Evidence Review Group Report
Ruxolitinib for the treatment of myelofibrosis

Produced by CRD and CHE Technology Assessment Group, University of York, Heslington, York, YO10 5DD

Authors Dawn Craig, Research Fellow, CRD
Aileen Neilson, Research Fellow, CRD
Micah Rose, Research Fellow, CRD
Ros Wade, Research Fellow, CRD
Lisa Stirk, Information Specialist, CRD
Rocio Rodriguez-Lopez, Information Specialist, CRD
David Bowen, Honorary Professor of Myeloid Leukaemia Studies and Consultant Haematologist, St James's Institute of Oncology
Nerys Woolacott, Senior Research Fellow, CRD

Correspondence to Nerys Woolacott, Senior Research Fellow, CRD, University of York, York, YO10 5DD

Date completed 4 December 2012

Source of funding

This report was commissioned by the NIHR HTA Programme as project number HTA 11/79.

Declared competing interests of the authors

Professor David Bowen has received payments from Novartis, although not directly related to ruxolitinib. He attended a Novartis UK Advisory Board in 2011 to discuss the management of myeloproliferative disease and the position of ruxolitinib within the current agents used. In addition he works closely with Novartis in other areas and has the following relationships with them: member of Trial Steering Committee for a global commercial phase 3 study (TELESTO) for which he is paid per hour, co-chair of SC for EUMDS Registry programme - an academic study funded by Novartis via University of Nijmegen and for which no direct funding is received by Professor Bowen personally, and he has attended Advisory Boards for various agents, for which he is paid honoraria.

Rider on responsibility for report

The views expressed in this report are those of the authors and not necessarily those of the NIHR HTA Programme. Any errors are the responsibility of the authors.

This report should be referenced as follows

Craig D, Neilson A, Rose M, Wade R, Stirk L, Rodriguez-Lopez R, Bowen D, Woolacott, N. Ruxolitinib for the treatment of myelofibrosis: A Single Technology Appraisal. CRD and CHE Technology Assessment Group, 2012.

Contributions of authors

Ros Wade and Nerys Woolacott wrote the clinical effectiveness sections of the report. Dawn Craig, Aileen Neilson and Micah Rose wrote the cost effectiveness sections of the report and conducted the economic analyses. Lisa Stirk and Rocio Rodriguez-Lopez wrote the sections on the search strategies. Professor David Bowen provided clinical advice and commented on drafts of the report.

Table of Contents

4.2 Critique of trials of the technology of interest, their analysis and interpretation 33

Table of Tables

Table of Figures

List of abbreviations

AE	Adverse event
Allo-HSCT	Allogeneic haematopoietic stem cell transplantation
AML	Acute myeloid leukaemia
BAT	Best available therapy
BCSH	British Committee for Standards in Haematology
CEA	Cost-effectiveness analysis
CHMP	Committee for Medicinal Products for Human Use
CI	Confidence interval
CML	Chronic myeloid leukaemia
COMFORT	Controlled myelofibrosis study with oral JAK inhibitor treatment
CSR	Clinical study report
CT	Computed tomography
DIPSS	Dynamic International Prognostic Scoring System
EMA	European Medicines Agency
EORTC-QLQ-C30	European Organisation for Treatment of Cancer Quality of Life Questionnaire
EQ-5D	EuroQol-5D
ERG	Evidence Review Group
ET	Essential thrombocythaemia
FACT-Lym	Functional Assessment of Cancer Therapy - Lymphoma
FDA	Food and Drug Administration
HMRN	Haematological Malignancy Research Network
HR	Hazard ratio
HRQoL	Health related quality of life
ICER	Incremental cost-effectiveness ratio
IPSS	International Prognostic Scoring System
IWG-MRT	International Working Group for Myelofibrosis Research and Treatment
LT	Leukaemic transformation
MDACC	MD Anderson Cancer Center
MF	Myelofibrosis
MFSAF	Myelofibrosis Symptom Assessment Form
MRI	Magnetic resonance imaging
MS	Manufacturer's submission
NHL	Non-Hodgkin lymphoma

NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
NR	Not reported
PET-MF	Post-essential thrombocythaemia myelofibrosis
PGIC	Patient's Global Impression of Change
PMF	Primary myelofibrosis
PV	Polycythaemia vera
PPV-MF	Post-polycythaemia vera myelofibrosis
QALY	Quality-adjusted life year
RBC	Red blood cell
RCT	Randomised controlled trial
SAE	Serious adverse event
SCT	Stem cell transplant
SPC	Summary of product characteristics
TSS	Total Symptom Score

1. SUMMARY

1.1 Critique of the decision problem in the manufacturer's submission

The objective of the NICE scope was that ruxolitinib be appraised within its licensed indication for the treatment of myelofibrosis (MF). The European Medicines Agency (EMA) granted marketing authorisation of ruxolitinib for the treatment of disease-related splenomegaly or symptoms in adult patients with primary myelofibrosis (PMF) (also known as chronic idiopathic myelofibrosis), post polycythaemia vera myelofibrosis (PPV-MF) or post essential thrombocythaemia myelofibrosis (PET-MF). However, the manufacturer's submission (MS) addresses the use of ruxolitinib in patients with intermediate-2 or high-risk MF, a narrower, higher risk population, which none the less reflects the likely use of ruxolitinib in clinical practice in the UK.

The manufacturer's decision problem matches the NICE scope in stating the comparator is 'standard therapy without ruxolitinib'. In patients refractory to the largely ineffective therapies available 'standard therapy without ruxolitinib' may well comprise no treatment (other than transfusional support). Given the lack of any one clearly effective treatment for MF this comparator seems appropriate. However, the therapies that comprise 'best available therapy' (BAT) in the MS can be questioned as it does not include all the treatments listed in the NICE scope, and does include lenalidomide, which is very expensive and rarely used in the UK.

The outcome measures specified in the NICE scope were very general: symptom relief (including pain and fatigue); overall survival; progression-free survival; response rate; changes in body weight; adverse effects of treatment; and health related quality of life (HRQoL). The manufacturer modified these to more closely reflect the clinical trials (and effects) of ruxolitinib. Most notably spleen size reduction (as a measure of response rate) is the first outcome stated in the MS decision problem. A ≥50% reduction in palpable spleen length is a criterion for 'clinical improvement' according to the International Working Group for Myelofibrosis Research and Treatment (IWG-MRT) consensus criteria for treatment response in myelofibrosis. However, other criteria for demonstrating 'clinical improvement', defined by the IWG-MRT relate to reductions in the haematological symptoms of MF; these were only assessed in terms of adverse events (AEs). Importantly, treatment response was not assessed against complete remission or partial remission criteria defined by the IWG-MRT.

1.2 Summary of clinical effectiveness evidence submitted by the manufacturer

The evidence presented in the MS was derived mainly from two multicentre, randomised controlled trials (RCTs) (COMFORT-II compared ruxolitinib with best available therapy (BAT) and

COMFORT-I compared ruxolitinib with placebo) and an uncontrolled Phase I/II trial. These trials demonstrated that ruxolitinib confers significant benefits in terms of spleen size reduction and improvement in symptom burden. In the COMFORT-II trial, a reduction in spleen volume of ≥35% was achieved in 28% of ruxolitinib-treated patients compared with 0% in the BAT group ($p < 0.001$) at 48 weeks, and there was a mean change in spleen volume of −30.1% versus +7.3% ($P < 0.001$). Median time to response was 12.3 weeks in the ruxolitinib group and responses were sustained, with 80% of patients in the ruxolitinib group still having a response at a median follow-up of 12 months. Similar results were reported from COMFORT-I for the comparison with placebo. Ruxolitinib provided significant improvements in MF-associated symptoms and HRQoL compared with worsening of most symptoms and HRQoL in the patients treated with BAT.

There was no evidence of an improvement in progression-free survival with ruxolitinib (70%) compared with BAT(74%).

A significant increase in overall survival was observed for ruxolitinib over placebo in the COMFORT-I trial at 51 weeks (hazard ratio (HR) 0.5 (95% confidence interval (CI) 0.25 to 0.98) and 102 weeks (HR 0.58; 95% CI 0.36 to 0.95, p = 0.03) and the survival benefit compared with BAT reached borderline significance at 112 weeks in COMFORT-II (HR 0.52; 95% CI 0.27 to 1.00). Neither RCT was powered to detect significant differences in overall survival between treatment groups. Furthermore, after the randomised phase of the trials, long-term follow-up was confounded by the permitted crossover from placebo/BAT to ruxolitinib, potentially diluting the treatment differences. The results at the later time points are subject to uncertainty due to the small number of patients at risk in the analyses. In addition, survival data for a cohort of patients from the Phase I/II trial showed a 69% survival rate at 32 months follow-up.

Safety data were consistent across the three studies and indicated that ruxolitinib is generally well tolerated. The incidence of AEs leading to discontinuation in both RCTs was low in the ruxolitinib groups (8% and 11%) and comparable to those for the control groups. Serious AEs (SAEs) were reported in approximately 30% of ruxolitinib-treated patients in both studies and this was comparable to the incidence of SAEs reported for BAT. All individual SAEs in the ruxolitinib groups occurred at an incidence of 5% or less. Anaemia was the most frequently reported grade 3 or 4 AE in both treatment groups in the two RCTs, reflecting the frequent manifestation of severe anaemia as part of the underlying disease, but was effectively managed with red blood cell transfusions. Thrombocytopenia was the only other grade 3 or 4 AE reported in 8% of patients or more (in either treatment group) in both studies and was effectively managed by dose reductions or interruptions.

1.3 Summary of the Evidence Review Group's (ERG) critique of clinical effectiveness evidence submitted

The MS included a reasonably good quality systematic review: the search strategy was appropriate, comprehensive and well documented, inclusion screening was done in duplicate to reduce error and bias, adequate data were presented for the included studies and quality assessment was appropriate. However, the systematic review methods were poorly reported in the MS. The flow chart of the study selection process was incorrect. However, no relevant studies of ruxolitinib were overlooked: all three studies presented to the licensing authorities (FDA and EMA) were included.

The two RCTs comparing ruxolitinib with BAT and placebo were appropriate for the decision problem and were both of good quality. In addition, some results from a phase I/II dose finding study were reported. There is no evidence base of RCTs to use as the basis for an indirect comparison of individual therapies with ruxolitinib.

Whilst the evidence from the two good quality RCTs demonstrates that ruxolitinib is more effective than BAT and placebo at achieving a ≥35% reduction in spleen volume, the ERG believes the use of this outcome may generate an optimistic response rate. Whilst the manufacturer claims that this endpoint equates to the spleen reduction criterion for 'clinical improvement' according to the IWG-MRT consensus criteria for treatment response in myelofibrosis (≥50% reduction in palpable spleen length for patients with a palpable spleen that is at least 10 cm at baseline), there is some uncertainty about the equivalence of MRI assessment and palpation assessment, and the application of the ≥35% cut-off across all baseline spleen sizes may be inappropriate.

The other criteria for demonstrating 'clinical improvement', defined by the IWG-MRT consensus criteria for treatment response in myelofibrosis, relate to reductions in the haematological symptoms of MF. Importantly, ruxolitinib does not have a favourable effect on haematological symptoms such as anaemia and thrombocytopenia; these are in fact worsened by treatment in some patients and were assessed only in terms of their being adverse events. In addition, treatment response was not assessed against complete remission or partial remission criteria defined by the IWG-MRT.

Ruxolitinib was associated with improvements in symptom scores and quality of life; however, data were missing for a large proportion of patients for some of these quality of life and symptom improvement results. There was no justification for the missing data, therefore, the reliability and generalisability of these results is unclear.

The clinical significance of a lack of an improvement in progression-free survival with ruxolitinib compared with BAT is not discussed in the MS.

Overall survival was found to be statistically significantly better with ruxolitinib compared with placebo but not compared with BAT at a median follow-up of around one year. Although the comparison with BAT reached borderline statistical significance at 112 weeks, there are some difficulties with the interpretation of these results because the crossover from placebo or BAT to ruxolitinib was not adjusted for in the analysis. Furthermore, only a small number of patients were at risk at the later time points. Overall, the ERG considers the COMFORT-II trial data to be the most reliable in terms of survival because it includes a relevant comparator group, and fewer control group patients crossed over to ruxolitinib or discontinued from the study than in the COMFORT-I trial.

1.4 Summary of cost effectiveness evidence submitted by the manufacturer

The manufacturer submitted a state-transition Markov model, which represented the base case scenario primarily using data from the COMFORT-I/II trials. The model is used to evaluate the incremental costs and outcomes of ruxolitinib treatment compared against a BAT strategy for patients with intermediate-2 and high-risk MF, as stratified by IPSS classification. The model structure consists of health states relevant to treatment (treatment responder, treatment non-responder, discontinuation, and death) which were informed by the pivotal trials with the simulation of longer-term survival being informed by data from a phase I/II trial in a cohort of patients treated with ruxolitinib. Treatment response as defined by a ≥35% reduction in spleen volume was the main outcome measure in the COMFORT trials and also the driver of the model structure. HRQoL was an endpoint in the COMFORT trials assessed with the EORTC QLQ-C30, but was not used as the basis to derive health utility values for the model. HRQoL estimates were instead drawn from a previous NICE appraisal of eribulin[1] in patients with metastatic breast cancer to inform the base case analysis. Resource use was primarily drawn from the COMFORT-II trial and augmented using other external data sources: the manufacturer's own assumptions and clinical experts' opinion.

The manufacturer's economic analysis suggests that the plausible incremental cost-effectiveness ratio (ICER) for ruxolitinib versus a BAT strategy was £73,980 in the base case. This finding is consistent across the probabilistic analysis (at a willingness-to-pay for a QALY of £30,000, ruxolitinib had a 0% chance of being cost-effective) and the vast majority of deterministic sensitivity analyses undertaken by the manufacturer showed the ICER rarely fell below £70,000. Exceptions to this finding were observed for the following sensitivity (deterministic) analyses including:

- +/- 20% variation in base case utility values for responders (ICER = £67,444 to £83,144)
- +/- 20% variation mortality risk for intermediate-2 risk MF ruxolitinib patients (ICER = £69,330 to £78,430)

1.5 Summary of the ERG's critique of cost effectiveness evidence submitted

The economic analysis presented by the manufacturer was inadequate to fully address the decision problem specified in NICE's scope. The structure of the model, although accommodating several key clinical outcomes in the short-term, does not fully characterise the progressive nature of MF nor adequately capture all symptomatic and HRQoL aspects of the disease. The model hinges on using treatment effectiveness defined as ≥35% reduction in spleen volume and the long-term model has been developed without consideration of disease progression. In addition, there is an absence of evidence concerning the long-term survival, discontinuation, costs and outcomes (including HRQoL, symptom burden) for this population. Further, the 35 year time horizon over which the model is evaluated seems implausible for this population. The MS reported results of several sensitivity analyses, including probabilistic analysis. However, these analyses do not overcome the basic flaws in the model (they rely on the same inadequate assumptions) and it is the opinion of the ERG that the data and structural uncertainty within the model is under-represented. All of the further analysis undertaken by the ERG suggests that the base case ICER presented in the MS is likely to represent the best case scenario; however given the lack of disease progression incorporated into the model it is very difficult for the ERG to draw any conclusions on the most plausible ICER.

1.6 ERG commentary on the robustness of evidence submitted by the manufacturer

1.6.1 Strengths

The evidence presented for the effectiveness of ruxolitinib was identified through a systematic review and comprised two good quality RCTs, with supportive data from one phase I/II uncontrolled trial. The effectiveness of ruxolitinib was compared with relevant comparators (BAT) in one of the RCTs (the COMFORT-II trial), which represents the best available evidence on ruxolitinib.

The de-novo model used was generally transparent. The model contained the functionality to assess the impact of changing parameter values and uncertainties on the ICER, and included a number of built-in additional scenarios and sub-groups.

1.6.2 Weaknesses and areas of uncertainty

Whilst the EMA granted marketing authorisation of ruxolitinib for the treatment of disease-related splenomegaly or symptoms in adult patients with MF, the RCTs of ruxolitinib were conducted in

patients with intermediate-2 or high-risk MF, therefore its clinical effectiveness has only been assessed in this narrower, higher risk population. The trials excluded patients with a platelet count $<100x10^9$/L although the licence permits the treatment of those with a lower platelet count of $<50x10^9$/L, albeit at a lower starting dose.

The primary outcome measure in both RCTs was the proportion of patients achieving a 35% or more reduction in spleen volume from baseline. Splenomegaly is only one symptom of MF; other symptoms include haematological symptoms, such as anaemia and thrombocytopenia, which are worsened with ruxolitinib treatment in some patients, at least in the short term.

The effect of ruxolitinib on MF symptoms was assessed in the placebo controlled COMFORT-I trial using the modified Myelofibrosis Symptom Assessment Form (MFSAF) version 2 and the Patient's Global Impression of Change (PGIC) instrument. However, symptom control has not been compared between patients taking ruxolitinib and patients taking BAT for MF.

The MS clearly states that the RCTs were not designed to be sufficiently powered to detect statistically significant differences in overall survival. Furthermore, long term follow-up data are confounded by the crossover of placebo and BAT patients to ruxolitinib treatment, whilst the analysis retained patients in their randomised treatment group. In addition, the small number of patients at risk at the later time points increases the uncertainty around the survival results.

There are a number of issues in the manufacturer's de novo economic evaluation:

- structural limitations
- lack of robust survival data
- no disease progression
- definition and measurement of response
- lack of robust utility data for the MF population
- not all comparators are included and of those included not all are used in a UK context
- complications are not incorporated appropriately.

1.7 Summary of exploratory and sensitivity analyses undertaken by the ERG

The ERG conducted a number of exploratory analyses to illustrate uncertainty in the appropriateness of model assumptions and to demonstrate how alternative assumptions affect the ICER of ruxolitinib. The ERG felt that the omission of leukaemic transformation (LT) from the base case model was not justified as over 20% of MF patients die from LT, so each exploratory and sensitivity analysis

conducted by the ERG includes analysis with equivalent rates of LT in addition to analysis without LT. The analyses undertaken by the ERG included:

- survival assumptions
    - ICER ranged from £74,274 to £79,303
- definition of response criteria
    - ICER ranged from £79,536 to £90,557
- discontinuation rates
    - ICER ranged from £74,616 to £88,622
- utility values
    - ICER ranged from £97,105 to £110,325
- resource use and cost
    - ICER ranged from £75,141 to £80,874

Generally the result of alternative values being substituted was an increase in the ICER. The ERG believes that LT should be included in the base case which indicates that most ICERs using plausible values from the COMFORT-I and II trials are over £80,000 with some values over £90,000. The general limitations of the model should be considered alongside these ICERs.

