	1 (< 1%)
Liver function analyses					
ALT increased	5 (< 1%)	4 (< 1%)	0	9 (< 1%)	5 (< 1%)
AST increased	4 (< 1%)	5 (< 1%)	0	9 (< 1%)	4 (< 1%)
Bilirubin increased	0	2 (< 1%)	0	2 (< 1%)	0
Hepatic enzyme increased	7 (1%)	3 (< 1%)	1 (< 1%)	11 (< 1%)	8 (2%)
LFT abnormal	6 (1%)	8 (2%)	3 (2%)	17 (1%)	6 (1%)
Transaminases increased	0	1 (< 1%)	0	1 (< 1%)	0
Peripheral vascular disorders NEC					
Flushing	0	2 (< 1%)	2 (1%)	4 (< 1%)	1 (< 1%)
Hot flush	0	1 (< 1%)	1 (< 1%)	2 (< 1%)	0

ALT, alanine aminotransferase; AST, aspartate aminotransferase; LFT, liver function test; NEC, not elsewhere classified

Serious adverse event data were not tabulated and were selectively reported. The MS states that cardiac disorders were the most common serious adverse events (incidence >1%) and were consistent for all treatment groups, including allopurinol (no numerical data provided). These adverse events were more frequent in patients who had a history of cardiovascular disease or cardiovascular risk factors.

In the combined pivotal phase III trials and the long-term extension studies, a total of 12 deaths (manufacturer's response to ERG queries suggest 13 deaths) in febuxostat treated groups have been reported, none of which were considered related to the study drug compared to zero deaths in the allopurinol group (no numerical data were provided by the manufacturer for death rates by study and febuxostat doses). Six of the 12 deaths were attributed to cardiovascular risk factors such as a history of coronary artery disease, cerebrovascular accident, congestive heart failure and myocardial infarction. A further evaluation of these deaths showed no statistically significant increases in cardiovascular adverse events or deaths in the febuxostat groups versus the allopurinol group.

A post hoc subgroup pooled analysis on the safety and tolerability of febuxostat (80 mg/d, 120 mg/d and 240 mg/d) compared to fixed dose allopurinol (300/100 mg/d) in the elderly was also reported in the MS (p17 manufacturer's response to ERG queries). Although no numerical data were provided, the MS states that no clinically meaningful differences in the incidences of treatment emergent and treatment related adverse events were found between age groups (<45, 45 to 65, >65 years) within febuxostat groups and no age related marked increase in the adverse event rate. Furthermore, in subjects over 65 years of age (>65 to 75, >75 to 80, >80 years), no marked increase in the adverse events rate were observed in the febuxostat groups. A summary of these results are provided in Table 11. For the mean change in baseline analysis, a statistically significant interaction between age and treatment was found for the mean cell haemoglobin concentration at week 52 and for T4 at week 28. The ERG notes that these post hoc analyses need to treated with great caution as no statistical comparisons were reported and the sample sizes in each group were very limited.

Table 11: Most frequent adverse events in subjects aged >65 years

	Treatment groups											
	Febuxostat 80 mg/d			Febuxostat 120 mg/d			Febuxostat 240 mg/d			Allopurinol 300/100 mg/d		
	>65 to 75 years (n=61)	>75 to 80 years (n=5)	>80 years (n=2)	>65 to 75 years (n=57)	>75 to 80 years (n=7)	>80 years (n=3)	>65 to 75 years (n=17)	>75 to 80 years (n=5)	>80 years (n=2)	>65 to 75 years (n=55)	>75 to 80 years (n=12)	>80 years (n=6)
Total patients with at least 1 adverse event	43 (70%)	5 (100%)	1 (50%)	45 (79%)	5 (71%)	2 (67%)	12 (71%)	4 (80%)	2 (100%)	49 (89%)	11 (92%)	6 (100%)
Diarrhoea (excluding infective)	4 (7%)	0	1 (50%)	5 (9%)	0	1 (33%)	3 (18%)	1 (20%)	1 (50%)	9 (16%)	0	1 (17%)
Non site specific injuries NEC	4 (7%)	1 (20%)	0	2 (4%)	0	0	1 (6%)	1 (20%)	0	4 (7%)	1 (8%)	0
Nausea and vomiting symptoms	4 (7%)	1 (20%)	0	6 (11%)	1 (14%)	0	0	0	0	2 (4%)	1 (8%)	0
Muscle related signs and symptoms NEC	2 (3%)	0	0	1 (2%)	0	0	0	0	0	1 (2%)	1 (8%)	0

NEC, not elsewhere classified

Critique of safety data reported

The interpretation of the safety and tolerability data is difficult, as the allopurinol SPC generally recommends the initiation of allopurinol at low dose (e.g. 100 mg/d) and up-titrate in order to reduce the risk of adverse events.[7] Higher doses than that used in the FACT and APEX trials (300 mg/d) are also used in practice.

Although the short to medium term (up to 52 weeks) adverse event profile for febuxostat appears to be similar to allopurinol, published long-term safety data for febuxostat are lacking. It is also well recognised that RCTs have a limited ability to assess drug toxicity. Febuxostat safety data need to be supplemented by other types of study, including post-marketing surveillance studies, which can follow-up larger numbers of patients for longer periods of time, and which generally collect data relating to the target population treated in normal clinical practice rather than to highly selected populations treated under specialised conditions. In addition, all deaths in the febuxostat group were judged to be unrelated to the study drug; however, it is still a reason for concern.

The rates of discontinuation from the FACT and APEX trials suggest that febuxostat is not well tolerated as allopurinol, even when the latter is not up-titrated from low doses. The higher incidence of gout flares with febuxostat (particularly 120 mg/d) during initial prophylaxis with colchicine or naproxen may lead to compliance issues in the short-term, although the results suggest the incidence of flares may be reduced by starting at lower doses of febuxostat.

Data checking also highlighted some errors in the reporting of data between (and within) the MS and supplementary data reported by the manufacturer, particularly in the rates of discontinuation for gout flare, personal reason and other. Other errors were also evident in the adverse events sections of the MS (e.g. summation of individual adverse events did not correspond to each group results).

4.3.2 Critique of submitted evidence syntheses

Meta-analysis

The MS relies on a pooled analysis of data from the FACT and APEX trials and treats them as one large study. However as noted in section 4.1.7, the ERG considers this type of data pooling to be inappropriate as it fails to preserve randomisation and introduces bias and confounding. A more satisfactory statistical technique involves combining the results from two or more separate studies in a meta-analysis.[20,17,18,19] However, as requested, the results of such meta-analyses in the form of relative and absolute risk reductions using the both the fixed and random effects models were not provided by the manufacturer. These meta-analyses have therefore been calculated (from data provided from the individual studies in the MS or data from the primary published peer reviewed clinical paper of the FACT study[10] minus pooled results in the MS for data on the APEX trial), using the Cochrane Collaboration Review Manager 4.2.10 software.

Continuous and dichotomous data were combined using the inverse variance method of meta-analysis to give a weighted average of the effect estimates from the individual studies. Effect estimates for continuous data were obtained by comparing least squares mean (±standard deviation, SD) percentage change in outcome measure for each treatment group, from baseline to study end and are presented as a weighted mean difference (WMD) between treatments. The treatment goal outcomes were assessed as relative risk (probability) of reaching goal in one treatment group relative to other, during the trial period. It should be noted that a higher relative risk (or probability) of the outcome is desirable in the case of reaching treatment goal. Heterogeneity between trial results was explored using the chi^2 test and the I^2 measure.

A summary of the results from the meta-analysis are presented in Table 12 and Table 13. The meta-analysis showed that the probability of reaching sUA targets was significantly higher for hyperuricaemic gout patients treated with febuxostat (80 mg/d and 120 mg/d) relative to that for patients receiving fixed dose (300/100 mg/d) allopurinol ($p<0.00001$ for all comparisons except post hoc subgroup febuxostat 80 mg/d analysis of patients with sUA <6 mg/dL at last 3 visits by baseline [<9 mg/dL] sUA, $p<0.0002$). This might be expected as febuxostat 80 mg/d significantly reduced sUA levels by -11.63% ($p<0.00001$) compared with fixed dose

allopurinol. Similarly febuxostat 120 mg/d significantly reduced sUA levels by -18.34% ($p<0.00001$) compared with allopurinol. A summary of these results are presented in Figure 1 and Figure 2. For important clinical outcomes, the probability of gout flares needing treatment during (week 1 to 8) or after (week 9 to 52) prophylaxis with febuxostat (80 mg/d) compared to allopurinol was not significant ($p>0.18$ for all comparisons). The probability of gout flares needing treatment during or after prophylaxis was significantly lower with allopurinol relative to that for patients receiving febuxostat 120 mg/d ($p<0.05$ for all comparisons). The test for statistical heterogeneity was not significant for any of these outcomes ($p>0.05$).