Beyond the analyses that were undertaken by the ERG, there were a number of analyses that were not feasible to conduct given available data and resources. The ERG felt that the assumption of perpetual treatment irrespective of continued effectiveness was illogical; however did not investigate alternative decision rules for at what point patients would discontinue due to lack of effectiveness. The introduction of such a rule is likely to increase the ICER.

The ERG found the handling of splenectomy within the model inappropriate. It is handled as a complication with a utility decrement and a cost assigned when it occurs. As constructed, individuals who have splenectomies remain in the non-responder state. It is likely that splenectomy increases mortality risk, and increases quality of life after the operation due to relief of splenomegaly related symptoms. The ERG feel that this should have been a state in the model rather than a complication.

The ERG also has significant reservations about the handling of other complications in the model. Only non-responders are subject to complications. Among complications listed in the 'other complication' classification within the model are infection and sepsis, which are not completely dependent on splenomegaly for their occurrence. These types of complications are relevant for both responder groups. The addition of complications for responders will increase the ICER.

It is likely that the impact of these additional analyses would benefit BAT rather than ruxolitinib. Due to the considerable uncertainty surrounding the estimates of the base case ICER the ERG conducted an analysis of an alternative scenario that executed the following changes:

- reduced the time-horizon of the model to 15 years;
- allowed equivalent rates (3.6%) of LT between ruxolitinib and BAT;
- allowed transfusion dependence;
- mapped utilities from Roskell et al.[2] in a MF population;
- and used the survival HR from long term COMFORT-II data (manufacturer's response document).

Including all of the above changes simultaneously to the model increases the ICER from £73,980 to £148,867. Given the considerable uncertainty surrounding the appropriateness of the model structure and parameters, the ICER of £148,867, whilst plausible, is still highly uncertain.

1.8 Conclusions or key issues

Evidence from two good quality RCTs demonstrates that ruxolitinib is effective at reducing splenomegaly and its associated symptoms. For patients who can tolerate ruxolitinib and remain on therapy, the evidence suggests that splenomegaly and its associated symptoms can be reduced. However, patients with MF without symptoms related to splenomegaly are less likely to benefit from ruxolitinib, whose primary treatment effect is to reduce spleen volume. Importantly, haematological symptoms of MF (in particular anaemia and thrombocytopenia) are worsened by ruxolitinib in some patients, at least in the short term, requiring dose interruptions and reductions, as well as blood transfusions. There is no evidence of any improvement in progression-free survival with ruxolitinib. There is some evidence that overall survival may be increased with ruxolitinib, although these data are uncertain.

The model presented in the MS does not fully capture disease progression. In addition to the structural issues, some of the underlying modelling assumptions are clinically dubious. The additional analyses undertaken by the ERG showed that the majority of plausible modifications to the model inputs resulted in an increase in the ICER. The alternative scenario presented demonstrates the effect of pooling a number of plausible modifications to undertake an alternative scenario. This scenario more than doubles the ICER presented by the manufacturer. The ERG feel that the lack of disease progression captured in the model and the lack of long term data make obtaining a more robust estimate of the ICER difficult. It is however very likely that the base case ICER presented by the manufacturer represents a best case scenario.

2 BACKGROUND

2.1 Critique of manufacturer's description of underlying health problem.

The description of the aetiology, epidemiology and treatment of MF is generally adequate. However, whilst the debilitating symptoms of MF and their effects on quality of life are summarised correctly, the impression given is that all symptoms are secondary to splenomegaly; the fact that anaemia and thrombocytopaenia are primary symptoms in many patients is not mentioned.

The description of MF below is taken from the recently published British Journal of Haematology guidelines from the British Committee for Standards in Haematology (BCSH)[3]

> "The clinical features of myelofibrosis are variable and include progressive anaemia, leucopenia or leucocytosis, thrombocytopenia or thrombocytosis and multi-organ extramedullary haemopoiesis, most commonly causing hepatomegaly and symptomatic splenomegaly. Patients with advanced disease experience severe constitutional symptoms, the consequences of massive splenomegaly (pain, early satiety, splenic infarction, portal hypertension and dyspnoea), progressive marrow failure, pulmonary hypertension, transformation to leukaemia and early death."

The MS describes the three types of MF (PMF, PPV-MF and PET-MF), which are clinically distinct disorders that share molecular and pathological similarities, the differences between them is clearly stated and the role of an over-activation of the JAK/STAT signalling pathway is described in detail.

The impact of MF on HRQoL is outlined with evidence cited that, as assessed using the baseline EORTC QLQ-C30 questionnaire, scores for MF closely matched those baseline scores previously reported for patients with recurrent/metastatic cancer or acute myeloid leukaemia (AML).[4]

Whilst the description appropriately draws attention to the importance of fatigue in reducing patients' quality of life, the link between anaemia and fatigue is not stated explicitly.

The MS states that as MF continues to advance, patients are at increased risk of evolution to AML and that approximately 20% of patients die following disease transformation to AML. The MS did not report the proportion of patients who suffer transformation to AML. It has been reported that across all severities of MF, LT will be the cause of death of 15% of patients within 5 years.[5] The rate of LT

at 10 years in patients classified as high-risk or intermediate-2 risk according to DIPSS is approximately 60% and 20% respectively.[6]

The mortality risk of patients with MF is detailed appropriately in the MS. The MS usefully summarises the various prognostic scoring systems. The BCSH guidelines indicate that the DIPSS Plus is the most relevant to clinical practice, but in the trials of ruxolitinib the IPSS is used (Table 2.1). The MS does not provide any information on the distribution of the different risk groups in the UK. It should also be noted that the product licence for ruxolitinib is not framed in terms of these levels of risk: all levels of risk are covered by the product licence provided patients have splenomegaly or symptoms. However, as stated earlier the BCSH guidelines suggest that ruxolitinib is suitable for patients with profound constitutional symptoms, which are usually associated with massive splenomegaly.

Table 2.1 The International Prognostic Scoring System used in myelofibrosis

Risk group	Number of factors	Median survival (years)
International Prognostic Scoring System for PMF[a]		
Low	0	11.3
Intermediate-1	1	7.9
Intermediate-2	2	4
High	≥ 3	2.3
Dynamic International Prognostic Scoring System[b]		
Low	0	—
Intermediate-1	1 or 2	14.2
Intermediate-2	3 or 4	4
High	5 or 6	1.5
DIPSS-Plus risk categories in PMF[c]		
Low	0	15.4
Intermediate-1	1	6.5
Intermediate-2	2 or 3	2.9
High	≥ 4	1.3

[a]Cervantes et al. 2009;[7] [b]Passamonti et al. 2010;[6] [c]Gangat et al. 2011[8]
PMF, primary myelofibrosis; DIPSS, Dynamic International Prognostic Scoring System.

The MS stated that all three types of MF are very rare disorders. On 3 April 2009, Jakavi® (ruxolitinib) was designated as an orphan medicinal product (EU/3/09/620) by the European Commission for the treatment of PMF, PPV-MF and PET-MF.[9]

The MS reported that the prevalence of primary MF has been estimated to be 2.7 per 100,000 population,[10] with MF secondary to PV or ET affecting 0.1 per 100,000 population,[11] and an estimated annual incidence of MF as 0.34–0.76 per 100,000,[12-14], equating to approximately 187– 420 individuals diagnosed with MF in England and Wales per year. These incidence and prevalence estimates reported in the MS appear reasonable. However it is unclear whether they are representative of the UK population. The prevalence of 2.7/100,000 population PMF was cited from the Orphanet Report Series, Rare Diseases collection, May 2012.[10] The Orphanet report was based on a systematic survey of the literature in order to provide an estimate of the prevalence of rare diseases in Europe. It was not based on a survey of the UK population and may well have included no UK information. Similarly the prevalence of PPV-MF and PET-MF cited in the MS of 0.1 cases/100,000 population,[11] was derived from an EMA report, itself based on data from Norway, Iceland and Liechtenstein. It should be noted that the directly comparable data for PMF from Norway, Iceland and Liechtenstein is not reported by the EMA. The proportions of the subtypes of MF do not reflect those seen in the trials of ruxolitinib, where the patients are more evenly distributed across the sub-types.

Response to treatment is appropriately discussed in the MS in the context of the International Working Group for Myelofibrosis Research and Treatment (IWG-MRT) consensus criteria (see Table 2.2). This consensus is important when considering the relevance of the outcomes used in the trials of ruxolitinib presented by the manufacturer in later sections.

2.2 Critique of manufacturer's overview of current service provision

The MS correctly states that there is no current well-accepted standard of care or clinical pathway for the treatment of MF and quotes the recently produced BCSH guidelines on the investigation and management of primary MF, PPV-MF and PET-MF.[3] The algorithm presented in the MS is not strictly correct. According to the BCSH guidelines, ruxolitinib is recommended as part of the medical management of splenomegaly; as a second-line treatment for patients with symptomatic splenomegaly who are unresponsive to hydroxycarbamide, or the immunomodulators thalidomide with prednisolone or, in patients with anaemia and adequate platelet count, lenalidomide. Ruxolitinib provides a first-line treatment option for the management of profound constitutional symptoms; patients with such symptoms are usually in the poor risk group.

Allogeneic haematopoietic stem cell transplantation (allo-HSCT) is mentioned in the MS but it is not considered as a comparator to ruxolitinib. The MS correctly states that allo-HSCT is the only potentially curative therapy, but that it is generally reserved for patients aged <45 years who have a poor prognosis if left untreated. In fact the BCSH guidelines have slightly broader recommendations than this in that they include a recommendation for reduced intensity allo-HSCT in patients aged over

45 years or who have a HSCT co-morbidity index ≥3, but only in those who are transplant eligible and at high or intermediate-2 risk. Across both age groups the BCSH guidelines recommend allo-HSCT (or reduced intensity allo-HSCT) especially if the patient is transfusion dependent and/or has adverse cytogenetic abnormalities. Thus, it might be reasonable to consider allo-HSCT as a comparator to ruxolitinib, although as identified in a recent survey of practice, it is only used in a small proportion of patients in the UK – less than 10%[15] and in reality might be much lower (clinicl advice ot ERG). This is discussed further in Section 3.3.

The MS states that,
"It is expected that treatment of patients with ruxolitinib will not require additional monitoring as compared with current practices. Indeed, because current therapies are associated with severe side effects, requiring close observation, it is likely that treatment with ruxolitinib will lead to a reduction in resource costs compared with the alternatives."

Given that ruxolitinib therapy is associated with significant anaemia and thrombocytopaenia, which the MS states are controlled by dose reductions and treatment interruptions,[16] the ERG feel that it is unclear if the resource use reductions discussed in the MS would be realised in clinical practice.

Through the MS the status of three therapies used in the management of MS is unclear: blood transfusions, splenectomy and splenic irradiation. None of these therapies is clearly described nor is any treated as comparators of ruxilitinib or included as a component of BAT. In the economic model splenectomy and splenic irradiation, and to a lesser extent, transfusions are incorporated as complications of MF, with only the disutilities or costs of treatment included, whilst any benefits are not.

Table 2.2 International Working Group for Myelofibrosis Research and Treatment consensus criteria for treatment response in myelofibrosis

Treatment response	Guidelines
Complete remission	• Complete resolution of **palpable splenomegaly and hepatomegaly** • Complete resolution of **disease-related signs and symptoms** • **Peripheral blood count remission** (defined as haemoglobin level ≥ 110 g/L, platelet count ≥ 100 x 10^9/L, and ANC ≥ 1.0 x 10^9/L. In addition, all 3 blood counts should be no higher than the ULN) • **Normal leukocyte differential** (including disappearance of nucleated RBCs, blasts and immature myeloid cells in the peripheral smear, in the absence of splenectomy) • **Bone marrow histologic remission** (defined as the presence of age-adjusted normocellularity, ≤ 5% myeloblasts, and an osteomyelofibrosis grade ≤ 1)
Partial remission	• Requires all of the above criteria for CR except the requirement for bone marrow histologic remission. (Repeat bone marrow biopsy is required in the assessment of PR and may or may not show favourable changes that do not fulfil the criteria for CR)
Clinical improvement[a]	Requires **one** of the following in the absence of disease progression and CR/PR assignment • **≥ 50% reduction in palpable splenomegaly** of a spleen that is at least 10 cm at baseline or a spleen that is palpable at more than 5 cm at baseline becomes not palpable • **≥ 20 g/L increase in haemoglobin level or becoming transfusion independent** (applicable only for patients with baseline haemoglobin level < 100 g/L) • **≥ 100% increase in platelet count and an absolute platelet count of ≥ 50 x 10^9/L** (applicable only for patients with baseline platelet count < 50 x 10^9/L) • **≥ 100% increase in ANC and an ANC of ≥ 0.5 × 10^9/L** (applicable only for patients with baseline ANC < 1 x 10^9/L)
Progressive disease	Requires **one** of the following: • **Progressive splenomegaly (**defined as appearance of a previously absent splenomegaly that is palpable at > 5 cm below the left costal margin or ≥ 100% increase in palpable distance for baseline splenomegaly of 5–10 cm, or ≥ 50% increase in palpable distance for baseline splenomegaly of > 10 cm) • **Leukaemic transformation** (confirmed by a bone marrow blast count of at least 20%) • **Increase in peripheral blood blast percentage** of ≥ 20% that lasts for ≥ 8 weeks
Stable disease	None of the above
Relapse	Loss of CR, PR or CI

Tefferi et al. 2006.[17]
[a]Response is validated if it lasts at least 8 weeks
CR, complete remission; PD, progressive disease; PR, partial remission; RBC, red blood cell; SD, stable disease; ULN, upper limit of normal.

3 CRITIQUE OF MANUFACTURER'S DEFINITION OF DECISION PROBLEM

3.1 Population

The population in the MS matched that specified in the NICE scope:
"Adults with disease-related splenomegaly or symptoms of primary myelofibrosis or myelofibrosis secondary to polycythaemia vera and essential thrombocythaemia".

This is within the licensed indication of ruxolitinib for the treatment of disease-related splenomegaly or symptoms in adult patients with PMF, PPV-MF or PET-MF.[9]

It should be noted that, although the stated population matches the NICE scope, the evidence presented in the MS is derived from clinical trials whose populations represent only a subset of the licensed population (see Section 4.2.1). However, whilst the licence is very broad, it is likely that in clinical practice patients treated with ruxolitinib will be those recommended in the BCSH guidelines,[3] i.e. patients with profound constitutional symptoms, which are usually associated with massive splenomegaly. The population considered in the economic model reflects that of the clinical trials (intermediate-2 and high-risk patients). The ERG sought clarification regarding the generalisability of the clinical trials to the licensed population. In their response the manufacturer stated that,

"..the efficacy of ruxolitinib is based on spleen size and not risk group. We would therefore expect ruxolitinib to be effective regardless of risk group and the trial data to be generalisable to the full licensed population, as per the summary of product characteristics (SPC)."

Patients recruited into the clinical trials had varying spleen size and the relationship between response to ruxolitinib and baseline spleen size has not been evaluated in the MS. This is discussed further in Section 4.2.

3.2 Intervention

The manufacturer's decision problem specifies only 'ruxolitinib', which matches the NICE scope. The licensed dose of ruxolitinib is 5 – 25 mg twice daily. The recommended starting dose is 5, 15 or 20 mg twice daily according to the patient's platelet count, with cautious upward titration. The maximum dose is 25 mg twice daily.

The doses used in the trials which comprise the evidence base for the submission and those used in the economic model are discussed in Sections 4 and 5. 3.3

Comparators

The manufacturer's decision problem matches the NICE scope in stating the comparator is 'standard therapy without ruxolitinib'. Given the lack of any one clearly effective treatment for MF[3] this comparator seems appropriate. In patients refractory to the largely ineffective therapies available 'standard therapy without ruxolitinib' may well comprise no treatment.

The NICE scope also stated that standard therapy includes any combination of splenectomy or splenic irradiation, chemotherapy, immunomodulators, erythropoietin or red blood cell transfusion, corticosteroids, androgenic steroids, allopurinol, dietary advice or no treatment, and that standard includes haematopoietic stem cell transplantation, for the subgroup of people with MF for whom it is appropriate. To further explore the possibility of a more specific comparator for ruxolitinib the ERG investigated trials of treatments used in MF using the manufacturer's search results (as reported in the manufacturer's response document) and also a pragmatic search of comparators conducted by the ERG, which identified studies of CEP-701, thalidomide, lenalidomide, pomalidomide, pacritinib (SB1516), histone deacetylase inhibitor Sb939, allo-HSCT, imatinib mesylate, splenic irradiation, hydroxyurea and interferon. However, almost all of these individual therapies have been studied in uncontrolled trials only.

Two RCTs were identified.[18-19] One RCT compared thalidomide with placebo. It demonstrated that tolerance of thalidomide was a significant problem (dropout from the trial was high) and treatment benefit was modest: no difference in the number of patients with anaemia or the number of blood transfusions. There was a statistically significant benefit in terms of spleen size but only in that the *increase* in size was less than in the placebo group. The second RCT compared pomalidomide +/- prednisone with placebo in patients with MF associated anaemia. Overall 20/84 patients achieved a response and 15 of the 20 responses involved the patient becoming transfusion independent. It is noteworthy that although more than half of the patients in each treatment group had splenomegaly of >10 cm no patient achieved a response in terms of spleen reduction.

These two small trials with their limited signs of any efficacy in MF indicate that there is no evidence base of RCTs to use as the basis of any form of indirect comparison of individual therapies with ruxolitinib. A full review of uncontrolled trials of the other treatments was not possible within the limited timescale and resource of this STA.

The MS dismisses allo-HSCT as a comparator, stating "There is insufficient evidence to allow a comparison between ruxolitinib and SCT in individuals for whom SCT is appropriate."

The ERG concurs with this statement. Based on the manufacturer's searches, the evidence base for allo-HSCT is comprised of only uncontrolled trials and case series, with the exception of an RCT comparing bone marrow sourced allo-HSCT with blood sourced allo-HSCT.[20] Furthermore, this therapy is not uniform, with a number of variants in terms of the conditioning regimens utilised to eradicate the tumour cells and the source of the cells blood or bone marrow.[3] On the other hand the RCTs of ruxolitinib have been conducted only in patients in whom allo-HSCT was either not appropriate or not available at the time. Therefore, it is not possible to make a meaningful comparison of these therapies.

It should also be noted that allo-HSCT might not be considered a valid comparator to ruxolitinib because, as identified in a recent survey of practice, it is only used in a small proportion of patients in the UK: less than 10% according to a recent survey,[15] and even this may be an overestimation according the ERG's clinical advisor.

The comparators used in the trials which comprise the evidence base for the submission and in the economic model are discussed in Sections 4 and 5.