The change in primary tophus size from baseline to end of study could not be calculated as appropriate data (i.e. mean ±SD) for meta-analysis were not provided (as requested) and data on the reduction of number tophi were not reported in the MS.

Table 12. Summary of ERG meta-analysis results – Febuxostat 80 mg/d versus allopurinol 300/100 mg/d (ITT analysis)

Outcomes	Febuxostat (80 mg/d)	Allopurinol (300/100 mg/d)	Relative risk (fixed effects, 95% CI)	Relative risk (random effects, 95% CI)	Risk difference (fixed effects, 95% CI)	Risk difference (random effects, 95% CI)
Primary efficacy endpoint						
sUA levels <6.0 mg/dL at last 3 visits	262/517 (51%)	113/519 (22%)	2.33 (1.94 to 2.80)	2.32 (1.93 to 2.79)	29% (+23% to +34%)	29% (+23% to +35%)
Secondary efficacy endpoints						
sUA levels <6.0 mg/dL at final visit [a,b]	368/502 (73%)	190/505 (38%)	1.95 (1.72 to 2.21)	1.95 (1.72 to 2.20)	36% (+30% to +41%)	36% (+30% to +41%)
Recurrent gout flares needing treatment						
Day 1 to week 8 (prophylaxis)	128/517 (25%)	113/519 (22%)	1.14 (0.91 to 1.42)	1.14 (0.91 to 1.42)	3% (-2% to +8%)	3% (-2% to +8%)
Week 9 to 52	269/451 (60%)	260/471 (55%)	1.08 (0.97 to 1.20)	1.08 (0.97 to 1.26)	4% (-2% to +11%)	4% (-4% to +12%)
Day 1 to week 52	312/517 (60%)	299/519 (58%)	1.05 (0.95 to 1.16)	1.04 (0.92 to 1.18)	3% (-3% to +9%)	3% (-4% to +10%)
Subgroup analysis						
sUA levels <6 mg/dL at last 3 visits (primary outcome) by baseline sUA						
<9 mg/dL	83/136 (61%)	54/142 (38%)	1.61 (1.26 to 2.06)	1.62 (1.26 to 2.07)	23% (+12% to +35%)	23% (+12% to +35%)
9 to <10 mg/dL	90/165 (55%)	41/177 (23%)	2.35 (1.74 to 3.18)	2.36 (1.74 to 3.19)	31% (+22% to +41%)	31% (+22% to +41%)
≥10 mg/dL	89/216 (41%)	18/200 (9%)	4.59 (2.88 to 7.32)	4.56 (2.85 to 7.28)	32% (+25% to +40%)	32% (+20% to +44%)
sUA levels <6 mg/dL at final visit [a,b] by baseline sUA						
<9 mg/dL	114/132 (86%)	79/140 (56%)	1.54 (1.31 to 1.80)	1.55 (1.32 to 1.81)	30% (+20% to +40%)	31% (+19% to +43%)
9 to <10 mg/dL	122/163	70/173	1.85	1.85	34%	34%

	(75%)	(40%)	(1.51 to 2.26)	(1.51 to 2.26)	(+24% to +44%)	(+25% to +44%)
≥10 mg/dL	132/207	41/192	2.99	2.99	43%	43%
	(64%)	(21%)	(2.24 to 4.00)	(2.24 to 4.00)	(+34% to +51%)	(+34% to +51%)
sUA levels at final visit [a,b]						
<5.0 mg/dL	234/502	65/505	3.62	3.62	34%	34%
	(47%)	(13%)	(2.83 to 4.63)	(2.83 to 4.63)	(+28% to +39%)	(+28% to +39%)
<4.0 mg/dL	96/502	10/505	9.68	9.47	17%	17%
	(19%)	(2%)	(5.10 to 18.36)	(4.99 to 17.97)	(+13% to +21%)	(+13% to +21%)

sUA, serum uric acid; NR, not reported; SD, standard deviation; CI, confidence interval; ITT, intention to treat

[a] Defined as last visit of a specific subject

[b] Available case analysis (not ITT analysis) – sensitivity analysis using worse case scenario in all randomised patients (ITT analysis) yielded similar results (data not reported)

Table 13. Summary of ERG meta-analysis results – Febuxostat 120 mg/d versus allopurinol 300/100 mg/d (ITT analysis)

Outcomes	Febuxostat (120 mg/d)	Allopurinol (300/100 mg/d)	Relative risk (fixed effects, 95% CI)	Relative risk (random effects, 95% CI)	Risk difference (fixed effects, 95% CI)	Risk difference (random effects, 95% CI)
Primary efficacy endpoint						
sUA levels <6.0 mg/dL at last 3 visits	329/519 (63%)	113/519 (22%)	2.91 (2.44 to 3.47)	2.91 (2.44 to 3.47)	42% (+36% to +47%)	42% (+36% to +47%)
Secondary efficacy endpoints						
sUA levels <6.0 mg/dL at final visit [a,b]	402/507 (79%)	190/505 (38%)	2.11 (1.87 to 2.38)	2.11 (1.87 to 2.38)	42% (+36% to +47%)	42% (+36% to +47%)
Recurrent gout flares needing treatment						
Day 1 to week 8 (prophylaxis)	187/519 (36%)	113/519 (22%)	1.65 (1.36 to 2.02)	1.65 (1.35 to 2.02)	14% (+9% to +20%)	14% (+9% to +20%)
Week 9 to end of study	279/455 (61%)	260/471 (55%)	1.12 (1.00 to 1.25)	1.11 (1.00 to 1.24)	7% (0% to +13%)	6% (0% to +13%)
Day 1 to end of study	348/519 (67%)	299/519 (58%)	1.16 (1.06 to 1.28)	1.15 (1.04 to 1.29)	9% (+3% to +15%)	9% (+3% to +15%)
Subgroup analysis						
sUA levels <6 mg/dL at last 3 visits (primary outcome) by baseline sUA						
<9 mg/dL	107/144 (74%)	54/142 (38%)	1.95 (1.55 to 2.46)	1.95 (1.55 to 2.46)	36% (+26% to +47%)	36% (+26% to +47%)
9 to <10 mg/dL	119/161 (74%)	41/177 (23%)	3.19 (2.40 to 4.23)	3.19 (2.40 to 4.23)	51% (+41% to +60%)	51% (+41% to +60%)
≥10 mg/dL	103/214 (48%)	18/200 (9%)	5.29 (3.33 to 8.40)	5.29 (3.33 to 8.39)	39% (+31% to +47%)	39% (+31% to +47%)
sUA levels <6 mg/dL at final visit [a,b] by baseline sUA						
<9 mg/dL	124/141 (88%)	79/140 (56%)	1.56 (1.33 to 1.82)	1.56 (1.33 to 1.83)	32% (+22% to +41%)	32% (+22% to +41%)
9 to <10 mg/dL	139/157	70/173	2.19	2.18	48%	48%

	(89%)	(40%)	(1.81 to 2.64)	(1.80 to 2.64)	(+39% to +57%)	(+39% to +57%)
≥10 mg/dL	139/209	41/192	3.11	3.10	45%	45%
	(67%)	(21%)	(2.33 to 4.15)	(2.32 to 4.14)	(+36% to +54%)	(+36% to +54%)
sUA levels at final visit [a,b]						
<5.0 mg/dL	331/507	65/505	5.07	5.07	52%	52%
	(47%)	(13%)	(4.01 to 6.42)	(4.01 to 6.42)	(+47% to +57%)	(+47% to +57%)
<4.0 mg/dL	200/507	10/505	19.92	19.53	37%	37%
	(39%)	(2%)	(10.68 to 37.14)	(10.47 to 36.42)	(+33% to +42%)	(+33% to +42%)

sUA, serum uric acid; NR, not reported; SD, standard deviation; CI, confidence interval; ITT, intention to treat

[a] Defined as last visit of a specific subject

[b] Available case analysis (not ITT analysis) – sensitivity analysis using worse case scenario in all randomised patients (ITT analysis) yielded similar results (data not reported)

Figure 1: Mean percentage change (fixed and random effects model) from baseline in sUA levels at final visit – Febuxostat 80 mg/d versus fixed dose allopurinol (300/100 mg/d)

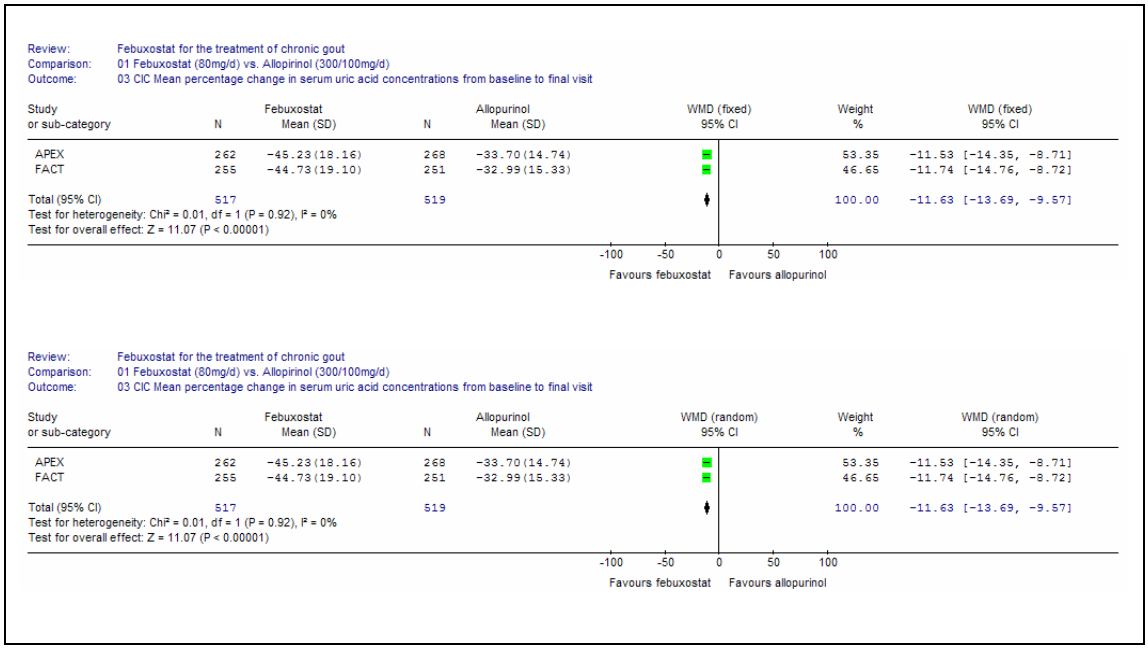
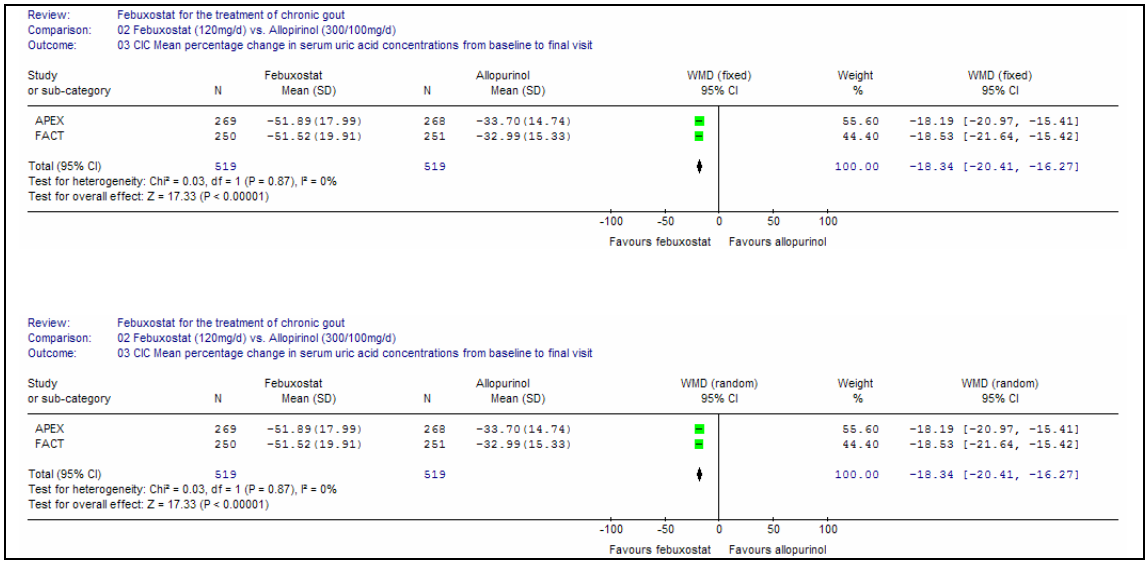


Figure 2: Mean percentage change (fixed and random effects model) from baseline in sUA levels at final visit – Febuxostat 120 mg/d versus fixed dose allopurinol (300/100 mg/d).



Indirect/ mixed treatment comparisons

No Indirect / mixed treatment comparisons have been conducted as only two treatments have been modelled, of which head-to-head trials do exist. If more interventions (including no active treatment) were modelled then a need for indirect comparisons may have become evident. Our clinical advisors have commented that the results for allopurinol therapy appear to be lower in the head-to-head trials than would have been expected from previous clinical trials. Reference to previous allopurinol trials could provide re-assurance that the head-to-head trials did not, by chance, favour or disfavour allopurinol.

4.3.3 Summary

Overall the MS contains a biased estimate of treatment efficacy for febuxostat in comparison to fixed dose allopurinol. The ERG has reservations over the appropriateness of the fixed dose comparator (gout management guidelines[5,6] and the allopurinol SPC[7] generally recommended initiation of allopurinol at low dose (e.g. 100 mg/d) and to up-titrate (maximum 900mg/d) according to therapeutic targets; however, the manufacturer and our clinical advisors note that dose escalation is rarely practiced by clinicians), and with the pooled treatment comparisons (as the approach fails to preserve randomisation and introduces bias and confounding). Whilst measures such as gout flares and tophi resolution were secondary outcomes, these are more clinically important. The evidence shows that, even though more febuxostat recipients achieved the recommended biochemical goal (<6 mg/dL), this did not translate into an advantage over allopurinol in clinically important outcomes.

5. ECONOMIC EVALUATION

5.1 Overview of manufacturer's economic evaluation

The manufacturer's present an estimation of the cost per QALY of febuxostat compared with allopurinol. The analyses presented only compare two strategies, the provision of allopurinol for the entire time horizon and the provision of febuxostat for the entire time horizon. For both interventions, treatment was assumed to persist regardless of the patient response. The number of gout flares, which have both cost and utility implications, occurring within the model were linked to the sUA level of the patient.

5.1.1 Natural history

The natural history of patients with gout who do not receive treatment is not modelled within the MS, which assumed that a patient would be prescribed either febuxostat or allopurinol. As such, the MS does not present incremental cost-effectiveness ratios compared with no treatment. No data has been presented on the likely sUA levels for patients who receive no treatment. The lack of no treatment data resulted in the inability to accurately model discontinuation of treatment, with the model assuming 100% compliance and concordance. This could potentially bias the model as data from the FACT and APEX trials (p44 and 47, MS) both showed that the discontinuation rate was greater on febuxostat than on allopurinol (the discontinuation rates for the APEX trial is marked commercial-in-confidence). Conversely, the percentages of patients remaining on treatment in the EXCEL trial reports show low persistence rates for all interventions (35% febuxostat 80 mg/d, 10% febuxostat 120 mg/d and 5% allopurinol) whilst the MS (p109) also reports randomised subset of the open-label EXCEL trial resulted in 76% of the patients on febuxostat 80 mg (n = 299), 71% on febuxostat 120 mg (n = 291) and 40% on allopurinol 300/100 mg (n = 145) remained on initial treatment after more than 24 months.[22]

It is noted that in the manufacturer's response to the NICE STA questions (p14, manufacturer's response to ERG queries) that the APEX study contained a placebo arm, which could be used to approximate the number of flares over the long-term for patients without treatment.