3.4 Outcomes

The outcome measures specified in the NICE scope were very general: symptom relief (including pain and fatigue); overall survival; progression-free survival; response rate; changes in body weight; adverse effects of treatment; and HRQoL. The manufacturer modified these to more closely reflect the clinical trials (and effects) of ruxolitinib. Most notably spleen size reduction (as a measure of response rate) is the first outcome stated in the MS decision problem. Other outcomes addressed were impact on symptom burden, overall survival, progression-free survival, changes in body weight, AEs and HRQoL.

The primary outcome in the MS was the proportion of patients achieving a 35% or more reduction in spleen volume from baseline as assessed by magnetic resonance imaging (MRI) or computed tomography (CT) (as a measure of response rate). The ERG has identified some issues with this outcome. Firstly, in the MS the manufacturer claims that a 35% reduction in spleen volume measured by MRI or CT equates to a 50% reduction in palpable splenomegaly, which is one of the criteria for clinical improvement in the IWG-MRT consensus criteria for treatment response in myelofibrosis.[17] This is not a generally accepted assumption but is based on data from 24 patients in the phase I/II

trial. Secondly, the IWG's criterion specifies, "≥ 50% reduction in palpable splenomegaly of a spleen that is at least 10 cm at baseline or a spleen that is palpable at more than 5 cm at baseline becomes not palpable". This is changed to be "≥35% reduction in spleen volume as measured using MRI". However, in the MS (and supporting trials) this 35% reduction cut-off is applied across all baseline spleen volumes, which may not be appropriate. This is discussed further in Section 4.2.

It should be noted that the IWG-MRT criteria include a number of symptom responses (see Table 2.2), complete or partial remission require a complete resolution of disease-related signs and symptoms, complete resolution of palpable splenomegaly and hepatomegaly as well as remission in terms of blood counts and a normal leucocyte differential. A 50% reduction in palpable splenomegaly is classified as a "clinical improvement" only.

Across the publications identified by the ERG's searches for studies of other comparator treatments for MF the ERG identified a number of outcome measures used, including IWG or EUMNET criteria, anaemia and transfusion requirement, reduced spleen volume alongside anaemia and other blood counts, bone marrow response and antiangiogenesis response.

The comparator trials (RCTs and others) make it clear that different therapies affect different aspects of MF and are assessed therefore using different outcome measures. The control of anaemia is clearly a major objective in some patients and for some therapies. The success of ruxolitinib in terms of spleen size reduction has to be interpreted in the context of its effect (or lack thereof) on other aspects of MF, as does any comparison with other therapy.

3.5 Other relevant factors

The MS states that MF is a highly rare orphan disease and is generally diagnosed in individuals over 60 years of age. Therefore, patients with MF may be less likely to receive extensive cancer treatment because of their age, and may also be at risk of receiving poorer treatment because of the rarity of their disease. On 3 April 2009, Jakavi was designated as an orphan medicinal product (EU/3/09/620) by the European Commission for the treatment of PMF, PPV-MF and PET-MF.[9]

4 CLINICAL EFFECTIVENESS

This section contains a critique of the methods of the systematic review presented in the MS, followed by a description and critique of the trials included in the review, including a summary of their quality and results. The ERG's conclusions on the clinical effectiveness of ruxolitinib for the treatment of MF are presented at the end of this section.

4.1 Critique of the methods of the review

The MS described a systematic review of interventions for patients with MF, including one trial comparing ruxolitinib with other treatments commonly used to treat MF, one trial comparing ruxolitinib with placebo and one dose finding study.

4.1.1 Search strategy

The manufacturer's submission described the search strategies used to identify relevant clinical effectiveness studies on the use of ruxolitinib for the treatment of PMF, PPV-MF and PET-MF. Strategies were only briefly described in the main body of the submission, however full details were provided in the Appendices.

The electronic databases MEDLINE, MEDLINE In Process, EMBASE, BIOSIS, Cochrane Database of Systematic Reviews, Cochrane Central Register of Controlled Trials and the Database of Abstracts of Reviews of Effects (DARE) were used to identify clinical studies on the use of ruxolitinib for the treatment of PMF, PPV-MF and PET-MF. In addition to this, abstracts of conference proceedings, clinical trial registers and The National Institute for Health and Clinical Excellence (NICE) website were reviewed.

Searches were conducted on 27 July 2011. A search update was also performed on 19 March 2012. Search strategies for each database were documented in Tables D1 - D4, as stated in the MS. The searches covered the period 1960 - 19 March 2012, were not limited by language of publication, and excluded animal-only studies.

Overall the searches were appropriate, comprehensive and well documented, and included the use of both subject indexing terms and free text searching. Field searching, Boolean operators and truncation were used where required. All the required databases were searched, as well as the additional databases BIOSIS and DARE.

The MS states that the search strategy was designed to be broad enough to capture all potential trials and other prospective studies of MF. An additional broad search for the drug name ruxolitinib (and related terms) may also have been useful in identifying drug adverse effects in any diseases. The search strategies used in the manufacturer's submission were limited to trials and multicentre studies only, however a search for other study designs such as cohort or case control studies may also have provided useful supplementary information on the safety profile of ruxolitinib.

For both the clinical and cost effectiveness searches, study design limits were applied to the Cochrane Library databases search strategies in order to be consistent with the usage of terms in the MEDLINE and EMBASE searches. These filters are however redundant as the Cochrane Library databases are already limited by study design. Further methodological filters are therefore superfluous, and may exclude potentially useful records.

4.1.2 Inclusion criteria

Two reviewers independently assessed the records against the inclusion and exclusion criteria, reducing the potential for error or bias. The inclusion criteria stated for the selection of studies for the systematic review in the MS were very broad, not specifying the intervention, comparators or outcomes of interest. The inclusion criteria relating to the population of interest appear to be appropriate being in line with both the decision problem and the marketing authorisation of ruxolitinib. Note only patients with PMF, patients with MF secondary to PV or ET and patients with PMF who were not suitable for allo-HSCT were eligible. Exclusion criteria were secondary MF associated with, for example, tuberculosis. Other diseases that were excluded were CML and antecedent PV or ET without MF.

The review criteria for intervention, outcomes and study design were appropriate to capture a wide range of trials appropriate for a systematic review of all treatments for MF. It appears that at the study selection stage these were narrowed such that only studies of ruxolitinib were included in the review.

Whilst only English language articles were selected for the review, creating a potential for language bias, it is unlikely that any relevant studies of ruxolitinib were excluded on the basis of language of publication. Although publication type was not specified, two conference abstracts were excluded from the review based on their publication type, suggesting that only fully published studies were included, increasing the potential for publication bias.

A flow chart of the study selection process was presented in the MS. Within this flow chart there were some anomalies:

- The reason for exclusion of 32 articles was "intervention", however, there were no inclusion/exclusion criteria relating to intervention, therefore, the ERG asked for clarification regarding this. The manufacturer responded giving details of the articles excluded at the full paper stage. None of the excluded articles related to the review question, so their exclusion was appropriate.

- The flow chart of the study selection process also stated that 4 RCTs, 1 RCT open-label follow-up study and 12 other long-term non-comparator studies were included in the review. However, only three studies were discussed further in the MS, and it was stated that these were selected because they related to ruxolitinib. The ERG requested details of the other 14 studies that were included in the review, and the rationale for excluding these articles at this stage of the process (i.e. after applying inclusion and exclusion criteria). The manufacturer responded giving details of the articles excluded at this stage of the review. None of the excluded studies related to ruxolitinib, therefore, their exclusion was appropriate, for a review of ruxolitinib.

4.1.3 Data extraction

The MS presented adequate data from the two included RCTs and the phase I/II trial.

4.1.4 Quality assessment

A table of the quality assessment results for the two randomised controlled trials was presented as an appendix in the MS (Table D6), which included all the quality criteria specified by NICE. Quality assessment results were checked by the ERG. The MS also included a more detailed description of some of the factors likely to affect the validity of the included studies, such as the method of randomisation (Interactive Voice Response System), blinded outcome assessment for the primary outcome (central reading of MRI and CT results by a reader unaware of the treatment allocation) and comparability of baseline characteristics between treatment groups. Results of the quality assessment of the phase I/II trial were also presented as an appendix in the MS (Table D12), using appropriate criteria.

4.1.5 Evidence synthesis

The manufacturer described the results of the individual studies separately, which appears appropriate in view of the differences in study design and participant and intervention characteristics.

4.1.6 Conclusions from critique of systematic review methods

The search strategy was appropriate, comprehensive and well documented, inclusion screening was done in duplicate to reduce error and bias, adequate data were presented for the included studies and quality assessment was appropriate. However, the systematic review methods were poorly reported in the MS. The stated inclusion criteria used to select studies for the review were very broad, not specifying the intervention, comparators or outcomes of interest and it appears that only fully published English language articles were included in the review. However, no relevant studies of ruxolitinib were overlooked: all three studies presented to the licensing authorities (FDA and EMA) were included.

4.1.7 Ongoing studies

The MS states that the following studies involving ruxolitinib are currently recruiting patients with myeloproliferative neoplasms (PMF, PPV-MF and PET-MF) and have estimated study timelines that may permit data reporting in the next 12 months.

- Study NCT01317875 is a phase Ib dose-escalation study of ruxolitinib (starting dose 5 mg twice daily) in patients with low platelet counts (< 100 x 10^9/L). The estimated completion date is October 2013.
- Study NCT01433445 is a phase Ib dose-finding study to determine suitable doses of ruxolitinib and panobinostat for use in combination. The estimated completion date is 2015.
- Study NCT01558739 is a phase II open-label study to evaluate the efficacy of ruxolitinib in patients with PMF, PPV-MF or PET-MF and uses a composite endpoint of reduction in splenomegaly and/or reduction in total symptom score. It is anticipated that this UK study will generate local health resource utilisation data.

In addition, the MS lists the following ongoing studies of ruxolitinib in patients with myeloproliferative neoplasms:

- Study NCT01348490 (INCB 18424-258) is a non-randomised phase II study assessing the efficacy and tolerability of individualized dose-optimized ruxolitinib (starting dose 5 mg twice daily) in patients who have low platelet counts (50 x 10^9/L to 100 x 10^9/L). The estimated study completion date is December 2012.
- Study NCT01445769 (INCB 18424-261) is an open-label phase II study exploring ruxolitinib given at a starting dose of 10 mg twice daily, with treatment increased to 20 mg twice daily according to efficacy and safety parameters. The estimated study completion date is February 2013.

- Study NCT01340651 (INCB 18424-260) is another phase II open-label study and is evaluating the effects of a sustained release formulation of ruxolitinib on platelet count. This study has been completed.

4.2 Critique of trials of the technology of interest, their analysis and interpretation

4.2.1 Trials included in the review

Two multi-centre parallel-group RCTs of ruxolitinib were included in the MS review; COMFORT-I[21] and COMFORT-II[22]. Both trials assessed ruxolitinib at starting doses of 15 mg or 20 mg twice daily (the starting dose was dependent on baseline platelet count) in patients with splenomegaly and intermediate-2 or high-risk PMF, PPV-MF and PET-MF. In the COMFORT-I trial included patients were refractory to all other therapies and the comparator was placebo.[21] In the COMFORT-II trial included patients were or were not refractory to other therapies and the comparator was best available therapy (BAT) which could be no therapy, where appropriate.[22] Both trials were directly relevant to the decision problem; COMFORT-I reflected the decision problem where ruxolitinib was end-of-line therapy.

In addition, a phase I/II uncontrolled trial was included in the review. The main objective of this trial was to identify the most effective and well-tolerated dose of ruxolitinib.[23] The rationale for including this trial in the review was that it included more long-term data for the efficacy and safety of ruxolitinib in patients with MF. This trial included patients who were or were not refractory to other therapies and had intermediate-1, intermediate-2 or high-risk disease. The starting dose of ruxolitinib ranged from 10 mg twice daily to 200 mg once daily.

The study design and patient characteristics of the two RCTs and the phase I/II trial are summarised in Tables 4.1 and 4.2 and the results of the RCTs are summarised in Table 4.3.

4.2.1.1 Randomised controlled trials

Table 4.1 Study design and patient characteristics of the included RCTs

Study details	COMFORT-I	COMFORT-II
Location	89 sites in the United States, Canada, Australia	56 sites in Europe (included United Kingdom)
Design	Randomised, double-blind, placebo-controlled	Open-label, randomised
Duration of core study	24 weeks	48 weeks
Method of randomisation	Interactive Voice Response System; 1:1 ratio	Interactive Voice Response System; 2:1 ratio
Method of blinding (care provider, patient and outcome assessor)	Patients received matching placebo tablets, unblinding could occur after week 24; investigators were blind to treatment assignment as database was frozen until primary analysis was complete; MRI and CT scans were assessed by a central review process that was blinded to treatment	None
Intervention(s)	Oral ruxolitinib tablet 15 mg or 20 mg twice daily (n = 155)	Oral ruxolitinib tablet 15 mg or 20 mg twice daily (n = 146)
Comparator(s)	Matched placebo (n = 154)	BAT (n = 73)
Primary outcome	Proportion of patients achieving a ≥ 35% reduction from baseline in spleen volume, assessed by MRI or CT scan	
Timing of primary outcome	Week 24	Week 24 (secondary) and 48 (primary)
Secondary outcomes	Duration of maintenance of reduction in spleen volume in patients initially randomised to receive ruxolitinib, assessed by MRI or CT scan Proportion of patients who had a ≥ 50% reduction from baseline in week 24 Total Symptom Score, measured by the modified MFSAF v2.0 diary Change from baseline in week 24 Total Symptom Score, measured by the modified MFSAF v2.0 diary Overall survival HRQoL assessments using EORTC QLQ-C30 and PROMIS Fatigue scale (exploratory endpoints)	Duration of maintenance of spleen volume reduction ≥ 35% reduction from baseline Time to achieve a first ≥ 35% reduction in spleen volume from baseline Progression-free survival Leukaemia-free survival Overall survival Transfusion dependency/independency Change in bone marrow histomorphology HRQoL assessments using EORTC QLQ-C30 and FACT-Lym (exploratory endpoints)
Duration of follow-up for reported analysis	Median, 32 weeks (51 weeks for additional analysis of overall survival)	Median, 12 months (for overall survival), and 61 weeks for a pre-planned safety update)
Patient inclusion criteria	Age ≥ 18 years	
	Life expectancy of ≥ 6 months	
	Diagnosis of PMF, PPV-MF or PET-MF according to WHO criteria (2008)	
	An IPSS score of 2 (intermediate-2 risk level) or ≥ 3 (high-risk)	
	Palpable spleen measuring ≥ 5 cm below the left costal margin	
	ECOG performance status of ≤ 3 (scale of 0 to 5)	
	Peripheral blood blast count of < 10%	
	Absolute peripheral blood CD34+ cell count > 20 x 10^6/L	Platelet count ≥ 100 x 10^9/L without assistance of growth or thrombopoietic factors, or platelet transfusions. Absolute neutrophil count ≥ 1 x 10^9/L
	Disease that was resistant or refractory to available treatment or intolerant of or not candidates for such therapy	
	Disease that required treatment defined by any of the following: IPSS prognostic score ≥ 3, palpable spleen length > 10 cm, score of > 3 on at least 2 items or score of 5 on 1 item on the MFSAF v2.0	

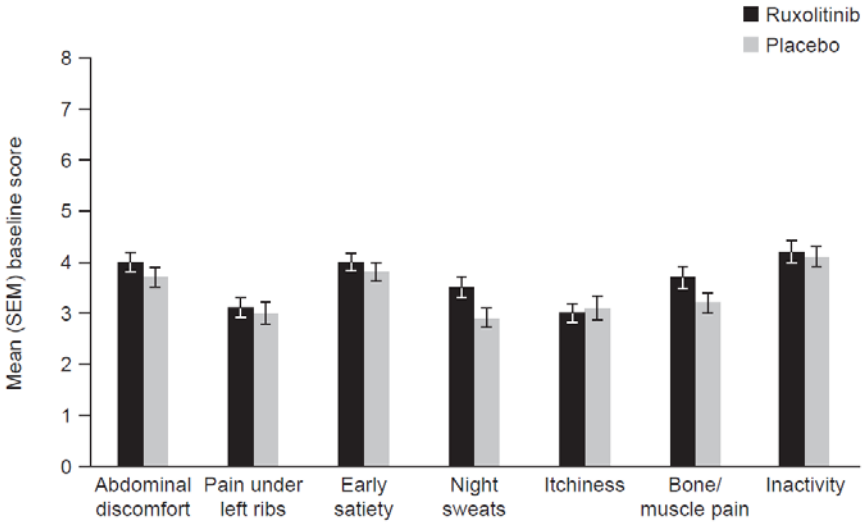
Study details	**COMFORT-I**		**COMFORT-II**	
	diary			
Patient exclusion criteria	Absolute neutrophil count ≤ 1 x 10^9/L or platelet count < 100 x 10^9/L)		History of ANC ≤ 0.5 x 10^9/L or platelet count < 50 x 10^9/L except during treatment for myeloproliferative neoplasm or cytotoxic therapy	
	Bilirubin ≥ 2 x ULN; alanine aminotransferase ≥ 2.5 x ULN; creatinine > 2.0 mg/L)			
	History of malignancy in past 5 years			
	Splenic irradiation within 12 months prior to randomisation/screening			
	Previous treatment with JAK inhibitor			
	Concurrent treatment with other prohibited medications			
			Pregnant or breastfeeding	
Characteristic	**COMFORT-I (n = 309)**		**COMFORT-II (n = 219)**	
	Ruxolitinib (n = 155)	**Placebo (n = 154)**	**Ruxolitinib (n = 146)**	**BAT (n = 73)**
Median age (range), years	66 (43–91)	70 (40–86)	67 (35–83)	66 (35–85)
Male, %	51.0	57.1	57	58
Disease type, %				
PMF	45.2	54.5	53	53
PPV-MF	32.3	30.5	33	27
PET-MF	22.6	14.3	14	19
IPSS risk status, %				
High	58.1	64.3	60	59
Intermediate-2	41.3	35.1	40	40
Prior hydroxy-carbamide use, %	67.1	56.5	75	68
Palpable spleen length, median (range), cm	16 (0–33)[a]	16 (5–34)	14 (5–30)	15 (5–37)
Spleen volume, median (range), cm^3	2598 (478–7462)	2566 (521–8881)	2408 (451–7766)	2318 (728–7701)
Platelet count, median (range), x 10^9/L	262 (81–984)	238 (100–887)	244 (–)	228 (–)
Haemoglobin				
Median (range), g/dL	10.5 (6.6−17.0)	10.5 (3.5−17.3)	–	–
< 10 g/dL, %	–	–	45	52
JAK2V617F mutation positive,%	72.9	79.9	75	67

Participants

The controlled trials of ruxolitinib were conducted only in patients with splenomegaly and intermediate-2 or high-risk MF,[21-22] therefore the clinical effectiveness of ruxolitinib has been assessed only in this narrower, higher risk population. In addition, patients with an absolute neutrophil count $\leq 1 \times 10^9$/L or platelet count $<100 \times 10^9$/L and patients suitable for allo-HSCT at the time of study enrolment were excluded from the trials, so the efficacy and safety of ruxolitinib has not been assessed in these patients. Within this narrower population, the trial inclusion criteria appear to have been appropriate, and were similar between the two trials, with the exception that patients in the COMFORT-I trial had disease that was refractory to available therapies, had side effects requiring their discontinuation, or were not candidates for available therapies, therefore, in this trial ruxolitinib was used in the second-line setting.[21]

There were no inclusion criteria related to symptoms of MF. Figure 4.1 displays participants' baseline symptom scores for the COMFORT-I trial assessed using the modified MFSAF version 2 (this is Figure 14 of the MS). The mean TSS at baseline was 18.0 for ruxolitinib-treated patients and 16.5 for placebo-treated patients (out of a potential maximum score of 60 indicating worst possible symptoms). Baseline symptom scores for the COMFORT-II trial were not reported in the MS, however 69% patients in the ruxolitinib group and 63% patients in the BAT group had constitutional symptoms at baseline, including weight loss, fever and night sweats. Therefore, fewer participants in the COMFORT-II trial appear to have had constitutional symptoms at baseline, than in the COMFORT-I trial (80.5% ruxolitinib patients and 83.6% placebo patients had night sweats at baseline, as displayed in Figure 4.1).