5.1.2 Treatment effectiveness within the submission

Gout Flares

- Initial 3 months

The numbers of gout flares within the initial three months of treatment have been taken from pooled data from the FACT and APEX trials, and are commercial-in-confidence (p105, MS). The flare rates have been reduced by 78% (in accordance with data from Borstad et al)[23] in the first three months, by assuming that patients received three months of prophylactic treatment with colchicine. However, this reduction will over-estimate the effect of colchicine, as eight weeks of colchicine prophylaxis was used in the FACT and APEX studies. This will introduce some inaccuracy, with the expected number of flares in the initial three months to be higher than that estimated in the MS.

- Beyond 3-months

After the first three months of treatment the proportion of gout flares has been assumed to be related to which of four sUA levels a patient is in. These categories are ≤360 μmol/L (6 mg/dL), >360 μmol/L (6 mg/dL) and ≤480 μmol/L (8 mg/dL), >480 μmol/L (8 mg/dL) and ≤600 μmol/L (10 mg/dL) and >600 μmol/L (10 mg/dL). The mean number of flares per category is indicated as commercial-in-confidence (p106, MS) and is presented in Table 14.

Table 14. Probability of experiencing flare(s) per sUA level, and calculated monthly number of flares, as at month 4 after treatment instalment

		Estimated monthly number of flares in the overall population[a]	
sUA level	**Prob. flare(s)**	**Mean**	**SE**
≤ 360 µmol/L (6 mg/dL)	0.6456	0.0874	0.00473
> 360 µmol/L (6 mg/dL) and ≤ 480 µmol/L (8 mg/dL)	0.7304	0.0989	0.00535
> 480 µmol/L (8 mg/dL) and ≤ 600 µmol/L (10 mg/dL)	0.8011	0.1085	0.00587
> 600 µmol/L (10 mg/dL)	0.8569	0.1161	0.00628

sUA, serum uric acid

[a] Includes patients who did not experience flare(s) during the observation period.

The increased odds of a gout flare in relation to sUA levels was provided (after a substantial delay) to the ERG in an accompanying document, IMSIII.[24] Data, as reported in the IMSIII report (p84) and reproduced in Table 15, claimed to be the results of a multivariate analysis. However within the text it states (p83) that "The significance of sUA level disappeared when controlling for other significant co-variates of the bivariate analysis." As the p-value was also the same as in the bivariate analysis of sUA level against the number of flares, the ERG believes that this is the bivariate analysis and that within this data set there was no significant link between sUA levels and the number of gout flares. However this does not mean that there would be no relationship detected were a bigger or different data set analysed.

Table 15. Parameter estimates of the variables significantly associated with the odds of experiencing a flare (logistic regression) - Results of multivariate analysis

Variable	P-value	Odds Ratio	Std. Error	95% CI	
				Lower	Upper
Constant	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
sUA level	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Patients on 80 mg/d febuxostat treatment are assumed to have a dose increase to 120 mg/d if the sUA level was not ≤360 µmol/L (6 mg/dL) after the initial three months of febuxostat treatment. The sUA levels of patients who changed dose from 80 mg/d to 120 mg/d were

assumed to be identical to a cohort of patients that had been prescribed 120 mg/d from the initiation of treatment. Dose titration of allopurinol was not permitted, regardless of the sUA level of the patient.

Beyond three months, the sUA levels for a cohort of patients were more favourable to those on febuxostat therapy than those on allopurinol; however, a sizeable proportion of patients are seen to respond on allopurinol treatment, which strengthens the case that an algorithm of interventions are required to be modelled in order to determine the most cost-effective strategy.

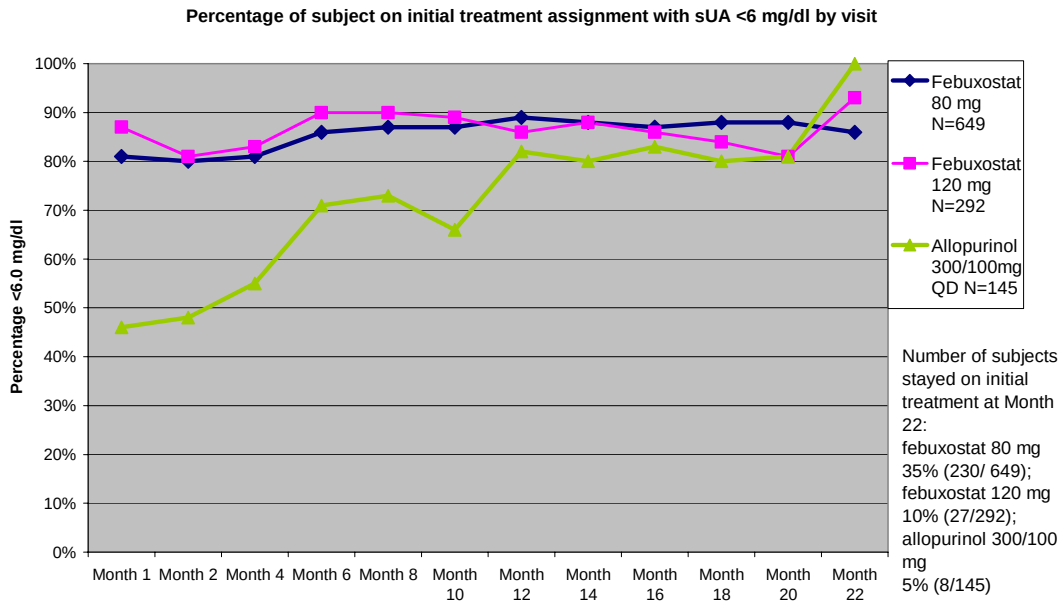
Table 16. Proportion of patients in each sUA category, by drug (allopurinol vs. febuxostat)—week 2 to month 12

	Allopurinol		Febuxostat 80 mg		Febuxostat 120 mg	
sUA level	**Mean**	**SE**	**Mean**	**SE**	**Mean**	**SE**
≤ 6 mg/dL	0.4159	0.0235	0.7340	0.0219	0.8498	0.0177
> 6 and ≤ 8 mg/dL	0.4614	0.0238	0.1970	0.0197	0.1010	0.0150
> 8 and ≤ 10 mg/dL	0.1023	0.0144	0.0567	0.0115	0.0443	0.0102
> 10 mg/dL	0.0205	0.0067	0.0123	0.0055	0.0049	0.0035

sUA, serum uric acid

It was assumed that patients remained in the sUA group based on evidence from the EXCEL study (p76-77, MS). Figure 3 shows that the percentage of patients with sUA < 360 μmol/L (6 mg/dL) remained relatively constant across time. It is noteworthy that only 735 patients were enrolled in the EXCEL trial (p74, MS); however, the data in figure 3 are based on 1086 patients.

Figure 3. Percentage of subjects on initial treatment with sUA < 360 µmol/L (6 mg/dL) by visit up to two years (EXCEL)[a]



sUA, serum uric acid

[a] Interim analysis of ongoing trial. The number of subjects in each time interval reflects the subjects exposure to date prior to switch or change in dose.

Health related quality of life

The utility for patients with gout is modelled in two ways, a chronic utility associated with sUA level and a decreased utility associated with a gout flare. The methodology for calculating utility scores by sUA category and a description of the data set used were provided in the IMSIII report.[24]

- Chronic utility score

Each sUA level was assigned a utility derived from the EQ-5D quality-of-life (QoL) assessment from individual patients. Using a multivariate analyses it was shown that an increase in level of sUA was associated with a decrease in utility of [redacted] (95% condidence interval, CI: [redacted] – [redacted]) for each increase in sUA level The average utility for patients in the sUA <360 µmol/L (6 mg/dL) level was estimated to be [redacted] (SE = [redacted]), which

was used within the model with a decrease in utility for the remaining sUA categories. The data is presented in Table 17.

Table 17. Baseline utility per sUA level

sUA level	Mean utility	SE
sUA ≤ 6 mg/dL	[redacted]	[redacted]
sUA > 6 mg/dL and ≤ 8 mg/dL	[redacted]	[redacted]
sUA > 8 mg/dL and ≤ 10 mg/dL	[redacted]	[redacted]
sUA > 10 mg/dL	[redacted]	[redacted]

sUA, serum uric acid

Within the Microsoft Excel spreadsheet model only the utility for patients with a level of sUA ≤ 6 mg/dL is sampled, from a beta distribution with parameters XXXXXXXX which has a 95% CI of [redacted] to [redacted]. The disutility associated with each increase of sUA level sampled from a beta distribution with parameters XXXXXXXX which has a 95% CI of XXXX to XXXX. It is noted that the underlying utility in patients with a level of sUA ≤ 6 mg/dL and for the disutility associated with increasing sUA levels are sampled independently, rather than sampled jointly in order that correlations within the data are explicitly incorporated. However the inaccuracy introduced by sampling independently is not expected to be large and would be anticipated to widen the uncertainty in the final results.