Figure 4.1 Modified Myelofibrosis Symptom Assessment Form version 2.0 individual symptom scores at baseline



Patients with symptoms at baseline (%):

Ruxolitinib	95.3	83.9	94.0	80.5	75.8	81.9	91.9
Placebo	94.1	80.9	91.4	83.6	74.3	82.9	92.8

The trials were conducted in the USA, Canada, Australia and Europe (including four recruiting sites in the UK); the manufacturer stated that the study findings are likely to be generalisable to the UK population of MF patients with splenomegaly and intermediate-2 or high-risk MF. However, the proportion of patients with the different subtypes of MF in the trials does not appear to reflect the prevalence data reported in the background section of the MS, which suggested PPV-MF and PET-MF were around 30 times less common than PMF, but in the trials patients with PMF make up only around 50% of the populations. The ERG asked the manufacturer to comment on the generalisability of the study findings to the UK population, given the difference in prevalence of subtypes of MF. The manufacturer stated that regardless of whether MF is primary or secondary, the disease is characterised by a clonal haemopoietic stem cell proliferation associated with a characteristic stromal pattern, a leuco-erythroblastic blood film and elevated levels of various inflammatory and pro-angiogenic cytokines; therefore, similar efficacy is not unexpected for a drug with a mode of action like ruxolitinib, regardless of MF subtype.

Intervention

In the controlled trials ruxolitinib was prescribed at a starting dose of 15 mg twice daily if baseline platelet count was $200x10^9$/litre or less or 20 mg twice daily if baseline platelet count was greater than $200x10^9$/litre. A protocol-specified dosing regimen required reductions in dose for reasons of safety (if neutropenia or thrombocytopenia developed) and permitted escalation of the dose to increase efficacy, although the dose could not exceed 25 mg twice daily.[21-22] The licensed dose of ruxolitinib is 5 – 25 mg twice daily, therefore, the use of the intervention in the trials presented in the MS was directly relevant to the decision problem.

A large proportion of participants required dose reductions or interruptions (63% of the ruxolitinib group in the COMFORT-II trial required dose reductions or interruptions, 32% of the ruxolitinib group in the COMFORT-I trial required dose interruptions; dose reductions were not reported); the effect of dose interruptions on TSS (assessed using the MFSAF version 2) was presented for patients in the COMFORT-I trial, but the effect of dose reductions or interruptions on other outcomes was not reported. The ERG asked the manufacturer for more information on the frequency and duration of dose reductions and interruptions and results for the primary outcome and survival in patients who had dose interruptions or reductions. The manufacturer stated that the primary efficacy outcome of the COMFORT-II trial was based on the measured mean dose intensity of 30.3 mg/day. This result took into account all dose interruptions and reductions and therefore the results of those patients who had interruptions and reductions were accounted for in the overall outcome. In the COMFORT-I trial the mean dose intensity was 30.82 mg/day. Therefore, these dose reductions and interruptions do not appear to have significantly reduced the mean dose intensity in the trials and furthermore, the results are likely to reflect the use of ruxolitinib in clinical practice.

Comparators

The COMFORT-II trial compared ruxolitinib with best available therapy (BAT), including observation alone (33% patients), antineoplastic agents (hydroxyurea and anagrelide; 51% patients), glucocorticoids (prednisone/prednisolone and methyprednisolone; 16% patients), anti-anaemia preparations (epoetin-alpha), immunomodulatory agents (thalidomide and lenalidomide), purine analogs (mercaptopurine and thioguanine), antigonadotropins and similar (danazol), interferons (PEG-interferon-alpha 2a and interferon-alpha), nitrogen mustard analogs (melphalan) and pyrimidine analogs (cytarabine).[22] These comparators were generally appropriate, although lenalidomide is not very rarely used in UK practice.[15] The mode of action of the comparator therapies and the symptoms they target is varied; few comparator treatments would have an effect on spleen size, with many of them targeting the haematological symptoms of MF, such as leucocytosis, thrombocytosis, cytopenias and anaemia.

The COMFORT-I trial compared ruxolitinib with placebo.[21] However, as patients in this trial were refractory to available therapies, had side effects requiring their discontinuation, or were not candidates for available therapies, there were no alternative therapies for these patients, therefore the comparator in this trials could be interpreted as a form of BAT for this population.

Outcomes

Spleen reduction

The primary outcome for both trials was the proportion of patients achieving a ≥35% reduction from baseline in spleen volume, assessed by MRI or CT scan which the manufacturer claims corresponds to a 50% reduction in palpable spleen length (a criterion for clinical improvement defined by the IWG-MRT). The justification for this primary outcome was that MRI/CT is a more robust and objective measurement of spleen size than physical examination by palpation. The phase I/II trial demonstrated that changes in spleen volume assessed by MRI paralleled changes in palpable spleen length; 24 patients receiving ruxolitinib 15 mg twice daily were assessed after six months of therapy, the median reduction in spleen volume was 33% and the median reduction in spleen length was 52%. Therefore, the 35% reduction in spleen volume assessed by MRI or CT appears appropriate, as this is a more objective measurement and can be read by blinded outcome assessors (as was done in both COMFORT trials). However, assessment of spleen length by palpation is more clinically relevant as this is how spleen size is assessed in clinical practice. Both COMFORT trials assessed both spleen volume (using MRI or CT) and palpable spleen length and compared the results (Figures 11 and 33 of the MS).

A 50% reduction in palpable spleen length is one of the criteria for demonstrating "clinical improvement" defined by the IWG-MRT consensus criteria for treatment response in myelofibrosis. However, the IWG-MRT criterion is a "≥50% reduction in palpable splenomegaly of a spleen that is at least 10 cm at baseline or a spleen that is palpable at more than 5 cm at baseline becomes non palpable" (see Table 2.2). The 35% reduction cut-off was applied across all baseline spleen volumes, despite almost a third of ruxolitinib patients in the COMFORT-II trial and a fifth of ruxolitinib patients in the COMFORT-I trial having a baseline palpable spleen length of less than 10cm.[9]

A 35% reduction in spleen volume for those patients with a smaller spleen at baseline may have little impact on patients' symptoms or HRQoL. For this reason, the emphasis on a 35% or more reduction in spleen volume as the primary outcome, above symptom relief, overall survival and HRQoL, does not appear to be appropriate.

Haematological symptoms

The other criteria for demonstrating clinical improvement defined by the IWG-MRT consensus criteria for treatment response in myelofibrosis, relate to reductions in haematological symptoms of MF (see Table 2.2). Haematological symptoms such as anaemia and thrombocytopenia were only assessed in terms of adverse events. In addition, treatment response was not assessed against complete remission or partial remission criteria defined by the IWG-MRT.

Symptom reduction

The COMFORT-I trial assessed symptom reduction using the modified Myelofibrosis Symptom Assessment Form (MFSAF) version 2, which was an appropriate tool to use. This tool is disease-specific and assesses seven symptoms of MF; abdominal discomfort, pain under the ribs on the left side, early satiety, night sweats, itchiness, bone/muscle pain and inactivity. The COMFORT-I trial assessed the proportion of patients achieving a ≥50% reduction in TSS using the MFSAF version 2 and the mean change from baseline in TSS; the 50% cut-off was chosen because a reduction of this magnitude correlated with a significant improvement in disease symptoms in the phase I/II trial. The COMFORT-I trial also assessed symptoms using the PGIC instrument, where patients rated the improvement or worsening of their condition, which also appears to be an appropriate tool to use. The COMFORT-II trial did not assess symptom reduction, other than in terms of HRQoL.

HRQol

Both COMFORT trials assessed HRQoL as an exploratory endpoint. The COMFORT-II trial used the disease-specific European Organization for Research and Treatment Quality of Life Questionnaire (EORTC QLQ-C30) and the Functional Assessment of Cancer Therapy-Lymphoma (FACT-Lym) scale. The COMFORT-I trial used the EORTC QLQ-C30 and the PROMIS Fatigue scale. These tools appear to have been appropriate.

Survival

Neither of the COMFORT trials were designed to be sufficiently powered to detect a significant difference in survival outcomes. The COMFORT-II trial assessed overall survival, progression-free survival and leukaemia-free survival. The COMFORT-I trial assessed overall survival.

4.2.1.2 Phase I/II trial

This uncontrolled, dose finding study was included in the MS, mainly as supportive evidence and initially as the only source of survival data (at the clarification stage the manufacturer supplied updated longer term survival data from the RCTs).

Participants

Participants in the phase I/II trial differed from those in the RCTs in that there was no requirement for a palpable spleen measuring ≥5cm below the left costal margin, and the trial also included patients with intermediate-1 risk MF,[23] who have a better prognosis than those with intermediate-2 or high-risk MF. Thus it was more reflective of the broad product licence for ruxolitinib than the UK decision problem.

Intervention

The dose of ruxolitinib administered in the phase I/II trial differed from the doses administered in the RCTs, ranging from 10 mg twice daily to 200 mg once daily.[23] The different doses used in this trial, compared with the COMFORT trials, is likely to have resulted in differences in results for spleen volume, symptom reduction and adverse events (AEs).

Table 4.2 Study design and patient characteristics of the included phase I/II trial

Study details	**Phase I/II**
Location	United States
Design	Open-label, Uncontrolled single arm trial
Duration of core study	Not stated
Method of randomisation	None
Method of blinding	None
Intervention(s)	Oral ruxolitinib, dose escalation
Comparator(s)	None
Primary outcome	Number of participants with adverse events (measured monthly starting at baseline) Change from baseline in spleen and liver volume by MRI (measured at baseline, 1-, 3- and 6 months, and subsequent 6-month intervals) Change from baseline in body weight (measured at baseline, 1-, 2- and 3 months, and subsequent 3-month intervals) Analysis of clinical response based on a reduction of 50% or more in palpable splenomegaly (measured quarterly starting at baseline)
Secondary outcomes	Change in symptoms as assessed using the modified MFSAF 6-minute walk test Transfusion independence Change in mean white blood cell count, platelets and haemoglobin Change in *JAK2V617F* allele burden
Duration of follow-up for reported analysis	Median, 14.7 months for 1st analysis. Further analyses reported for a follow-up of 32 months and 42 months
Patient inclusion criteria	Age ≥ 18 Life expectancy ≥ 12 weeks Diagnosis of PMF, PPV-MF or PET-MF Patients could have: 1) newly-diagnosed MF requiring therapy and must be classified as intermediate- or high-risk according to Lille (Dupriez) Scoring System, or with symptomatic splenomegaly > 10 cm below the costal margin, or

Study details	**Phase I/II**
	2) previously-treated disease and must have experienced a relapse, have disease refractory to previous therapy or have had severe side effects from therapy ECOG performance status of 0, 1 or 2 Adequate bone marrow reserve (absolute neutrophil count > 1500/μL; platelet count > 100,000/μL) Adequate liver and renal function (total bilirubin ≤ 2.0 mg/dL; alanine aminotransferase ≤ 2.5 x institutional ULN; creatinine ≤ 2.5 mg/dL)
Patient exclusion criteria	Pregnant or breastfeeding Clinically significant viral infection (hepatitis, HIV) Current diagnosis of another malignancy New York Heart Association Criteria Class IV impairments Prior treatment with any JAK inhibitor or concurrent treatment with other prohibited medications Any current or planned therapy with CYP3A4 and CYP1A2 inhibitors or inducers Treatment with intermediate or high dose steroids greater than the equivalent of 10 mg prednisone per day Unresolved toxicity greater or equal to Grade 2 from previous anticancer therapy Incomplete recovery from any prior surgical procedures or who had surgery within 4 weeks prior to study entry Acute active infection requiring antibiotics Uncontrolled intercurrent illness
Characteristic	
Median age (range), years	65 (40–84)
Male, %	63
Disease type, % PMF PPV-MF PET-MF	 53.0 31.8 15.2
IPSS risk status, % High Intermediate-2 Not determined	 27.5 65.4 7.1
Median time since diagnosis (range), years	6.0 (0.1-36.0)
Previous MF therapy	86%
Palpable spleen length, median (range), cm	19 (2–36)
Platelet count, median (range), x 10^9/L	263 (101–1195)
Haemoglobin Median (range), g/dL < 10 g/dL, %	104 (72–169)
JAK2V617F mutation positive,%	82

4.2.2 Summary of the quality of the included trials

4.2.2.1 Randomised controlled trials

In general, both trials were well conducted; results of the quality assessment were presented in Table D6 of the MS. The placebo-controlled trial (COMFORT-I) was double-blinded, although patients were eligible for early unblinding if they had a 25% or greater increase in spleen volume from baseline. In addition, for early unblinding to occur before week 24, patients also had to demonstrate worsening early satiety accompanied by weight loss or worsening splenic pain accompanied by increased narcotic requirements. The trial was unblinded when all patients had completed the week 24 evaluation or discontinued treatment and 50% patients had completed the week 36 visit. The RCT comparing ruxolitinib with BAT (COMFORT-II) was not double-blind, although for the primary outcome of ≥35% reduction in spleen volume assessed by MRI or CT, the outcome assessors were blinded: images were read centrally by a reader unaware of the treatment group, so both trials can be considered to be at low risk of bias for the primary outcome.

Both trials clearly described the eligibility criteria, had adequate sample sizes, an appropriate method of randomisation and adequately reported the participants' baseline characteristics, which were broadly similar between the two trials, except that the COMFORT-I trial only included patients who had disease that was refractory to available therapies, had side effects requiring their discontinuation, or were not candidates for available therapies. The proportion of patients with baseline palpable spleen length less than 10 cm in the ruxolitinib group was greater than that in the comparator group in both trials (32.2% versus 23.3% in the COMFORT-II trial and 20.6% versus 17.5% in the COMFORT-I trial[9]), therefore, a smaller absolute reduction in spleen volume would be required to achieve a ≥35% reduction from baseline in the patients in the intervention group with smaller spleens than those in the comparator group. The proportion of patients with baseline palpable spleen length less than 10 cm was not reported in the MS.

In both COMFORT trials analyses were on an intention-to-treat basis for the primary endpoint; patients who discontinued therapy, crossed-over before 24 weeks (in the COMFORT-I trial) or did not have a 48-week assessment of spleen volume (in the COMFORT-II trial; due to discontinuation or entering the open-label extension phase of the trial) were counted as non-responders for change in spleen volume and symptom score, which seems appropriate. Pre-planned and post hoc subgroup analyses were performed for the primary outcome in both COMFORT trials, according to patient characteristics (for example gender, MF subtype, IPSS risk category and JAK2 mutation status), although the trials were not designed to be sufficiently powered to detect a statistically significant difference in spleen volume reduction for subgroups.

For some outcomes the number of patients included in the analyses was not reported in the MS. The ERG asked for clarification on the number of patients included in each of the analyses, which was provided by the manufacturer (Table 4.3). For some outcomes (actual change in spleen volume, actual change in symptoms, Global Health Status/QoL) there appears to be missing data, which has not been accounted for by the manufacturer. Therefore there is some question over the reliability of these results and how representative the responses of the patients who were included in the analyses are.

4.2.2.2 Phase I/II trial

This was an uncontrolled phase I/II trial. Study quality was adequate, as presented in Table D12 of the MS, although some aspects of external validity were rated as unclear, such as whether study participants were representative of the entire population from which they were recruited, and whether the staff and facilities where the patients were treated were representative of the treatment the majority of patients receive. It was also unclear whether there was adequate adjustment for confounding in the analyses from which the main findings were drawn. It should be noted that whilst the quality of this study was adequate, there is a high risk of bias associated with this study design relative to an RCT.

4.2.3 Summary of the results of the included trials

Table 4.3 Summary of results of the included RCTs

Outcome	COMFORT-II	COMFORT-I
Spleen volume		
Patients achieving ≥ 35% spleen volume reduction		
at week 12	30% vs 1% (n=144/146 ruxolitinib, n=72/73 BAT)	39% vs 0% (n=155/155 ruxolitinib, n=153/154 placebo)
at week 24	32% vs 0%, p < 0.001 (n=144/146 ruxolitinib, n=72/73 BAT)	42% vs 1%, p < 0.001[a] OR: 134.4, 95% CI 18.0 to 1004.9 (n=155/155 ruxolitinib, n=153/154 placebo)
at week 48	28% vs 0%, p < 0.001[a] (n=144/146 ruxolitinib, n=72/73 BAT)	–
Mean change in spleen volume		
at week 24	−29.2% vs +2.7%, p < 0.001 (n=125/146 ruxolitinib, n=45/73 BAT)	−31.6% vs +8.1% (n=139/155 ruxolitinib, n=106/154 placebo)
at week 48	−30.1% vs +7.3%, p < 0.001 (n= 98/146 ruxolitinib, n=34/73 BAT)	–
Symptoms		
Patients achieving ≥ 50% reduction in TSS at week 24	–	45.9% vs 5.3%, p < 0.001 OR: 15.3, 95% CI 6.9 to 33.7 (n=149/155 ruxolitinib, n=152/154 placebo)

Mean change from baseline in TSS at week 24	–	46.1% vs −41.8%, p < 0.001 (n=129/155 ruxolitinib, n=103/154 placebo) Mean absolute change in symptom score: -8.6 vs 3.2
PGIC: patients rating condition much/very much improved at week 24, %	–	66.9% vs 11.2% (n=139/155 ruxolitinib, n=107/154 placebo)
HRQoL		
Mean change from baseline in Global Health Status/QoL (EORTC QLQ-C30)	At week 48: +9.1 vs +3.4 (n= 66/146 ruxolitinib, n=27/73 BAT)	At week 24: +12.3 vs −3.4, p < 0.001 (n=136/155 ruxolitinib, n=104/154 placebo)
Mean change from baseline in FACT-Lym total score at week 48	At week 48: + 11.3 vs −0.9 (n= 70/146 ruxolitinib, n=29/73 BAT)	–
Survival		
Overall survival	At median follow-up of 61 weeks: 92.0% vs 95.0%, (HR, 1.01, 95% CI 0.32 to 3.24) At median follow-up of 112 weeks: 86% vs 78%, (HR, 0.52; 95% CI 0.27 to 1.00)	At median follow-up of 51 weeks: 91.6% vs 84.4%, (HR, 0.50; 95% CI 0.25 to 0.98; p = 0.04) At median follow-up of 102 weeks: 27 ruxolitinib patients died vs 41 placebo patients, (HR, 0.58; 95% CI 0.36 to 0.95; p=0.028)
Progression-free survival	At week 48: 69.9% vs 74.0%, (HR, 0.81, 95% CI 0.47 to 1.39)	–
Adverse events		
Anaemia	40.4% vs 12.3%[b]	96.1% vs 86.8%[c]
Grade 3/4 anaemia	42% (40.4%[e]) vs 31% (23.3%[e])	45.2% (52.2%[d]) vs 19.2%
Thrombocytopenia	44.5% vs 9.6%[b]	69.7% vs 30.5%[c]
Grade 3/4 thrombocytopenia	8% (9.6%[e]) vs 7% (9.6%[e])	12.9% (16.2%[d]) vs 1.3%
Non-haematological adverse events affecting >20% of either treatment group:		
Diarrhoea	23% vs 12%	23.2% vs 21.2%[c]
Peripheral oedema	22% vs 26%	18.7% vs 22.5%[c]
Fatigue	-	25.2% vs 33.8%[c]
Abdominal pain	-	10.3% vs 41.1%[c]
Bruising	-	23.2% vs 14.6%[c]
Blood transfusion	51% vs 38%	-

[a]Primary endpoint; [b]Data from Harrison et al., 2011[24]; [c]Data from Verstovsek et al., 2012[21]; [d]Updated data from Verstovsek et al., 2012 ASH abstract[25]; [e]Updated data from Cervantes et al., 2012 ASH abstract[26]; CI, confidence interval; COMFORT, controlled myelofibrosis study with oral JAK inhibitor treatment; EORTC QLQ-C30, European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire; FACT-Lym, Functional Assessment of Cancer Therapy-Lymphoma; HR, hazard ratio; HRQoL, health-related quality of life; PGIC, Patient's Global Impression of Change; TSS, Total Symptom Score.

Spleen reduction

As shown in Table 4.3, a statistically significantly greater proportion of patients in the ruxolitinib groups had a ≥35% reduction in spleen volume, compared with placebo or BAT.

For both of the COMFORT trials this primary variable result reported in the MS was based on an unadjusted analysis reported in the MS. However the clinical study report (CSR) reported a logistic regression model with baseline spleen volume, sex (male or female), disease subtype (PMF, PPV-MF, and PET-MF), hydroxyurea usage in the period before entering the study and study treatment included as model effects. For both trials the adjusted results were reported to be similar to the unadjusted results. In COMFORT-I they were 40% versus 0.65% for ruxolitinib versus placebo, with an odds ratio (95% confidence interval) of 125.4 (16.8, 398.1) and $p < 0.0001$. The actual results for COMFORT-II were not provided to the ERG. These results are as expected in randomised trials which should have balanced for such effects.

In addition, the CSRs reported for both trials a linear model with baseline, treatment, and baseline*treatment interaction as the model effects was explored to see if there was a significant baseline*treatment interaction in the percent change in spleen volume. In the COMFORT-I trial the baseline*treatment interaction was significant at the pre-specified 0.20 significance level ($p = 0.0712$). The results indicated that the treatment effect significantly depended on the baseline spleen volume. To further explore the nature of the significant interaction, a simple regression analysis was produced by treatment. The treatment difference remained in the same direction across the baseline spleen volume range; however, the effect size, in terms of percent change from baseline to week 24 in spleen volume, was larger for smaller baseline spleen volumes and smaller for larger spleen volumes. Whether this finding was the same in the COMFORT-II trial was not reported.

Amongst patients who achieved a ≥35% spleen volume reduction, this was maintained for a year or more in the majority of patients. The median time to first observation of a reduction of ≥35% from baseline was 12.3 weeks in the COMFORT-II trial, and in the COMFORT-I trial most of the patients who achieved a ≥35% spleen volume reduction had achieved this by week 12 (the median time to first observation was not reported for the COMFORT-I trial).

Changes in palpable spleen length over time were also reported in the MS: in the COMFORT-II trial there was a 56% decrease from baseline in the ruxolitinib group at week 48, compared with a 4% increase from baseline in the BAT group, although this analysis only included around half of the patients in either treatment group and the reason for this was not reported. In the COMFORT-I trial there were also reductions in spleen length in the ruxolitinib group and increases in spleen length in

the placebo group, however the percentage decrease/increase from baseline was not reported. The mean percentage change from baseline in spleen length was shown as Figure 11 of the MS for the COMFORT-II trial and Figure 33 for the COMFORT-I trial; however, this only appeared to reach a 50% reduction in palpable spleen length from the 24 week assessment in the COMFORT-II trial and from the week 48 assessment in the COMFORT-I trial.

A 50% reduction in palpable spleen length (which the manufacturer claims corresponds to a 35% reduction in spleen volume) is one of the criteria for demonstrating clinical improvement in the IWG-MRT consensus criteria for treatment response in myelofibrosis, however, this is for patients with a spleen at least 10 cm at baseline; a spleen that is more than 5 cm at baseline should become non-palpable for a clinical improvement to have been achieved. Therefore, some patients with a baseline palpable spleen length of less than 10 cm may not have met the IWG-MRT criteria for clinical improvement.

Spleen volume subgroup analyses

The RCTs were not designed to be sufficiently powered to detect a statistically significant difference in spleen volume reduction for subgroups (according to gender, MF subtype, IPSS risk category or JAK2 mutation status, amongst others). However, for the primary outcome (≥35% reduction in spleen volume) two subgroup differences did reach statistical significance in the COMFORT-I trial: females versus males; and patients on a 20mg twice daily versus 15 mg twice daily starting dose. None reached statistical significance in the COMFORT-II trial. These data were not presented in the MS, but figures showing results of subgroup analyses were presented in the manufacturer's response to the ERG's clarification letter. The figures are shown below as Figures 4.2 and 4.3.

Figure 4.2 Subgroup analysis of spleen volume for COMFORT-II trial

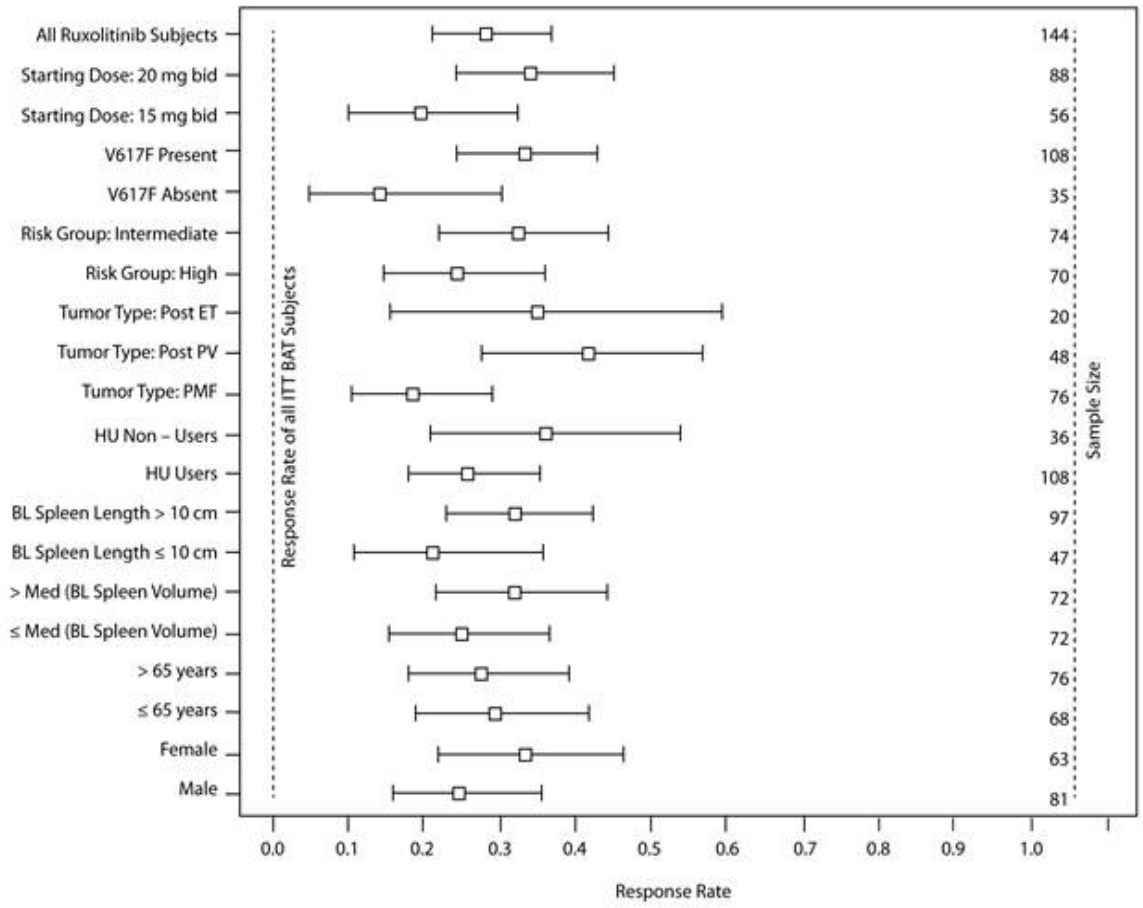
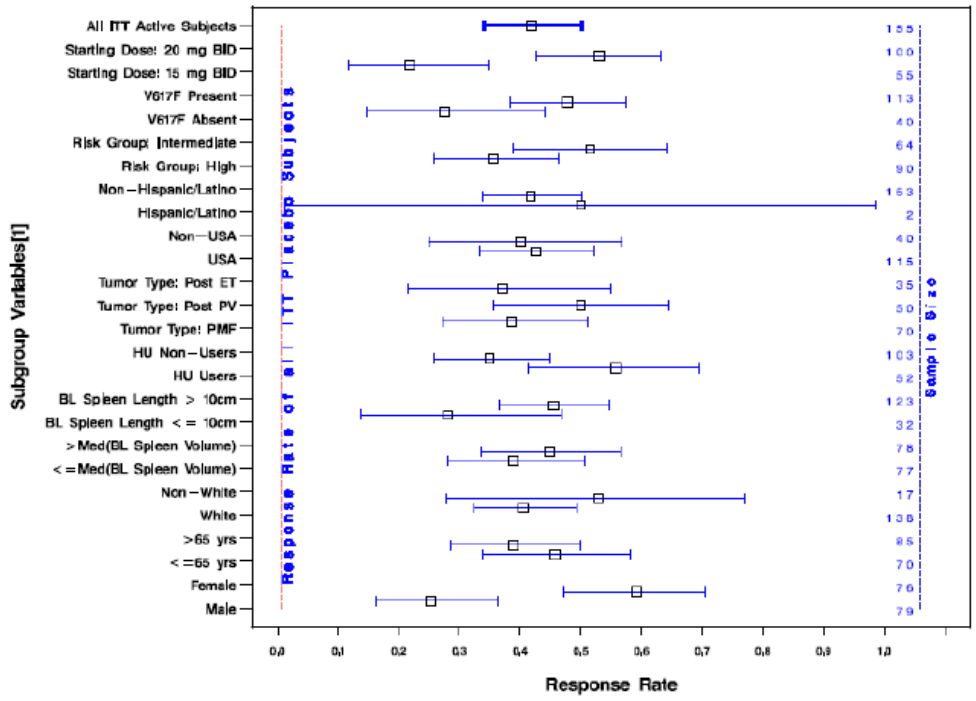


Figure 4.3 Subgroup analysis of spleen volume for COMFORT-I trial

Figure 23: Proportions of Subjects in Each Subgroup Who Achieved a ≥ 35% Reduction From Baseline in Spleen Volume at Week 24 (Ruxolitinib-Treated ITT Subjects)



Note: The vertical line to the left indicates the response rate of all ITT placebo subjects.
Note: The box represents the point estimate and the error bars represent the 95% confidence intervals based upon binomial exact method.
BID = twice daily; BL = Baseline; HU = hydroxyurea use; Med = median
Source: Figure 14.4.2.13

The suggestion of possible differential effects by MF subtype may warrant exploration in adequately powered clinical trials.

Symptom reduction

Only the COMFORT-I trial assessed symptom reduction. Significantly more patients in the ruxolitinib group achieved a ≥50% reduction in TSS, assessed using the MFSAF version 2, at week 24 than in the placebo group (45.9% vs 5.3%); the tool used to measure this outcome was appropriate and the analysis included over 95% of randomised patients, therefore this result is likely to be reliable. The ERG requested further information from the manufacturer to support the clinical validity of the 50% reduction cut off.

The manufacturer's response stated that the initial choice of a 50% reduction in TSS as the definition of a response was based on the assumption that this would represent a robust hurdle that would be likely to represent clinical benefit to patients, and that would be unlikely to be achieved by an inactive treatment. The COMFORT-I trial was designed to allow for validation of this assumption. An anchor-based method was employed in order to conclude that a 50% reduction in the TSS from

baseline to week 24 reflected a genuine patient benefit. The results for this exercise found that 91.2% of treatment group subjects labelled as responders (i.e. showed a greater than 50% reduction in their TSS from baseline to week 24) also characterised their own condition as either “much” or “very much” improved (according to assessment using the PGIC instrument). Additionally, 73.6% of the placebo group labelled as non-responders characterised their condition as unchanged or worsening. Whilst the results presented (see manufacturer’s response document question A5) suggest there is a reasonable correlation between a 50% or greater reduction in the TSS and meaningful benefit reported by patients (in terms of PGIC scores) there is uncertainty around this, with both responders failing to record improvements and non-responders reporting improvements.

In terms of actual scores the mean TSS reduced to almost half the baseline level in the ruxolitinib group at 24 weeks compared with an increase in TSS in the placebo group at 24 weeks, therefore, this result is also likely to be clinically significant (see Table 4.4 below, which is Table 19 of the manufacturer’s response document).

Table 4.4 Baseline, change and percent change from baseline in week 24 Total Symptom Score (observed cases)

Table 19: Baseline, Change, and Percent Change From Baseline in Week 24 Total Symptom Score (Observed Cases)

	Ruxolitinib (N = 155)	Placebo (N = 154)
Baseline total symptom score		
N	131	105
Mean (SD)	18.0 (10.9)	16.5 (11.5)
Median	17.8	14.8
Min, Max	0.0, 50.1	0.0, 52.7
Week 24 total symptom score		
Mean (SD)	9.4 (9.7)	19.7 (13.7)
Median	6.6	17.6
Min, Max	0.0, 49.1	0.0, 59.7
Change from Baseline to Week 24[a]		
N	131	105
Mean (SD)	-8.6 (10.0)	3.2 (9.4)
Median	-6.9	2.0
Min, Max	-35.8, 22.1	-27.1, 30.3
p-value[b]	<0.0001	
Percent Change from Baseline to Week 24[c]		
N	129	103
Mean (SD)	-46.1 (48.6)	41.8 (99.3)
Median	-56.2	14.6
Min, Max	-100.0, 108.3	-100.0, 511.6
p-value[b]	<0.0001	

[a] This analysis only includes subjects who had a non-missing change from Baseline to Week 24. Data collected after the date of treatment crossover were not included in this analysis.
[b] The p-value was calculated using the Wilcoxon Rank-Sum test.
[c] This analysis only includes subjects who had a non-missing percent change from Baseline to Week 24. Data collected after the date of treatment crossover were not included in this analysis.
Source: Table 14.2.5.1 and 14.2.6.1

The proportion of patients rating their condition as much improved or very much improved (assessed using the PGIC instrument) at week 24 also supports this finding. However, data were missing for many of the placebo group patients in these analyses, at baseline and week 24 which undermines the reliability of the results.

Symptom reduction has not been compared between patients taking ruxolitinib and patients taking best available therapy for MF.

Given the importance of symptom reduction for patients with MF the ERG requested further information on the correlation between spleen reduction and symptom improvement. The manufacturer provided the results of an analysis of data from COMFORT-I (Table 4.5), which indicated that as spleen volume reduces, patients concomitantly report improvements in their MF symptoms.

Table 4.5 Modified MFSAF v2.0 diary change scores by spleen volume reduction (COMFORT-I trial)

Spleen Volume	**MFSAF Total Symptom Score Change**[1]		
	N	Mean	SD
No Reduction (<10%)	117	-1.97	9.17
10% to < 35% Reduction	63	6.64	10.53
35% to < 50% Reduction	38	12.54	10.24
> 50% Reduction	25	10.50	10.24
Total	243[2]	3.82	11.6
	F=28.043, p<0.001		

[1] Positive (+) mean change scores indicate improvement; negative (-) scores indicate worsening. This is different than what is reported in the Clinical Study Report results which bases change on (Week 24 – Baseline)

[2] 243 (of the 250 subjects with paired Baseline Total score and Week 24 Total score) have a valid spleen volume change score.

The manufacturer also stated that the relationship between spleen length and percentage improvement on TSS is also being analysed in a phase II exploratory analysis that is currently underway in the UK (NCT01558739). The primary endpoint of this study is the assessment of ruxolitinib using a composite measure that is composed of >50% reduction in splenomegaly at 48 weeks and >50% reduction in TSS derived from MFSAF at 48 weeks. This should give a clearer indication of clinical improvement by identifying patients who will derive some clinical benefit, whether spleen size reduction or symptom control. It is anticipated that this UK study will generate local health resource utilisation data.

HRQoL

In the COMFORT-II trial greater improvements in Global Health Status/QoL were observed in the ruxolitinib group than the BAT group at 48 weeks, although there were missing data for many patients in the analysis, with only 66/146 ruxolitinib patients and 27/73 BAT patients included (see Table 4.3) reducing the reliability of the results. Mean Global Health Status/QoL also improved in the BAT group at 48 weeks, but the improvements were not as great as those seen in the ruxolitinib group; it is unclear whether the difference in the level of improvement is clinically significant.

In the COMFORT-I trial Global Health Status/QoL was statistically significantly better with ruxolitinib than placebo at week 24, with ruxolitinib patients' Global Health Status/QoL improving

from baseline and placebo patients' worsening, although this analysis only included 104/154 patients in the placebo group.