- Disutility associated with a gout flare

Each flare (with an assumed duration of one week) was assigned a utility penalty of XXXX (SE = XXXX) which is roughly equivalent to a utility of 0.5 for the seven day duration of the flare. This disutility of a gout flare is sampled within the model by a beta distribution with parameters XXXXXXXXX which has a 95% CI of XXXX to XXXXX.

5.1.3 Resources and costs

The resource costs used in the model are summarised in Tables 6-12 to 6-16 of the MS (p118-119). The methodology used to calculate costs were provided in a report that was received, after delay, by the ERG and does not appear to be incorrect.

5.1.4 Discounting

Both costs and QALYs have been discounted at 3.5% per annum in accordance with the NICE reference case.

5.1.5 Sensitivity analyses

A number of univariate sensitivity analyses were conducted that involved changing the time horizon of the analyses, the protective effect provided by colchicine prophylaxis, discount rates, the assumed cost of febuxostat, the disutility associated with each incremental level of sUA and the proportion of patients < 360 μmol/L in months 4 to 24 for febuxostat. These indicated that the parameters which made the most difference to the results in univariate analyses were the assumed cost of febuxostat, the disutility associated with each incremental level of sUA and the proportion of patients < 360 μmol/L in months 4 to 24 for febuxostat.

Only one set of PSA results have been provided by the manufacturer which were those accompanying the initial submission. Despite acknowledgement by the manufacturer that the prices of interventions should not be included in the PSA, (p39, manufacturer's response to ERG queries) and that the price of allopurinol was overestimated (p38 and 39) and that the prophylaxis success rate should be included in the PSA, no further runs were conducted, although a deterministic evaluation of reducing the price of allopurinol and of using 0% and 100% for prophylaxis success rates were conducted which did not markedly affect the results. However, a re-run of the analyses would have been much more preferable.

The variables altered in the PSA, as reported in the MS, are given in Table 18. It is noted that this table, which implies that the numbers of flares for each sUA category are independently sampled, does not replicate the PSA actually conducted within the model, which samples the number of flares in patients within the sUA ≤ 6 mg/dL group, and then assumes a constant proportional increase in other sUA levels. As the model assumes a fixed ratio, this will underestimate the uncertainty in the results. A similar logic is applied to the distribution of patients amongst sUA levels; only the percentage of patients in the sUA ≤ 6 mg/dL level are sampled with the remaining patients divided amongst the remaining categories using fixed proportions. This will again underestimate the uncertainty within the model.

Table 18. Parameters applied in the probabilistic sensitivity analysis - distribution, mean and SE

Parameter	Distribution	Mean	SE
Clinical data			
First 3 months after treatment initiation			
Incidence of flares (N flare/ 3 month)			
Allopurinol	[redacted]	[redacted]	[redacted]
Febuxostat pooled (80 mg + 120 mg)	[redacted]	[redacted]	[redacted]
From month 4 after treatment initiation			
Proportion of patients in sUA level ≤ 6 mg/dL			
Allopurinol	[redacted]	[redacted]	[redacted]
Febuxostat pooled (80 mg + 120 mg)	[redacted]	[redacted]	[redacted]
Monthly number of flares, by sUA level	[redacted]	[redacted]	[redacted]
sUA ≤ 6 mg/dL	[redacted]	[redacted]	[redacted]
6 mg/dL > sUA ≤ 8 mg/dL	[redacted]	[redacted]	[redacted]
8 mg/dL > sUA ≤ 10 mg/dL	[redacted]	[redacted]	[redacted]
sUA > 10 mg/dL	[redacted]	[redacted]	[redacted]
Utility data			
Flare penalty: utility decrease per experienced flare	[redacted]	[redacted]	[redacted]
sUA ≤ 6 mg/dL	[redacted]	[redacted]	[redacted]
Decrease in utility per sUA level:	[redacted]	[redacted]	[redacted]
Cost data (UK)			
Allopurinol (£)	Gamma	0.065	0.0098
Febuxostat (£)	[redacted]	[redacted]	[redacted]
Cost of 1 flare (£):	[redacted]	[redacted]	[redacted]

5.1.6 Model validation

The MS states that "All data entries and formulas were checked by a modeller at IMS not involved in the model building" (p125, MS). A small number of errors were found which included the inappropriate omission and inclusion of variables within the PSA, the fixed proportion of increased number of flares per sUA levels and patients apportioned between remaining sUA levels once the number in the sUA ≤ 6 mg/dL level are sampled. The methodology of calculating median cost per QALY was incorrect, although this metric is not commonly used, additional an erroneous cell reference in a formula was detected. As described earlier the independent sampling of the utility level for patients of sUA ≤ 6 mg/dL level and the utility decrease per increase in sUA category will also add inaccuracy. As re-analyses were not undertaken the ERG cannot determine the level of inaccuracy.

5.2 Critique of approach used

5.2.1 Critique of the modelling structure used

The ERG has serious concerns regarding the cost-effectiveness analyses undertaken, primarily due to the model structure and the strategies compared. The analyses presented only compare two strategies, the provision of allopurinol for the entire time horizon and the provision of febuxostat for the entire time horizon. For both interventions, treatment was assumed to persist regardless of the patient response. In the clarification letter sent to the manufacturer the ERG suggested that the following strategies, where patients progress to the next intervention following lack of response, should be evaluated as a minimum. Allopurinol – Febuxostat – No Treatment; Febuxostat – Allopurinol – No-Treatment; Allopurinol – No Treatment and Febuxostat – No Treatment. Assumptions regarding the efficacies of second-line interventions would be necessary, with the base-case assuming that the effect on sUA levels and gout flares remains at the level seen in first-line treatments.

These strategies allow the option for patients to be prescribed the cheaper, generic intervention (allopurinol) and to only progress to the more expensive, non-generic treatment (febuxostat) if a lack of response was identified. Often such a strategy is estimated to be the most cost-effective approach. The inclusion of a no treatment arm is to determine whether any of the two active interventions are cost-effective, without this the more cost-effective treatment of two treatments where neither are cost-effective could be selected. The manufacturer declined to meet this request, claiming that their 'cost-effectiveness analysis has been performed from the perspective of a first-line treatment' (p3, manufacturer's response to ERG queries). This approach is not as appropriate as an approach that looks at sequential treatment in the form of an algorithm. It is claimed that "A second-line treatment with febuxostat was not considered, since first line treatment appears to be cost-effective. It would be ethically unjustified to deprive patients of an effective first line treatment that is cost-effective" (p3, manufacturer's response to ERG queries). The ERG disputes this claim, as febuxostat is only likely to be the most cost-effective strategy as a first –line treatment since an algorithm of allopurinol followed by febuxostat was not modelled. The manufacturer claims that "An estimation of the impact of a sequential approach is unfeasible, since there are no clinical data available regarding: (1) The efficacy of febuxostat in patients who are not (or insufficiently) responding to allopurinol, (2) The efficacy of allopurinol in patients who are

not (or insufficiently) responding to febuxostat (3) The clinical evolution of patients who have stopped febuxostat or allopurinol treatment. Whilst this data is not available, modelling assumptions can be made in order that estimation of the cost-effectiveness of treatment strategies could be made. Since our clinical advisors were not aware of data that showed that the efficacy of an intervention would be dependent on previous treatments, a base-case analyses assuming constant efficacy regardless of placement in a treatment algorithm should have been conducted. The lack of a no-treatment arm in the submitted modelling also has the consequence that it is not possible to determine whether either of the interventions are cost-effective. Given these reservations, it is believed that the model is not appropriate for determining the position of febuxostat in the treatment of patients with gout.