In the COMFORT-II trial improvements in individual symptom scores on the EORTC QLQ-C30 for MF-associated symptoms were seen for ruxolitinib patients at 48 weeks, whereas BAT patients had worsening scores, in terms of fatigue, pain, dyspnoea, insomnia and appetite loss; again, there were missing data for many patients in these analyses, reducing the reliability of the results (see Appendix 1). Diarrhoea was the only symptom where BAT patients had an improvement in scores, this improvement was greater than the improvement seen for ruxolitinib patients (EORTC QLQ-C30 symptom score: -6.2 versus -4.1; this result was not reported in the MS, but was reported in the publication of the COMFORT-II trial.[22]).

Individual FACT-Lym scores were also improved in the ruxolitinib group in the COMFORT-II trial, whilst BAT patients had worsening scores at 48 weeks, although there were missing data for many patients in the analysis, with only 70/146 ruxolitinib patients and 29/73 BAT patients included, again reducing the reliability of the results. The differences in scores at week 48 were clinically significant.

Progression-free survival

Progression-free survival was an outcome in COMFORT-II only. At 48 weeks there was slightly higher in the BAT treated group 74% compared with 70% in the ruxolitinib group, but the difference was not statistically significant (HR 0.81, 95% CI 0.47 to 1.39). Thus there was no evidence of an improvement in progression-free survival with ruxolitinib.

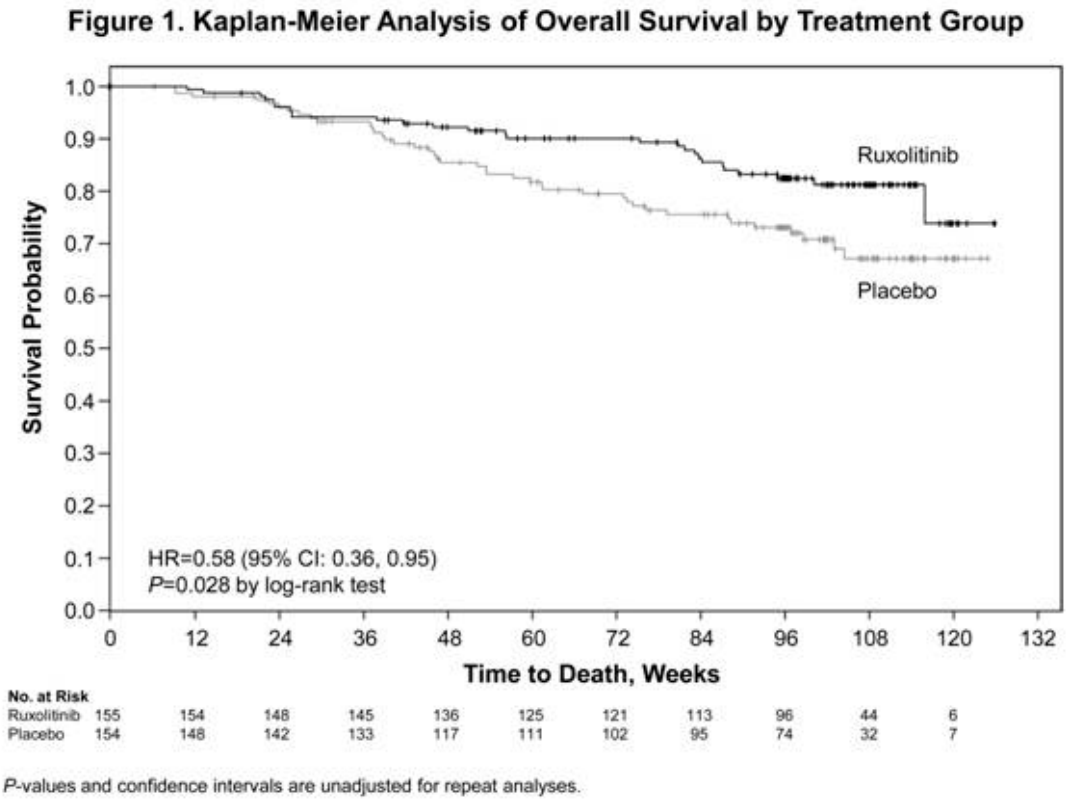
Survival

Overall survival data were presented for both of the RCTs and the phase I/II trial. The ERG considers the COMFORT-II trial data to be the most relevant and reliable in terms of survival; this is because it includes a relevant comparator group, with fewer control group patients having crossed over to ruxolitinib or discontinued from the study than in the COMFORT-I trial.

The RCTs were not designed to be sufficiently powered to detect a statistically significant difference in survival outcomes between treatment groups. Although at shorter follow-up no survival benefit was seen with ruxolitinib compared with the BAT group (Table 4.3), it reached borderline statistical significance; 86% versus 78% (HR 0.52, 95% CI 0.27 to 1.00) at a median of 112 weeks of follow-up (updated survival data submitted to the ERG as part of the manufacturer's response to the clarification letter). A similar improvement in survival was found when ruxolitinib was compared with placebo (Table 4.3). The survival analyses were undertaken on an intention-to-treat basis, with patients

analysed in the treatment group to which they had been randomised. As many patients had crossed over from placebo/BAT to ruxolitinib or discontinued from the study at the time of the survival analyses; the manufacturer claims that this means the survival benefit of ruxolitinib is likely to be underestimated in these trials, which could be the case. Further difficulties for interpretation are due to the small number of patients at risk at the later time points, due in part to mortality events and patient withdrawals but also due to the fact that many patients still in the study have not yet reached the later time points analysed. Thus for example, from the latest update from COMFORT-I, at 120 weeks only 44 of the 155 patients randomised to ruxolitinib are at risk and only 32/154 placebo patients. At 120 weeks the numbers are 6 and 7 respectively. The equivalent 'at risk' numbers for the 112 week time point in COMFORT-II, when the survival benefit of ruxolitinib over BAT reaches borderline statistical significance, were not reported with these results (Manufacturer's response document). The updated Kaplan-Meier analysis of overall survival by treatment group for the COMFORT-I trial is presented in Figure 4.4 below.

Figure 4.4 Updated Kaplan-Meier analysis of overall survival by treatment group **(COMFORT-I)**



The MS presented survival data from the uncontrolled Phase I/II trial, which had follow up of 32 months. The phase I/II trial survival data cannot be considered directly relevant as the doses of ruxolitinib used in this trial do not reflect the licensed dose, long-term survival data are presented separately for the two centres in the trial, meaning very small numbers of patients were included in the analyses. Data were compared for responders versus non-responders, and for trial patients versus a historical control group which was not described in the MS.

Updated survival results from the phase I/II trial were reported separately for the MD Anderson Cancer Centre (MDACC) and the Mayo Clinic. Overall survival for the subset of 107 patients treated at the MDACC was 69% after a median follow-up of 32 months. The Patients who had a ≥50% reduction in palpable spleen length had significantly longer survival compared with patients with a reduction in palpable spleen length of 25% or less (HR 0.22; 95% CI 0.10 to 0.51, p=0.0001), although very small numbers of patients were included in this analysis (20 patients with <25% reduction in spleen length from baseline, reducing to 10 patients at 16 months and 2 patients at 36 months). These were the survival data used in the manufacturer's decision model.

The overall survival of patients followed-up at the MDACC was compared with the overall survival of a historical control group of patients identified from three large databases and matched on the basis of study inclusion criteria. After a median follow-up of 32 months, 30.8% ruxolitinib patients and 60.9% control patients had died (HR 0.58, 95% CI 0.39 to 0.85, p=0.005). In contrast, there was no significant difference in overall survival for the subset of 51 patients who were treated at the Mayo Clinic compared with a historical control group at a median follow-up of 42 months.

Owing to the small numbers of patients in the survival analyses, the fact that overall survival was not a study end point and the design of the trial, these survival results are subject to great uncertainty.

Absolute mortality rates

The mortality rate of the MDACC seems rather high compared with that in the COMFORT-I and COMFORT-II trials; 14% ruxolitinib patients and 22% BAT patients had died at a median follow-up of 112 weeks in the COMFORT-II trial, 17% ruxolitinib patients and 27% placebo patients had died at a median follow-up of 102 weeks in the COMFORT-I trial, whereas 31% ruxolitinib patients and 61% historical control patients had died after a median follow-up of 32 months in the phase I/II trial. The MS states that the phase I/II trial included more patients with IPSS high-risk disease than the COMFORT trials (MS page 100), which may explain the higher mortality rate in this trial, however the manufacturer's response document also states that this trial included 10% intermediate-1 risk patients. However, the MS does not present any baseline data on the historical control patients, other

than stating that they were matched on the basis of study inclusion criteria; therefore, the reason for their high mortality rate is not clear.

Adverse events

The MS stated that ruxolitinib was generally well tolerated, with the most frequently occurring grade 3 or 4 AEs (anaemia and thrombocytopenia) being generally managed by dose modifications and/or blood transfusions. The MS states that anaemia and thrombocytopenia were expected given the mechanism of action of ruxolitinib. The ERG confirms that haematological AEs were very common with ruxolitinib, particularly thrombocytopenia and anaemia.

In the COMFORT-II trial 51% patients taking ruxolitinib required at least one blood transfusion, compared with 38% patients taking BAT, although the mean number of transfusions per month was similar in the two treatment groups. During the COMFORT-I study 41.2% of patients in the ruxolitinib group who were transfusion-dependent at baseline became transfusion independent (according to the IWG-MRT criteria), but this was lower than in the placebo group: 46.9%. These results are likely to have been achieved whilst using dose reductions and interruptions to minimise adverse effects.

It is important to note that anaemia is a common symptom associated with MF and, whilst some treatments specifically target haematological symptoms of MF, ruxolitinib exacerbates these symptoms in some patients, at least in the short term. The ERG requested clinical effectiveness data for subgroups of patients with and without anaemia or thrombocytopenia at baseline. The manufacturer provided data from the COMFORT-I trial that indicated that ruxolitinib-treated patients with new-onset grade 3 or 4 anaemia experienced improvements in spleen volume (see Figure 4.5 below) and symptoms (TSS; see Figure 4.6 below) that were similar to ruxolitinib-treated patients without anaemia. No data were presented relating to thrombocytopenia, or from the COMFORT-II trial.

Figure 4.5 Symptom score results by presence of new-onset grade 3 or 4 anaemia (COMFORT-I)

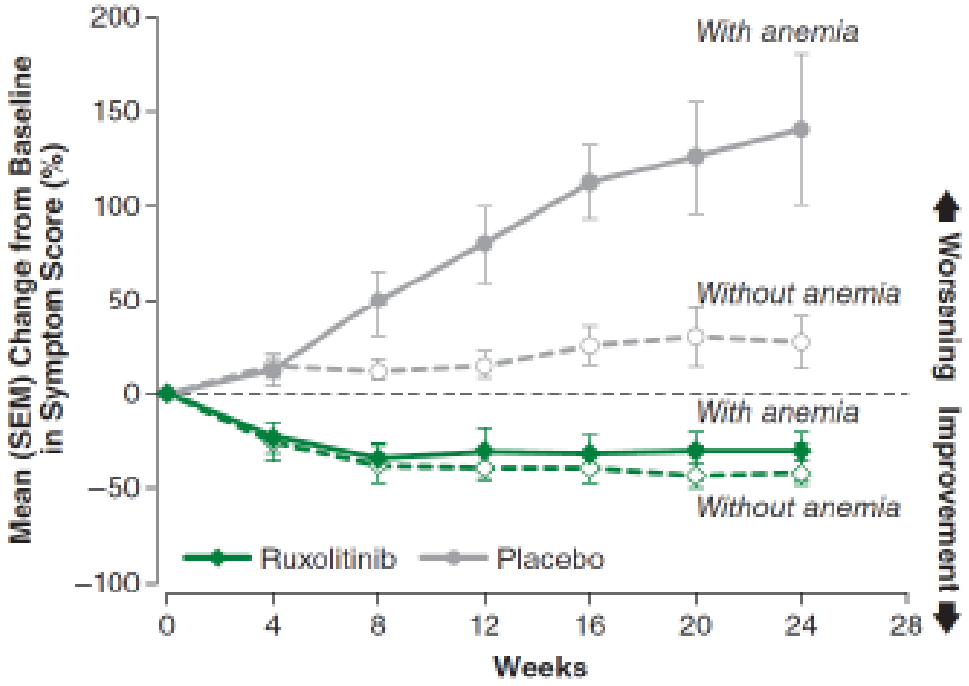
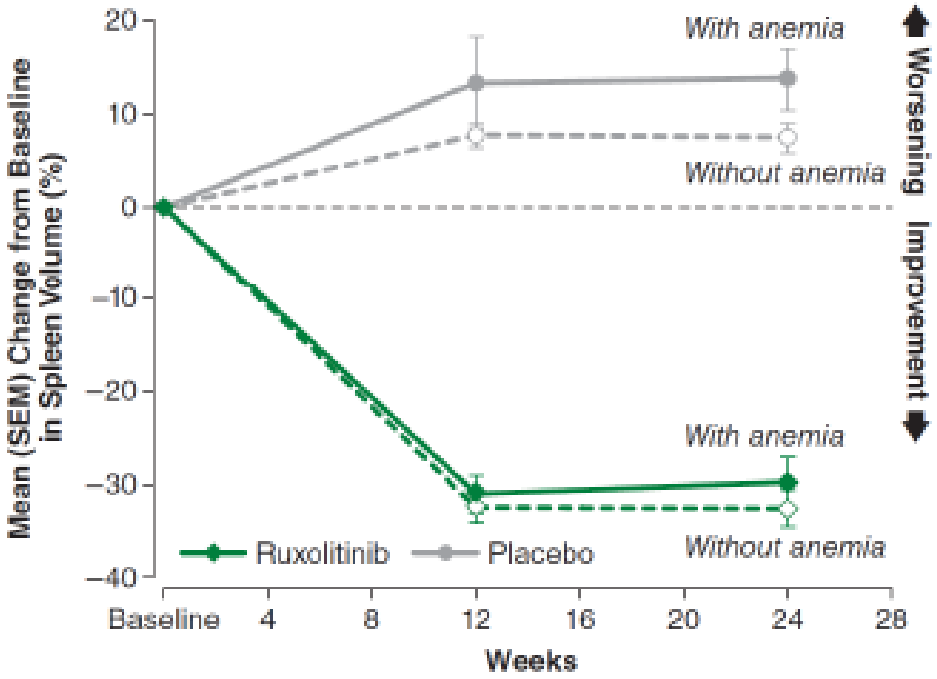


Figure 4.6 Spleen volume results by presence of new-onset grade 3 or 4 anaemia **(COMFORT-I)**



Grade 3 or 4 AEs were reported in 42% of the ruxolitinib group and 25% of the BAT group. 63% ruxolitinib patients in the COMFORT-II trial required dose interruptions or reductions due to AEs, most commonly thrombocytopenia. In the COMFORT-I trial 32% ruxolitinib patients and 36% placebo patients required dose interruptions; symptoms (TSS) returned to baseline levels over the period of approximately one week in these patients.

The safety and effectiveness of ruxolitinib has not been assessed in patients with a platelet count <100 x 10^9/L or absolute neutrophil count <1 x 10^9/L, therefore, given the haematological adverse effects associated with ruxolitinib, these patients may not be suitable for ruxolitinib treatment. The Summary of Product Characteristics states that there is limited information to recommend a starting dose for patients with platelet counts between 50,000/mm^3 and <100,000/mm^3. The maximum recommended starting dose in these patients is 5 mg twice daily and the patients should be titrated cautiously. Treatment should be discontinued for platelet counts less than 50,000/mm^3 or absolute neutrophil counts less than 500/mm^3.[27]

The most common non-haematological AE occurring in the ruxolitinib group was diarrhoea, which occurred in 23% of the ruxolitinib group and 12% of the BAT group.

Discontinuation rates

At the time of data cut-off for the COMFORT-II trial, 18% ruxolitinib patients had discontinued treatment (8% due to AEs), 20% ruxolitinib patients continued in the extension phase of the trial after a disease progression event qualified them for entrance into the extension phase, leaving 62% ruxolitinib patients ongoing in the randomised treatment phase. In the BAT group 33% patients discontinued treatment (12% due to withdrawal of consent, 5% due to AEs) and 25% crossed over to ruxolitinib, leaving 42% BAT patients ongoing in the randomised treatment phase. At a median follow-up of 112 weeks the primary reasons for discontinuation were AEs (11.6% ruxolitinib patients and 6.8% BAT patients), consent withdrawal (4.1% ruxolitinib patients and 12.3% BAT patients) and disease progression (2.7% ruxolitinib patients and 5.5% BAT patients). Overall, 72.6% (106/146) patients in the ruxolitinib arm and 61.6% (45/73) patients in the BAT arm entered the extension phase to receive ruxolitinib, and 55.5% (81/146) of those originally randomised to ruxolitinib remained on treatment at 112 weeks of follow-up. The primary reasons for discontinuation from the extension phase were progressive disease (8.2%), AEs (2.1%) and other (4.1%).[26]

During the course of the COMFORT-I trial, 14% ruxolitinib patients discontinued treatment (5% due to AEs) and 86% continued on randomised treatment. In the placebo group 24% patients discontinued treatment (8% due to disease progression, 5% due to AEs) and 23% crossed over to ruxolitinib, leaving 51% placebo patients continuing on randomised treatment. At 24 weeks 11% ruxolitinib patients and 10.6% placebo patients had withdrawn from the COMFORT-I trial due to AEs (Table B12 of MS). At a median follow-up of 102 weeks, 100/146 patients originally randomised to ruxolitinib remained on treatment.[25]

Figures 8 and 9 of the MS present details of patient disposition during the course of the trials. However, the information provided from the RCTs and follow-up studies do not present clear estimates of the long-term disposition of patients on ruxolitinib.

Discontinuation rates were high in the phase I/II trial; 46% at the MDACC at a median follow-up of 32 months and 92% at the Mayo Clinic at a median follow-up of 3.5 years; reasons for discontinuation included progressive disease, patient withdrawal of consent and physician decision to discontinue.

4.3 Conclusions of the clinical effectiveness section

The MS evaluation of ruxolitinib was primarily based on two RCTs; one comparing ruxolitinib with BAT and one comparing ruxolitinib with placebo. In addition, some results from a phase I/II dose finding study were reported. Whilst the methods of the systematic review were not well reported in the MS, it is unlikely that relevant studies of ruxolitinib were overlooked. In addition, there is no evidence base of RCTs to use as the basis of any form of indirect comparison of individual therapies with ruxolitinib.

The objective of the NICE scope was that ruxolitinib be appraised within its licensed indication for the treatment of MF. The EMA granted marketing authorisation of ruxolitinib for the treatment of disease-related splenomegaly or symptoms in adult patients with PMF, PPV-MF or PET-MF. However, the RCTs of ruxolitinib were conducted in patients with splenomegaly and intermediate-2 or high-risk MF, therefore its clinical effectiveness has only been assessed in this narrower, higher risk population. In addition, patients with an absolute neutrophil count $\leq 1 \times 10^9$/L or platelet count $<100 \times 10^9$/L and patients suitable for allo-HSCT at the time of study enrolment were excluded from the trials, so the efficacy and safety of ruxolitinib has not been assessed in these patients.

Evidence from the two good quality RCTs demonstrates that ruxolitinib is more effective than BAT and placebo at reducing spleen volume in patients with splenomegaly, with up to 42% patients taking ruxolitinib achieving a ≥35% reduction in spleen volume, which is maintained for over a year in the majority of responders. Whilst the manufacturer claims that this equates to the spleen reduction criterion for clinical improvement according to the IWG-MRT consensus criteria for treatment response in myelofibrosis (≥50% reduction in palpable spleen length for patients with a palpable spleen that is at least 10 cm at baseline), the ERG believes it may generate an optimistic response rate.

Importantly, ruxolitinib does not have a favourable effect on haematological symptoms such as anaemia and thrombocytopenia which are also criteria for demonstrating clinical improvement, defined by the IWG-MRT consensus criteria for treatment response in myelofibrosis; these are in fact worsened by treatment in some patients, at least in the short term, and were assessed only in terms of their being AEs.

It should also be noted that, treatment response was not assessed against complete remission or partial remission criteria defined by the IWG-MRT.

Ruxolitinib was also associated with improvements in symptom scores and quality of life; however, there were missing data for these quality of life and symptom improvement results. There was no justification for the missing data, therefore, the reliability and generalisability of these results is unclear.

There was no evidence of an improvement in progression-free survival with ruxolitinib compared with BAT (COMFORT-II trial only).

Whilst the RCTs were not designed to be sufficiently powered to detect statistically significant differences in overall survival, overall survival was found to be statistically significantly better for ruxolitinib patients than patients taking placebo at a median follow-up of 51 weeks. Overall survival appears to be similarly improved with ruxolitinib compared with BAT, although the difference did not reach borderline statistical significance until a median of 112 weeks of follow-up, at which time point the number of patients in the analysis was very small and so this result is subject to uncertainty.

Comparison of survival in patients who had a ≥50% reduction in palpable spleen length with those who achieved a reduction in palpable spleen length of 25% or less generated a significant difference (HR 0.22; 95% CI 0.10 to 0.51, p=0.0001). However, as this analysis was based on only a subgroup of patients from an uncontrolled study and included only very small numbers of patients, this result is subject to great uncertainty.

Haematological AEs were very common with ruxolitinib; 63% of patients required dose interruptions or reductions due to AEs, most commonly thrombocytopenia and anaemia. Dose interruptions or discontinuation resulted in worsening of symptoms of MF reverting to baseline levels over the period of around one week. 8% and 11% ruxolitinib patients discontinued treatment due to AEs from the COMFORT-II trial and COMFORT-I trial, respectively.

Higher proportions of patients discontinued treatment in the control arm than the ruxolitinib arm for both COMFORT trials (33% BAT patients discontinued treatment and 25% crossed over to ruxolitinib whilst 18% ruxolitinib patients discontinued treatment at the time of data cut-off for the COMFORT-II trial; 24% placebo patients discontinued treatment and 23% crossed over to ruxolitinib whilst 14% ruxolitinib patients discontinued treatment during the course of the COMFORT-I trial). In the long-term, discontinuation rates were high as indicated by data from the phase I/II trial; 46% at the MDACC at a median follow-up of 32 months and 92% at the Mayo Clinic at a median follow-up of 3.5 years.

In conclusion, for patients who can tolerate ruxolitinib and remain on therapy, the evidence suggests that splenomegaly and its associated symptoms can be reduced. However, patients with MF without symptoms related to splenomegaly are less likely to benefit from ruxolitinib. There is uncertainty over the very long-term effectiveness and tolerability of ruxolitinib, with limited data beyond two years.

5 COST EFFECTIVENESS

The focus of this section is the economic evidence initially submitted by the manufacturer in their report and the additional information provided by the manufacturer in their responses to requests for clarification from the ERG. A critical review is undertaken on the basis of this evidence and through the direct examination of the electronic version of the economic model. The critical appraisal is undertaken with the aid of a checklist, which will aid an assessment regarding the quality of the economic evaluation and guide a narrative review, which will highlight the key assumptions and potential limitations. Where possible, the issues highlighted are further explored in additional analyses undertaken by both the manufacturer during the clarification stage and by the ERG. The ERG's further analyses are presented in Section 6.

The manufacturer's initial economic submission included:

a) A description of the databases and websites searched in the literature review of cost-effectiveness studies, resource use studies and quality-of-life studies (MS, pg.262, Section 10.10), along with details of the systematic search strategy used to identify these studies.

b) A report on the de novo economic evaluation conducted by the manufacturer. The report outlined the intervention; comparators and patient population; the modelling methodology; the resource components and unit costs; data input sources and assumptions; the base case results; and sensitivity analysis (MS, pg. 125 to 207).

c) The manufacturer's electronic Excel-based de novo model.

A short outline of the economic elements of the submission is presented prior to a more detailed summary and critique. A summary of the manufacturer's approach and signposts to the relevant sections in the MS are reported in Table 5.1.

The manufacturer undertook a systematic review of cost-effectiveness studies including ruxolitinib. Based on their findings they developed a de novo economic decision model. The model presented is a state-transition Markov model, comprising four mutually exclusive health states, which reflect the treatment of MF (responder, non-responder, discontinuation and death). The time horizon for the base case was 35 years. The model uses a spleen volume reduction of 35% as the response criterion. Response defines quality of life, costs, and survival, so is the key factor in the model. The model allows for patients to be treated for a period of 24 weeks prior to response status being defined; patients are essentially incurring cost and accruing benefits during the 24 week period but their response status is not fixed until 24 weeks. The only possible transition during this initial treatment phase is to the discontinuation state. This discontinuation state is only utilised during the initial 24

week treatment phase. Patients who make this transition remain in the discontinuation state until the initial 24 week period is over, at which point they then move to the non-responder state. For all patients who remain on treatment a stopping rule is employed at 24 weeks. At 24 weeks all patients who have not achieved a 35% reduction in spleen volume discontinue ruxolitinib and transition directly to the non-responder state, where they remain for the duration of the model. The discontinuation state is not used outside of the initial treatment phase. Patients who have achieved a response at 24 weeks move to the responder state, where they remain until discontinuation or death; those responders who discontinue treatment move directly to the non-responder state. Death is a terminal state, which absorbs patients dying for any reason. A 25% spleen volume reduction with a 12 week initial treatment period was explored in sensitivity analysis.

The model population evaluated are defined by the IPSS criteria as intermediate-2 and high-risk MF patients. The ERG has comments regarding the appropriateness of the defined population which are outlined in Section 4.2.1 of the report. The key clinical model inputs including: response rates, survival, discontinuation during the initial treatment phase, and AEs will be discussed along with their data sources in Section 5.2.5, Model inputs. Costs included drug acquisition, management of MF (GP visits, inpatient stay, and outpatient attendance), blood transfusions, palliative care costs and those costs associated with the included complications. Details of the resource use and costing assumptions will be fully discussed in Section 5.2.7, Resources and costs.

Model outcomes were measured in quality-adjusted life-years (QALYs) based on utility weights identified from the literature. The MS states that no EQ-5D data were available in a MF population. Further, they state that attempts to map from HRQoL data collected as part of the relevant MF trials proved unviable. Finally, the MS utilised a comparative study of baseline EORTC scores from COMFORT-II and other commonly known cancers that demonstrated a similar symptomatic burden and HRQoL in patients with MF to that reported for AML, CML and breast cancer patients. The manufacture then identifies and selects metastatic breast cancer patient utilities for use in the base case; other utilities are presented for sensitivity analysis within the model.

Base case results were presented as an incremental cost-effectiveness ratio (ICER) for ruxolitinib compared with BAT. The results showed that ruxolitinib has an ICER of £73,980 per QALY compared with BAT. Sensitivity analysis conducted by the manufacturer for this population failed to significantly reduce the ICER of ruxolitinib compared to BAT. The full range of sensitivity analysis conducted by the manufacturer will be presented in Section 5.2.9, Sensitivity analyses.

Table 5.1 Summary of the manufacturer's economic evaluation (and signposts to MS)

	Approach	Source/MS Justification	Signpost (location in MS)
Model	Cost-utility analysis using a 35 year Markov model.	The model was used to simulate the treatment of MF.	Sections 7.2.2 and 7.2.3, pp. 129-30
States and Events	The model was comprised of four mutually exclusive health states. Complications were added to non-responders. Patient response by reduction in splenomegaly informed state transitions.	The manufacturer indicates that the choice of response criteria is not ideal, but feels that it is the best available evidence from the COMFORT-II trial.	Sections 7.2.3 and 7.2.4, pp. 129-32
Comparators	Ruxolitinib was compared to BAT.	There is no currently established care pathway for MF, so a combination of currently used treatments comprises BAT.	Section 7.2.7, pp. 135-6
Subgroups	The manufacturer conducted a subgroup sensitivity analysis using individual and combinations of BAT study drugs.	The manufacturer cites BCSH guidelines indicating drug therapies in use for the choice of subgroups.	Sections 7.2.7 and 10.14.4, pp. 135-6 and pp. 288-9
Treatment Effectiveness	The manufacturer used splenomegaly reduction, overall survival, presence of adverse events, and complications of MF to show clinical effectiveness.	The manufacturer did not have reliable data on symptomatic burden for patients in COMFORT-II.	Section 7.3.1, pp. 133-46
HRQoL	The manufacturer uses metastatic breast cancer utilities in their base case and evaluates CML, and NHL utilities in sensitivity analysis.	No EQ5D data was available for MF. The manufacturer cites a study comparing EORTC-QLQ30 similarities between the disease areas as justification.	Sections 7.4.1-7.4.8 and 7.4.9, pp. 153-160 and pp. 160-1
Adverse Events	Grade 3 and 4 adverse event data was gathered from the COMFORT-II trial and input into the model.	COMFORT-II is the primary efficacy data for ruxolitinib.	Section 7.3.1, pp. 144-5
Resource Utilisation and Costs	The manufacturer derives costs from COMFORT-II cross-referenced with the BNF, NHS reference costs and PSSRU costs.	UK costs were used to value trial events.	Section 7.5, pp. 164-178
Sensitivity Analysis	The manufacturer conducts a suite of deterministic and scenario analysis. PSA is presented for the base case and an alternative scenario.	Guidelines and clinical consultation are cited as justifications for deterministic and scenario sensitivity analyses.	Sections 7.6, 7.7.7-7.7.11, 7.9.4, and 10.14; pp. 179-82, pp. 194-198, pp. 203-4, and pp. 284-9

5.1 ERG comment on manufacturer's review of cost-effectiveness evidence

The manufacturer conducted a systematic literature review to identify studies of the cost-effectiveness of any intervention in patients with MF. The inclusion and exclusion criteria are reported in Table 5.2.

The manufacturer's submission described the search strategies used to identify relevant economic evaluations/resource use studies and quality of life studies on the treatment of primary myelofibrosis, post-polycythaemia vera myelofibrosis and post-essential thrombocythaemia myelofibrosis. Strategies were only briefly described in the main submission, however full details were provided in the Appendices.

The electronic databases MEDLINE, MEDLINE In Process, EMBASE, BIOSIS, EconLit, the Cochrane Central Register of Controlled Trials, the NHS Economic Evaluation Database (NHS EED) and the Health Technology Assessment database were searched. In addition to this, abstracts of conference proceedings, clinical trial registers and The National Institute for Health and Clinical Excellence (NICE) website were reviewed.

Searches were conducted on 27 July 2011. A search update was also performed on 19 March 2012. Search strategies for each database were documented in Tables D13 - D17 as stated in the manufacturer's submission. The searches covered the period 1960 - 19 March 2012, were not limited by language of publication, and excluded animal-only studies. For identification of cost-effectiveness and resource use studies, an additional publication date filter of 1 January 2000 to 19 March 2012 was applied to eliminate older studies which were unlikely to be relevant to current treatment practices, resource patterns and costs. For quality-of-life studies, no date limit was applied.

Overall the searches were appropriate and comprehensive, and included the use of both subject indexing terms and free text searching. Field searching, Boolean operators and truncation were used where required. All the required databases were searched, as well as the additional databases Biosis, the Cochrane Central Register of Controlled Trials and the Health Technology Assessment database.

For both the clinical and cost effectiveness searches, study design limits were applied to the Cochrane Library databases search strategies in order to be consistent with the usage of terms in the MEDLINE and EMBASE searches. These filters are however redundant as the Cochrane Library databases are already limited by study design. Further methodological filters are therefore superfluous, and may exclude potentially useful records.

Within the documentation of the cost effectiveness searches, it is not clearly marked which results sets relate to the economic studies, and which relate to the quality of life studies. The two sets of searches appear to have been conducted within one search for each database and subsequently separated so that date limits could be applied. This could have been more clearly indicated or highlighted within the manufacturer's submission for transparency.

There also appears to have been some error in documenting the results for the 2012 update searches, as these do not tally with the record numbers provided in Figures 40 and 41. For the economic systematic review update searches, the number of records documented in the strategy write-up totals 44 additional records. In Figure 40, this is listed as 29 records found. For the quality of life update searches, the number of records documented in the strategy write-up totals 55 additional records. In Figure 41, this is listed as 28 records found.

No clear objective for the review was stated; although the identification of resource use was clearly one aim. However, the ERG anticipate that the aims of the review included formulating the structure of a de novo model of a disease, and populating the parameters used to generate cost-effectiveness results in that model. Further, comparing the results obtained to previously published results can demonstrate the external validity of a model.

Given these likely objectives, limiting by date may not have been appropriate. In addition, a wider search of modelling studies may have aided the development of the MS *de novo* model. Given no clear objective was stated, if the objective was to identify relevant modelling methodology a wider search strategy/inclusion criteria may have been warranted. A clear objective would have allowed a better assessment regarding the appropriateness of the review. However, the ERG does not consider the conduct of the review to be a major flaw.

It would appear from the PRISMA diagram presented (MS, pg.276, Appendix 10, Section 10.10) that the screening of cost-effectiveness and resource use studies were done concurrently. It is therefore not possible to clearly define how many cost-effectiveness studies were identified, but later excluded. The manufacturer stated that no relevant cost-effectiveness studies were identified (MS, pg.125). In the absence of published studies, the manufacturer’s de novo model was the focus of their submission.

5.2 Summary and critique of manufacturer’s submitted economic evaluation by the ERG

The evaluation conducted by the manufacturer combines the clinical and economic data to evaluate the cost-effectiveness of ruxolitinib for the treatment of patients diagnosed with intermediate-2 or

high-risk MF. The NICE scope clearly states that ruxolitinib should be evaluated within its licensed indication. It should be noted that the population considered is narrower than that of the licensed population, but does reflect the relevant UK decision problem (see Section 3.1).

The remainder of this section provides a summary and critique of the de novo model presented in the MS. A summary of the NICE reference checklist with the ERG's comments on whether the manufacturer's de-novo model has been judged to fulfil the NICE reference case is presented in Table 5.2.

5.2.1 NICE reference case checklist

Table 5.2 NICE reference case

Elements of the economic evaluation	Reference case	Does the de-novo evaluation match the reference case?	ERG comment on whether de-novo evaluation meets requirements of NICE reference case
Comparators	Therapies routinely used in the NHS, including technologies regarded as current best practice	Partially	The ERG's clinical expert advised that lenalidomide is a rarely used treatment for MF in the NHS. Otherwise, the treatments comprising BAT are representative. The NICE scope[28] indicated that a comparison with allogeneic haemopoietic stem cell transplantation would be desirable.
Type of economic evaluation	Cost-effectiveness analysis	Yes	
Perspective on costs	NHS and PSS	Yes	
Perspective on outcomes	All health effects on individuals	Yes	
Time horizon	Sufficient to capture differences in costs and outcomes	Yes	The length of the model is sufficient to capture differences in costs and outcomes. However it is not clear to the ERG whether a 35 year time frame is too long for this population.
Synthesis of evidence on outcomes	Systematic review	Yes	
Measure of health effects	QALYs	Yes	
Source of data for measurement of HRQoL	Reported directly by patients and/or carers	Partially	Utilities were derived from metastatic breast cancer, not MF.
Source of preference data for valuation of changes in HRQoL	Representative sample of the public	Yes	
Discount rates	Annual rate of 3.5% on costs and health effects	Yes	
Equity weighting	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Yes	
Sensitivity analysis	Probabilistic sensitivity analysis	Yes	

5.2.2 Population

The NICE scope defined the population of interest as adults with disease-related splenomegaly or symptoms of PMF, PPV-MF and PET-MF, which is within the licensed population. In the UK, ruxolitinib is licensed for the treatment of splenomegaly and symptoms in MF patients at a dosage of 5-25 mg twice daily.[27]

As stated in Section 4.2.1, although the stated population matches the NICE scope the evidence presented in the MS is derived from clinical trials whose populations may represent only a subset of the licensed population. The three main trials from which data were derived reflect three different populations:

- COMFORT-I - intermediate-2 or high-risk patients who were refractory to all other therapies and the comparator was placebo;
- COMFORT-II - intermediate-2 or high-risk patients who were or were not refractory to other therapies and the comparator was BAT which could be no therapy;
- Phase I/II - patients who were or were not refractory to other therapies and had intermediate-1, intermediate-2 or high-risk disease.

Data from these populations has not been synthesised, but rather used selectively to independently inform model inputs. The appropriateness of these data are discussed in Section 5.2.5, Model inputs. The intermediate-2 and high-risk groups have been evaluated as one population.

5.2.3 Interventions and comparators

The MS includes two interventions in the base case analysis, ruxolitinib and BAT. Ruxolitinib therapy consists of twice daily ruxolitinib at a 5-25 mg dosage, while BAT consists of a variety of treatments for MF. The details of the treatments comprising BAT are summarised and evaluated in this section. The ERG will undertake further analysis on the options available within BAT, which will be presented in Section 6.

A summary of treatment included in BAT in the COMFORT-II trial is presented in Table 5.3. It is not clear from the data in what order the treatments were received, how long patients remain on each treatment, nor how many treatments each patient might receive. Information the ERG found from the CSR for COMFORT-II indicates that 33.9% of patients in the trial received no active treatment; this does not appear to have been accounted for in the model. In addition, the clinical expert on the ERG team indicated that lenalidomide is very rarely used in UK practice and the HMRN audit appears to

confirm this assertion: no patients in the HMRN audit received lenalidomide.[29] The ERG questions the inclusion of lenalidomide as a treatment option within the BAT bundle, and more particularly the presentation of a sensitivity analysis of ruxolitinib compared with thalidomide/lenalidomide and HU. The ERG considers these two sensitivity analyses to be unjustified and inappropriate.