The manufacturer has implicitly provided data to show that an algorithm of allopurinol followed by febuxostat is likely to be the most cost-effective algorithm, as it is assumed within the model that patients who have a sUA level ≤ 6 mg/dL at the end of three months are estimated to remain in this category, thus those patients who respond to allopurinol, which is markedly cheaper than febuxostat, are also likely to remain adequately treated. This is confirmed on page 35 of the MS where it is stated that "There is strong evidence from the EXCEL extension study that the proportion of patients in low sUA levels continue over the longer time period for patients who remain on active treatment. This supports the model assumption that the sUA levels achieved at one year can continue out to the two year horizon (and possibly beyond)." In response to the ERG comment that data from Roddy et al[25] states that 77% of patients on allopurinol had a sUA level ≤ 6 mg/dL, the manufacturer's comment (p36-37 manufacturer's response to ERG queries) that this figure has also been reported as varying from 19%-56%[26,27] (reference was also made to Dalbeth et al. 2006; however, no further bibliographic details were provided by the manufacturer). This confirms that a sizeable proportion of patients can be treated successfully with allopurinol.

5.2.2 Critique of the data set and methodology used to estimate a relationship between sUA levels and the number of gout flares

The ERG have serious concerns with both the methodology and interpretation of data contained in the MS and the manufacturer's response to ERG queries.[24] Unfortunately, as this additional data were provided to the ERG only five working days before the deadline for the

submission of the ERG report it was not possible to request clarification from the manufacturer on contentious issues.

5.2.3 Concern regarding discarded data

The original intention was for IMS to recruit physicians in the UK, France and Germany to collect data on 140 patients with gout in each country. However "Physician and patient recruitment appeared to be very difficult, causing delays in the anticipated project delivery. For that reason, it was decided … to have Medimix involved in the recruitment and data collection process for France and the UK." The data collected by the 3rd party (Medimix) was subsequently discarded in favour of data collected by IMS. This was justified by the statement "we therefore suspected that the data collected by Medimix did not reflect reality and would cause substantial bias in the outcomes of the analyses" (p33, IMSIII). This resulted in over half (51%) of the collected data and over three-quarters (77%) of UK data being discarded. The paraphrased, stated reasons for discarding the data collected by Medimix are as follows:

1) Medimix collected data mostly on patients for whom IMS expected, a priori, that there would be a higher resource use, which were patients with at least one of the following: currently having tophi, patients with an sUA level > 10 mg/dL and patients intolerant of allopurinol. This was not borne out in the UK results where the numbers of gout flares, proportion of patients having a gout flare, hospitalisations, laboratory tests, diagnostic tests and outpatient visits were all substantially lower in the Medimix patient group than in the IMS data set.
2) That more patients could be allocated to one of the four sUA levels in the IMS data than in the Medimix data.
3) That the IMS data allows an insight into the situation within Germany.
4) That IMS were "confident with its own data collection methods" and the reliability of their collected data.
5) That IMS had contact with the physicians they recruited and could thus explain the data collection methods in the case of uncertainties.

These points are addressed individually.

1) Whilst it may be true that there could be proven differences between the data collected by the two different agencies this would be best highlighted in a multivariate analysis using categorical variables for agency (such as 0 for IMS and 1

for Medimix). This would allow a much better insight on whether there was indeed a statistically significant difference having accounted for confounders. It is noted that for the only other country (France) where there were data collected by both IMS and Medimix that the data were as suspected a priori, and that the number of gout flares and hospitalisations were higher in the Medimix set. It is noted that whilst the proportion of patients having a gout flare was significantly higher in France for the Medimix data than the IMS data (95% CI: 5% to 46%) there were no significant difference in the proportion of patients with gout flares in the UK (95% CI: -27% to 4%) – ERG calculations. It appears from the IMSIII report[24] that the mean number of flares is significantly lower in the UK for the Medimix data and it would be interesting to know whether the number of flares per patient who suffers at least one flare can be explained by patient variables. Any such confounders may result in a relationship between sUA levels and an increase in the number of gout flares to become hidden within broad summary statistics and does not mean the data sets are necessarily incompatible.

2) Whilst it would be advantageous within the cost-effectiveness model to have a complete data set, the presence of unclassified sUA levels is not a reason in itself to prefer the IMS data. As the status of 13 more patients were undefined in the Medimix data set (p7, IMSIII) but Medimix collected data on 11 more patients (p5, IMSIII) the difference in the numbers of patients with classified sUA levels between collection agency is small. The exact reasons for a patient being unclassified are not provided and some investigation should be undertaken to establish whether this could be related to the true sUA level.
3) We believe this is not a valid reason for a study focussed on UK patients. The Medimix data set contains substantially more UK patients, and in isolation would be preferred for that reason. Since country has shown to be a significant predictor of the number of gout flares (p43, IMSIII) the results for Germany may not be applicable to the UK.
4) These collection methods have not been adequately described, nor have the suspected deviances from these, and their implications in the data collection methodology undertaken by Medimax.
5) Whilst contact with the clinicians may prove beneficial in clarifying uncertainties there is also potential to introduce recall bias, or even potentially to affect the population analysed. It is noted in the IMSIII report that they found physician and patient recruitment very difficult (p4) which may have affected the underlying population. For example, were only those patients in the UK with most persistent

gout, who have increased flare rates independent of sUA levels recruited, then the results would not be surprising.

Other considerations:

The ERG are not convinced by the reasons presented that the Medimix data should be discarded, indeed there may be an argument that since IMS only recruited 32 patients in the UK compared with 108 by Medimix that if one data set were to be discarded it be that from the IMS. The ERG would like to have seen additional data analysed using multivariate techniques for both all data pooled, and for the Medimix data set alone. Furthermore, other models in addition to linear should be explored between the ordinal sUA levels and the increased number of flares. Table 55 in the IMSIII report (p45) shows the increased odds of experiencing a flare and an increase in categorical sUA level. It is noted that there is a much larger mean increase for those patients with a sUA level >10 mg/dL compared to those below this threshold, and it may be that an exponential model be more appropriate than a linear model. It is also noted that Medimix data was used to inform the relationship between sUA level and utility. There is thus some inconsistency regarding to what extent these data are appropriate, and there is a strong concern that the data have been excluded for the relationship between sUA levels and the expected number of gout flares purely as the results were not as desired.

5.2.4 Interpretation of the multivariate analyses conducted

Irrespective of our concerns regarding the data analysed, we have concerns regarding the multivariate analysis reported (p83-85, IMSIII). It is stated that the results are presented in Table 100 of the IMSIII report, "but that the significance of sUA level disappeared when controlling for other significant co-variates of the bivariate analysis". It is thus implied that the sUA levels are not significantly related to the number of gout flares once confounders are incorporated. The full multivariate analyses should be presented, along with the output from the run, for instance were a backward stepwise analyses undertaken, the order of variables removed from the analysis should be given along with descriptive statistics with the final model ultimately presented. It is also recommended that all variables be included, not just those which were significant in the bivariate analysis. The implication of sUA levels being non-significant when confounders are included is questioned by Table 100 of the IMSIII report (p84); however, it is believed by the ERG that Table 100 is not a true multivariate analysis, but investigates only the relationship between sUA levels and the odds of

experiencing a gout flare. This would reproduce the bivariate analysis and it is noted that the p-value from Table 100 of the IMSIII report (p84) is, to the level of accuracy reported, identical to that in Table 52 of the IMSIII report (p43) which details the results for the bivariate analysis of sUA level compared with odds of experiencing a flare.

It is noted that clinical guidelines[5,6] indicate that there is a relationship between sUA level and the probability of a gout flare, with patients with higher sUA levels being more prone to gout flares. The lack of a statistically significant relationship within this data set does not mean that such a relationship does not exist.

5.2.4 Critique of the relationship between sUA level and utility

Whilst the analyses appear to be appropriately conducted it is noted that a small proportion of patients, *without a flare*, have a score below zero, indicating that they experience their health status as being "worse than death".

It is additionally noted that there is a possibility that confounders such as hypertension, may have been omitted from the data collection and that the relationship between sUA level and disutility may manifest itself only in a gout flare. This hypothesis has some support as gout is often viewed as an asymptomatic disease unless a gout flare occurs.

5.3 Results included in manufacturer's submission

As previously stated in section 5.1.7, despite the recognition of potential errors within the submitted PSA results, further analyses were not undertaken and thus the robustness of the data is questionable. Nevertheless, if the PSA results are assumed to be correct then the mean incremental cost per QALY was £16,324 at two years (Table19)

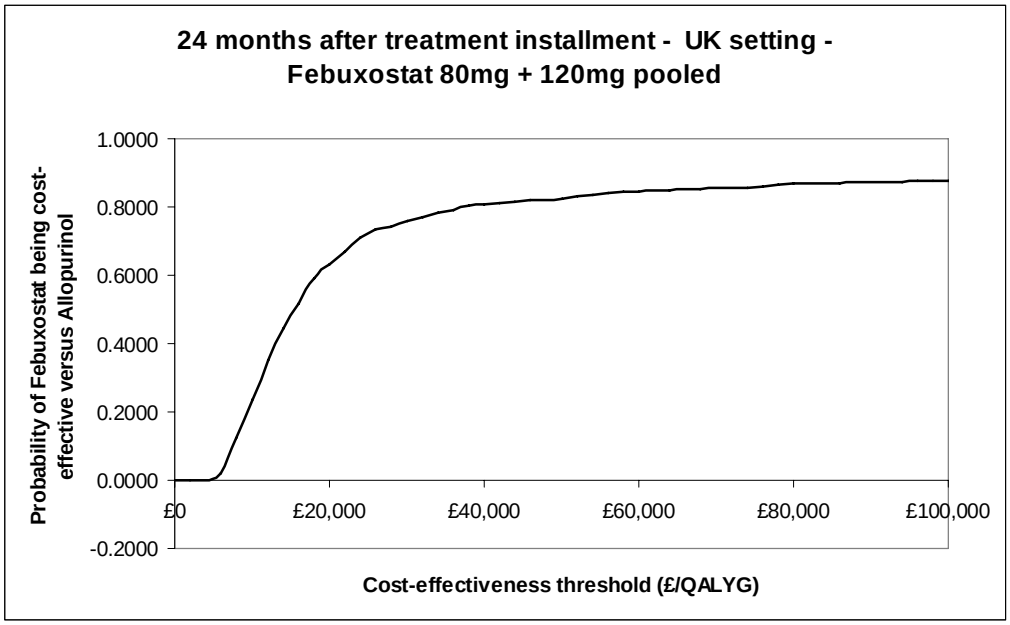
Table 19 Cost and QALY based on Monte Carlo analysis at 24 months in the UK

		95% CI			95% CI	
Treatment	**Cost**	**Lower**	**Upper**	**QALY**	**Lower**	**Upper**
Allopurinol	£2,606	£2,102	£3,223	1.399	1.291	1.510
Febuxostat	£3,145	£2,612	£3,770	1.432	1.346	1.510
Incremental	£539	£347	£776	0.033	-0.017	0.083
ICER	£16,324	£6,281	£239,928	-	-	-

QALY, quality-adjusted life-year

The cost-effectiveness acceptability curve (CEAC), as reported in the MS, is reproduced in figure 4. The MS reports that the percentage of times in which 80 mg per day of febuxostat, titrated to 120 mg per day where appropriate, had an incremental cost per QALY of < £20,000 compared with allopurinol was 63%.

Figure 4. Cost-effectiveness acceptability curve: UK setting



5.4 Comment on validity of results presented with reference to methodology used

As stated in section 5.2, the ERG does not believe the structure of the model is correct to determine the positioning of febuxostat in the treatment of patients with gout where cost-effectiveness is a consideration. As such, the results produced by this model have little

relevance in determining the use of febuxostat as a first-line therapy. The ERG requested a different set of analyses to be conducted in order that appropriate and relevant results were produced, but these were not undertaken by the manufacturer.

The ERG also believes that the data presented for the increased rate of gout flares with reference to the sUA category is highly uncertain. The ERG do not accept the reasoning for discarding 77% of the UK data set, and 51% of the overall data set, based on the evidence presented. The ERG also believes that the full multivariate analysis would not show a statistically significant linear relationship between sUA levels and the number of gout flares; however, clinical advice suggests that some relationship, not necessarily linear, is likely to exist between sUA levels and the number of gout flares. As this relationship is a pivotal driver of the cost-effectiveness ratios produced, uncertainty in this variable will translate into an uncertain cost per QALY of febuxostat compared with allopurinol. In the extreme case, where no relationship is assumed, allopurinol would dominate febuxostat as the interventions would have the same outcomes with the febuxostat patients having a higher acquisition price.

5.5 Summary of uncertainties and issues

The manufacturer provided data to show that where patients are not allowed to change treatments that febuxostat is more cost-effective than allopurinol. It is further commented that the PSA results provided by the manufacturer are known to omit variables, and includes the price of febuxostat, which casts doubts on the validity of this analysis. It is also noted that there is great uncertainty in the relationship between sUA level and the number of gout flares. The manufacturer's have used the odds of a gout flare associated with a bivariate analysis and overlooked those from a multivariate analysis that does not show a statistically significant relationship. This will add great uncertainty to the results produced by the cost-effectiveness analyses undertaken.

The model structure used by the manufacturer means that there is no estimation of the cost per QALY of algorithms containing febuxostat, allopurinol and no active intervention. This thus excludes the possibility of patients being initially treated with allopurinol and then switching to febuxostat or no active treatment, were the clinical response to allopurinol deemed insufficient. Thus, irrespective of concerns regarding the appropriateness of the assumed

relationship between sUA levels and the number of gout flares, key data that would be required by a funding agency is not provided.

6. ADDITIONAL WORK UNDERTAKEN BY THE ERG

The ERG believed that the pooling of clinical efficacy data was inappropriate (see Section 4.2.2) and conducted an appropriate meta-analysis. The ERG did not conduct any additional modelling work as it was believed that the fully sequential approach to treatment was necessary to calculate robust cost-effectiveness analyses, which was outside the remit of this report.

7. DISCUSSION

7.1 Summary of clinical effectiveness issues

The clinical evidence in the submission is derived from two head-to-head, phase III, multi-arm, randomised, double blind, controlled trials comparing the efficacy and safety of febuxostat with fixed dose allopurinol in patients with hyperuricaemia (sUA levels ≥8 mg/dL) and gout.

The main results, based on a simple pooled analysis of the patient level data from the FACT and APEX trials showed that a daily dose of febuxostat 80 mg or 120 mg was significantly more efficacious than allopurinol at the commonly used fixed daily dose of 300 mg in lowering sUA levels to therapeutic targets (<6 mg/dL). However, a large percentage of patients on febuxostat did not achieve the primary endpoint and the fixed dose regimen employed for allopurinol patients may have introduced bias. In general, there were no differences between treatments in more clinically important outcomes such as gout flares and tophi resolution after 52 weeks of treatment. No subgroup analyses were conducted for patients with renal impairment, non responders to allopurinol or patients with severe disease. In the ERG's opinion, the statistical approach for combining the data appears to be inappropriate as it fails to preserve randomisation and introduces bias and confounding. Therefore, resulting pooled data should therefore be treated with caution. Supplementary data, from a two year open label extension study (EXCEL) of the FACT and APEX trials, were also provided but were difficult to interpret (no statistical analysis undertaken) and poorly reported.

7.2 Summary of cost-effectiveness issues

The ERG has serious concerns regarding the cost-effectiveness analyses undertaken, primarily due to the model structure and the strategies compared. The analyses presented only compare two strategies, the provision of allopurinol for the entire time horizon and the provision of febuxostat for the entire time horizon. For both interventions, treatment was assumed to persist regardless of the patient response. In the clarification letter sent to the manufacturer the ERG suggested that the following strategies, where patients progress to the next intervention following lack of response, should be evaluated as a minimum. Allopurinol – Febuxostat – No Treatment; Febuxostat – Allopurinol – No-Treatment; Allopurinol – No

Treatment and Febuxostat – No Treatment. Assumptions regarding the efficacies of second-line interventions would be necessary, with the base-case assuming that the effect on sUA levels and gout flares remains at the level seen in first-line treatments.

These strategies allow the option for patients to be prescribed the cheaper intervention (allopurinol) and to only progress to the more expensive treatment (febuxostat) if a lack of response was identified. Often such a strategy is estimated to be the most cost-effective approach. The inclusion of a no treatment arm is to determine whether any of the two active interventions are cost-effective, without this the more cost-effective treatment of two treatments where neither are cost-effective could be selected. In its present form the model does not answer the key question facing a funding body, namely which treatment algorithm is most cost-effective for treatment of patients with gout.

The relationship between the sUA level of a patient and the expected number of flares that has been calculated in the submission is of concern both in the reduction of the data set and in the interpretation of the results. Whilst clinicians believe that higher sUA levels are likely to be associated with a greater number of flares, this could not be proven from the data set collected and the true relationship remains uncertain.

The relationship between sUA level and underlying utility also remains uncertain. Whilst the data appear to suggest that increased sUA levels lead to lower overall utility, the possibility that confounders not collected in the data set could explain this relationship cannot be ruled out.

7.3 Implications for research

The MS leaves a fundamental question unanswered, namely which treatment algorithm is most cost-effective for treatment of patients with gout. This is an area for future research; however, the experience of the ERG would suggest that providing allopurinol as a first line therapy would be the most cost-effective strategy for all patients that are not contraindicated to allopurinol. The cost-effectiveness of prescribing febuxostat to patients who do not clinically respond to allopurinol compared to no further treatment has not been evaluated, and should be the subject of further research.

Further research is required to determine the relationship between sUA levels and the expected number of gout flares.

Further research is required to determine if there is a relationship between sUA levels and the utility of a patient during periods between gout flares.

In addition to the short-term implications of gout flares and their associated resource use and complications, the longer-term impact of high sUA levels which may manifest in chronic tophaceous gouty arthritis, joint erosion and permanent damage and disability could markedly affect the cost-effectiveness ratios. Such relationships would benefit from further research, even if only from observational studies.

8. REFERENCES

1. Annemans, L., Spaepen, E., Gaskin, M., Bonnemaire, M., Malier, V., Gilbers, T., and Nuki, G. Gout in the UK and Germany: prevalence, comorbidities and management in general practice 2000-2005. *Ann Rheum Dis* 2007 (EPub ahead of print)

2. Mikuls, T. R., Farrar, J. T., Bilker, W. B., Fernandes, S., Schumacher, H. R., and Saag, K. G. Gout epidemiology: results from the UK General Practice Research Database, 1990-1999. *Ann Rheum Dis* 2005; **64** 267-272.

3. Catton, M., Schlesinger, N., Zhou, X., and Schumacher, R. Allopurinol for chronic gout (Protocol). *Cochrane Database of Systematic Reviews* 2006; **Issue 3.** Art. No.: CD006077. DOI: 10.1002/14651858.CD006077.

4. Ferraz, M. B. and O'Brien, B. A cost effectiveness analysis of urate lowering drugs in nontophaceous recurrent gouty arthritis. *J Rheumatol.* 1995; **22** 908-914.

5. Jordan, K. M., Cameron, S., Snaith, M., Zhang, W., Doherty, M., Seckl, J., Hingorani, A., Jaques, R., and Nuki, G. British Society for Rheumatology and British Health Professionals in Rheumatology Guideleine for the Management of Gout. *Rheumatology* 2007; **46** 1372-1374.

6. Zhang, W., Doherty, M., Bardin, T., Pascual, E., Barskova, V., Conaghan, P., Gerster, J., Jacobs, J., Leeb, B., Liote, F., McCarthy, G., Netter, P., Nuki, G., Perez-Ruiz, F., Pignone, A., Pimentao, J., Punzi, L., Roddy, E., Uhlig, T., and Zimmermann-Gorska, I. EULAR evidence based recommendations for gout. PartII: Managements. Report of task force of the EULAR Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). *Ann Rheum Dis* 2006; **65** 1312-1324.

7. Zyloric. Summary of product characteristics (GlaxoSmithKline). Available at: *http://emc.medicines.org.uk/emc/industry/default.asp?page=displaydoc.asp&documentid=4565* Accessed 10th March 2008

8. Joint Formulary Committee British National Formulary. [54] ed. London: British Medical Association and Royal Pharmaceutical Society of Great Britain; 2007

9. Schumacher, H. R., Taylor, W., Joseph-Ridge, N., Perez-Ruiz, F., Chen, L. X., Schlesinger, N., Khanna, D., Furst, D. E., Becker, M. A., Dalbeth, N., and Edwards, L. N. Outcome evaluations in gout. *The Journal of Rheumatology* 2007; **34** 1381-1385.

10. Becker, M. A., Schumacher, R., Wortmann, R. L., MacDonald, P. A., Eustace, D., Palo, W. A., Streit, J., and Joseph-Ridge, N. Febuxostat compared with Allopurinol in patients with Hyperuricemia and Gout. *N Engl J Med* 2005; **353** 2450-2461.

11. Schumacher, H. R., Becker, M. A., Wortmann, R. L., MacDonald, P. A., Hunt, B., Streit, J., Lademacher, N., and Joseph-Ridge, N. Febuxostat vs Allopurinol and Placebo in Subjects with Hyperuricemia and Gout: the 28-week APEX study. Poster presented at the Annual Scientific meeting of the American College of Rheumatology; San Diego, CA. 16-17 Nov 2005.

12. Becker, M. A., Schumacher, H. R., Jr., Wortmann, R. L., MacDonald, P. A., Palo, W. A., Eustace, D., Vernillet, L., and Joseph-Ridge, N. Febuxostat, a novel nonpurine selective inhibitor of xanthine oxidase: a twenty-eight-day, multicenter, phase II, randomized, double-blind, placebo-controlled, dose-response clinical trial examining safety and efficacy in patients with gout. *Arthritis Rheum.* 2005; **52** 916-923.

13. Schumacher, H. R., Becker, M. A., Wortmann, R. L., LLoyd, E., MacDonald, P. A., and Joseph-Ridge, N. The FOCUS trial 48-month interim analysis: long-term clinical outcomes of treatment with febuxostat in subjects with gout in an ongoing phase 2, open-label extension study. *Ann Rheum Dis* 2006; **65** (Suppl II) 93

14. Fergusson, D., Aaron, S. D., Guyatt, G., and Hebert, P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. *BMJ* 2002; **325** 652-654.

15. Hollis, S. and Campbell, F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ* 1999; **319** 670-674.

16. Taylor, W., Schumacher, H. R., Singh, J. A., Grainger, R., and Dalbeth, N. Assessment of outcomes in clinical trials of gout - a review of current measures. *Rheumatology* 2007; **46** 1751-1756.

17. Egger, M., Smith, G. D., and Phillips, A. N. Meta-analysis: principles and procedures. *BMJ* 1997; **315** 1533-1537.

18. Lievre, M., Cucherat, M., and Leizorovicz, A. Pooling, meta-analysis, and the evaluation of drug safety. *Current Controlled Trials in Cardiovascular Medicine* 2002; **3** 6

19. Bravata, D. M. and Olkin, I. Simple pooling versus combining in meta-analysis. *Eval Health Prof* 2001; **24** 218-230.

20. Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.0.0 [updated February 2008]. The Cochrane Collaboration, 2008. Available from www.cochrane-handbook.org.

21. National Institute of Health and Clinical Excellence (NICE) Single Technology Appraisal (STA) specification for manufacturers/sponsor submission of evidence. London, NICE 2006 Available at: http://www.nice.org.uk/niceMedia/pdf/STASpecManufacturerSubofEvidence.pdf

22. Becker, M. A., Schumacher, H. R., MacDonald, P. A., Lademacher, C., and Joseph-Ridge, N. Urate-lowering therapy in subjects with gout: interim results from the febuxostat/allopurinol comparative extension long-term study (EXCEL). Poster presented at the European League Against Rheumatism. *Ann Rheum Dis.* 2007; **66** (Suppl.2): 230.

23. Borstad, G. C., Bryant, L. R., Abel, M. P., Scroggie, D. A., Harris, M. D., and Alloway, J. A. Colchicine for prophylaxis of acute flares when initiating allopurinol for chronic gouty arthritis. *J Rheumatol*. 2004; **31** 2429-2432.

24. IMSIII. Study report - A health economic assessment of febuxostat in the management of gout. Part III: Cross sectional assessment of medical resource use in

patients with gout. IMS World Publications Ltd, 2008 (provided as Commercial in Confidence)

25. Roddy, E., Zhang, W., and Doherty, M. The changing epidemiology of gout. *Nat.Clin.Pract.Rheumatol*. 2007; **3** 443-449.

26. Reinders, M. K., van Roon, E. N., Houtman, P. M., Brouwers, J. R., and Jansen, T. L. Biochemical effectiveness of allopurinol and allopurinol-probenecid in previously benzbromarone-treated gout patients. *Clin.Rheumatol.* 2007; **26** 1459-1465.

27. Reinders, M. K., van Roon, E. N., Jansen, T. L., Delsing, J., Griep, E. N., Hoekstra, M., van de Laar, M. F., and Brouwers, J. R. Efficacy and tolerability of urate lowering drugs in gout: a randomised controlled trial of benzbromarone versus probenecid after failure of allopurinol. *Ann.Rheum.Dis.* 2008 [Epub ahead of print]