It is also clear from the published literature that there are other treatments used in the UK which are not included in the BAT bundle. Of note, the BCSH guidelines indicate that splenectomy and splenic irradiation are treatments for splenomegaly.[3] It should be noted that rather than being included as comparators, these have been modelled as complications of splenomegaly in the MS. The ERG feel that these excluded comparators should have been considered either within the BAT bundle or as alternative comparators. At a minimum, sensitivity analysis was warranted.

Table 5.3 Interventions comprising BAT in the MS (Table B26, pg.170 in MS)

Treatment	Proportion of patients (%)
Hydroxyurea	50.4%
Anagrelide	6.0%
Prednisone	9.9%
Methylprednisolone	2.8%
Prednisolone	0.8%
Epoetin alfa	8.3%
Thalidomide	3.2%
Lenalidomide	2.0%
Mercaptopurine	3.6%
Thioguanine	2.0%
Danazol	5.6%
Peginterferon alfa 2a (Pegasys)	1.2%
Melphalan	2.4%
Acetylsalicylic acid	2.8%
Cytarabine	0.8%
Colchicine	0.8%
Deferasirox (Exjade)	0.4%
Folic acid	1.2%
Lysine acetylsalicylate	0.8%
Interferon alfa 2a (Roferon-A)	1.2%

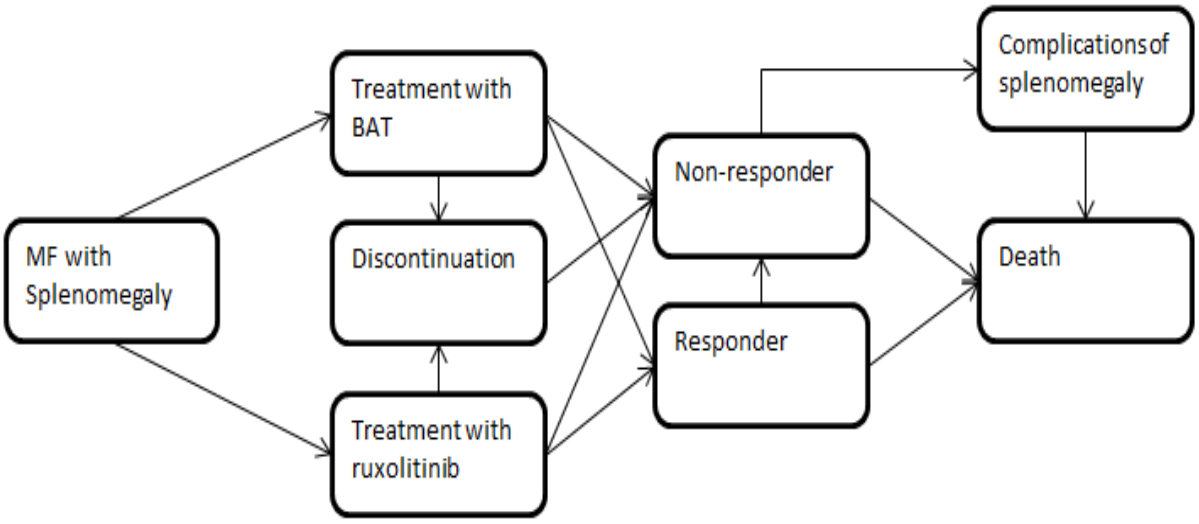
5.2.4 Model structure

The manufacturer created a state-transition Markov model to simulate the natural course of MF. The model uses 12 week cycles and a 35 year time horizon designed to simulate a lifetime time horizon. The model currently allows results to be generated for 24, 48, 96 and 144 week time horizons, in addition to the 35 year base case. Markov models are appropriate and commonly used for this type of analysis due to their ability to capture effects that happen over long time horizons and to extrapolate beyond shorter term trial data. The use of a Markov model is appropriate for chronic progressive conditions such as MF, although in this instance the ERG feels that the progression of MF has been oversimplified to the point where disease progression is not captured. As a result the ERG feel that they model structure is inappropriate to fully address the decision question.

The base case model contains four mutually exclusive health states: 'responder', 'non-responder,' 'discontinuation,' and 'death.' Figure 5.1 presents a diagrammatical representation of the model.

All patients begin the model at diagnosis of MF with splenomegaly and are assumed non-responders to treatment. Patients receive either ruxolitinib or BAT therapy for an initial treatment phase of 24 weeks. During this initial 24 week treatment phase all patients are treated with either ruxolitinib or BAT and treatment continues regardless of response. Patients move through two 12 week cycles and accrue costs and utilities associated with their response status. Any patient who discontinues treatment during this initial treatment phase (due to AEs, withdrawal of consent and clinical decision to stop treatment) moves to the discontinuation state and accrues costs and utilities of a non-responder. At 24 weeks (2 cycles of the model) a stopping rule is applied. All progression through the model beyond this point is determined by the patients' response status at 24 weeks. The discontinuation state is inactive for the duration of the longer-term stage of the model.

Figure 5.1 Simplified diagram of the model structure (Figure 26, pg. 128 in MS)



The stopping rule means that at 24 weeks:

- all patients achieving a response, defined as a 35% reduction in spleen volume are considered to be responders, move into the responder state and are treated with ruxolitinib therapy;
- all patients not achieving a response move to the non-responder state and are treated with BAT;
- all patients who discontinued treatment during the initial 24 week treatment phase move from the discontinuation state to the non-responder state and are treated with BAT.

The model was constructed to allow the time at which the stopping rule was applied to be varied to 12 weeks; this scenario was presented as sensitivity analysis. After the stopping rule has been applied the transition of patients through the model health states is subtly different.

Responders are assumed to remain treatment responders for the duration of the model, unless they discontinue treatment or die. Any responder discontinuing treatment moves to the non-responder state and is assumed to be treated with BAT for the duration of the model. Those patients who die move to the death state. The model assumes that patients who achieve a response (i.e. those in the responder state) do not experience complications of MF.

Non-responders are assumed to remain in the non-responder state, receiving BAT treatment, for the duration of the model unless they die and are absorbed into the death state. There is no option to discontinue BAT treatment after the stopping rule has been applied (24 weeks base case). Complications can be experienced by those in the non-responder state; such complications included portal hypertension, sepsis, infection and the need for interventions such as splenectomy or splenic irradiation. Experiencing one of these complications incurs a utility decrement and resource use. Complications were considered separately from adverse effects. This will be discussed later in Section 5.2.5. LT was not included in the base case, but a sensitivity analysis exploring its impact was presented in the manufacturer's sensitivity analysis.

Whilst the ERG appreciate that the lack of robust data has limited the modelling that could be undertaken, some of the basic underlying assumptions of the model may be too limiting. These include the assumptions that

- the inability of the responder and non-responder health states to capture disease progression
- patients responding to treatment maintain the level of spleen volume reduction observed at week 24 and the associated utility benefits (i.e. outcomes are constant over time)
- all rates after the initial treatment phase (24 weeks), overall survival, discontinuation and rates of complications are assumed constant over time
- only non-responders are at risk of the complications of MF, some of which may not be eliminated by a reduction in spleen size
- non-responders remain on BAT for the duration of the model (base case 35 years)

The ERG believes that these simplifications make the result of the modelling presented in the MS highly spurious. The assumptions appear to have been dictated by data availability rather than clinical plausibility. The use of a 35 year time horizon in the base case meets the NICE reference case, but the ERG feel the time-horizon may actually be too long given the average survival of this population. Clarification was sought from the manufacturer in the points for clarification document.
The evidence presented suggests that there is uncertainty surrounding the survival data for this population. The MS highlights that the median survival for patients with intermediate-II and high-risk MF is less than 5 years. However the manufacturer argues quite strongly that a time horizon of less than 35 years for the model would be inappropriate. In the manufacturer's response document the manufacturer provides base case results for two alternative time horizons (10 and 15 years). These are presented in Table 5.4. As you would expect the decrease in the time horizon results in an increase in the ICER. The uncertainty surrounding the time frame over which ruxolitinib patients achieve a benefit needs to be reduced. Currently, no long term data are available but the ERG feel that it may be

unrealistic to suppose that patients will continue on treatment and achieve a treatment response/benefit for 35 years.

Table 5.4 MS base case over 10 and 15 years' time horizons

Time horizon	35% response (24 weeks)
Lifetime (35 years)	**£73,980**
10 years	£81,308
15 years	£77,036

Figures in bold are the results of the manufacturer's base case

In addition, the need to extrapolate from such short-term uncertain data introduces a high-level of uncertainty. To further compound the issue several assumptions regarding constant rates have been made, many of which the ERG feel are implausible. The model inputs and associated assumptions along with their strengths, limitations and implications of their use will be discussed in the next section (Section 5.2.5). Where possible the implications of relaxing assumptions will be tested and presented in Section 6. However, it should be noted that the model structure is inadequate and does not capture disease progression; therefore, the additional analysis undertaken by the ERG should be considered with this major limitation in mind.

5.2.5 Model inputs

This section evaluates the inputs within the model that define clinical and cost-effectiveness. Clinical events such as response to treatment, AEs, and complications affect QALY and cost data within the model.

Response rates

A key driver of the model is treatment response. The MS defines response as a 35% reduction in spleen volume. This is a much narrower definition of response than that used in clinical practice, or defined within the IWG-MRT clinical improvement criteria, see Table 2.2, Background Section. The IWG-MRT consensus criteria for treatment response include relief of splenomegaly as one possible clinical improvement criteria; patients may also improve by other means.

As ruxolitinib treatment alleviates splenomegaly it consequently alleviates some of the symptoms of MF associated with splenomegaly. However, other symptoms which are not associated with splenomegaly are not necessarily alleviated by ruxolitinib. To some extent it is these other quality of life symptoms which the comparator treatments included in BAT aim to relieve. Ruxolitinib is the

only licensed treatment which specifically targets reduction in spleen volume; however reduction in spleen is not the only outcome of clinical significance in this population. The MS highlights issues with deriving a composite measure, but sensitivity analysis around alternative definitions of response such as symptom relief could have been and should have been investigated. The ERG feels that the use of a composite measure, as acknowledged in the MS, would have been more appropriate. The ERG will undertake an analysis using TSS and PGIC to inform response rates. See Section 6 for further explanation and results.

Other issues with the measurement of splenomegaly response include choosing MRI over palpation as a measurement tool, and the inclusion of patients who begin with a palpable spleen length between 5 cm and 10 cm as responding if they had a ≥50% reduction in palpable spleen length, which does not meet IWG-MRT criteria,[17] and the existence of a baseline by treatment interaction previously discussed in Section 4.2. Their impact is explored in Section 6.

The base case model uses data from COMFORT-II to inform response rates during initial treatment and to estimate transitions to the responder and non-responder model states (MS, Table B15, pg. 140). Response rates were derived from what the MS refers to as a per protocol analysis of COMFORT-II data. They state that per protocol analysis was used due to crossover confounding response rates in the BAT arm (MS, pg.139). The percentages of patients achieving response according to this definition are summarised in Table 5.4. However, it should be noted that the analysis undertaken to derive these data does not correspond to the per protocol analysis defined or undertaken in the clinical trial report. Rather it represents an analysis of all patients for who spleen reduction outcomes were available at a given time point. This equates to less patients than the per protocol analysis. These data will be discussed further in Section 6.

Table 5.4 Percentage of patients ≥ 35% reduction in spleen volume, per protocol analysis of COMFORT-II trial, IPSS classification

Treatment up to week	High-risk ruxolitinib	High-risk BAT	Intermediate-2-risk ruxolitinib	Intermediate-2-risk BAT
0	[redacted]	[redacted]	[redacted]	[redacted]
12	[redacted]	[redacted]	[redacted]	[redacted]
24	[redacted]	[redacted]	[redacted]	[redacted]

Discontinuation rates

The model utilised two alternative discontinuation rates, one for the initial 24 week treatment phase of the model and one for the responder state post 24 weeks. The rate used in the initial phase of the model was derived from the COMFORT-II trial. The number of patients discontinuing was reported in the MS as 13 patients out of 146 at 48 weeks (MS, pg143). This number did not match that reported in the flow diagram presented in the MS, pg. 75. This in turn did not match the data reported in Harrison et al.,[22] which was referenced in the MS as the data source. The ERG will undertake sensitivity analysis using the alternative data which will be presented in Section 6.

The discontinuation rate applied to the responder state was derived from the Phase I/II study.[30] The use of these rates rather than those from COMFORT-I and II was not justified in the MS. The long-term follow-up data from the COMFORT trials, indicated higher discontinuation of treatment rates for ruxolitinib than the rates demonstrated in the Phase I/II trial.[30] Given that COMFORT-II informs the primary efficacy inputs used in the model, the ERG feels that discontinuation rates from COMFORT-II may have been more appropriate and at a minimum should have been considered in sensitivity analysis. The effect on the ICER of using these alternative discontinuation rates in the post initial treatment phase of the model will be explored in the ERG's additional analysis; see Section 6.

Overall survival

The MS states that overall survival data from the COMFORT trials was not mature enough, insufficiently powered, plagued by missing values, and too confounded by crossover to be able to demonstrate any mortality benefits between ruxolitinib and the comparator treatment groups (BAT or placebo). However, the ERG requested updated survival data from the COMFORT trials in the points for clarification document and the manufacturer provided the updated survival data from both the COMFORT-I and II trials. These data will be used by the ERG to undertake an alternative analysis which will be presented in Section 6.

The remainder of this section on survival will discuss the data used in the MS base case. The base case model required data on survival split by responder status. Overall survival for those in the non-responder/BAT therapy state was based on a large multicentre study by Cervantes et al.[7] which involved the development of a highly discriminative prognostic system (IPSS) in 1054 patients consecutively diagnosed with PMF at 7 centres (UK, France, Spain, Italy (3), the USA) with an overall median survival of 69 months. Patients with PPV-MF or PET-MF were not included in this study. As discussed in Section 4, in the COMFORT-II trial, 52.7% of patients had PMF, 32.9% had PPV-MF, and 14.4% had PET-MF in the ruxolitinib arm, with similar proportions in the BAT arm. Cervantes et al. estimated a median survival of 27 months and 48 months for high and intermediate-2

risk MF patients, respectively, who had received treatment with BAT. However, the population does not include PPV-MF or PET-MF patients,[7] and the treatments received in the population are accumulated from 1980 to 2007. Hence it is not clear if the survival data is relevant to current best available therapy. The population from the Cervantes et al.study[7] only includes PMF patients which is only a subgroup of the patients in the efficacy trials. Whether the differences in the populations create different results is not known at this time, but indicators from COMFORT-I and II[21-22] suggest that PPV-MF and PET-MF patients may have a higher median response rate (although this could be by chance), and the Phase I/II study[30] demonstrated that spleen reduction correlated with longer survival (see Section 4.2.3).

The 12-week probability of death for non-responders was calculated from the median survival assuming an exponential distribution. The MS states that it was not possible to examine the impact of other parametric distributions. Whilst the ERG are uncertain why other distributions could not be assessed, the graph depicting the exponential distribution provided by the manufacturer in response to the ERG's clarification letter looks reasonable, however the ERG will undertake sensitivity analysis to highlight the uncertainty. The results of these analyses will be discussed and presented in Section 6.

Table 5.5 Mortality probability applied to patients in the 'non-responder' health state (MS, Table B17, pg. 143)

	IPSS classification [7]	
	Median survival	**12-week probability of death**
High-risk	27 (95% CI 23 to 31) months	6.86% (95% CI 8.01 to 6.00)
Intermediate-2	48 (95% CI 43 to 59) months	3.92% (95% CI 4.37 to 3.20)

CI, confidence interval; IPSS, International Prognostic Scoring System

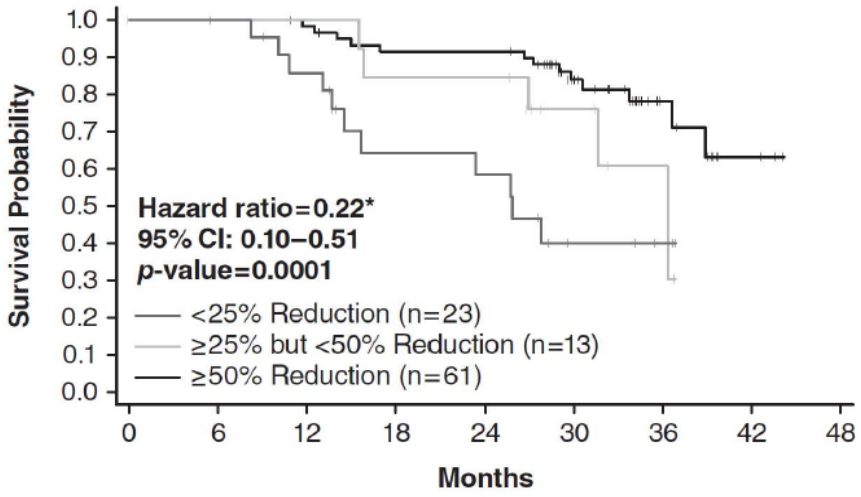
The MS applied the same probability of death at each 12 week cycle for all non-responders, regardless of whether the non-responder had initially receive ruxolitinib as a treatment or BAT; this was justified by the manufacturer on the basis that ruxolitinib patients who were non-responders moved to the BAT arm after just 24 weeks (see Table 5.8). The ERG feels that until more mature survival data are reported it is unclear whether this assumption is conservative or not. On one hand, patients who are treated with ruxolitinib and who continue to respond for several years before having their treatment stopped might actually achieve a slightly better survival compared with patients who never achieve a response with ruxolitinib, or whose duration of treatment response is short/shorter. On the other hand, patients treated with ruxolitinib have a higher risk of AEs (compared to BAT) and this risk as well as

deaths related to AEs and their associated treatment might increase with treatment duration. In light of the lack of evidence the assumption made in the MS is reasonable.

Survival in the MS base case for those in the responder state was calculated based on the Phase I/II study. Response in this study was defined as a 50% or greater reduction in palpable spleen length, which, as discussed in Section 4, approximates a ≥35% MRI volume reduction. The study by Verstovsek et al.[30]had a cohort of 107 ruxolitinib treated patients with a median follow-up of 32 months. The approximate HR for mortality between responders and non-responders was 0.3. A HR of 0.22 (95% CI: 0.01 to 0.51) was reported for the original comparison of responders with ≥ 50% reduction versus non-responders with <25% reduction; this was modified by the manufacturer to include all patients who did not achieve a ≥ 50% reduction in palpable spleen length (Figure 5.2). The manufacturer assumed identical HRs for high and intermediate-2 risk groups.

Alternative HRs were explored in sensitivity analysis but only within the 95% confidence interval of the Phase I/II trial. Alternative survival data such as that from COMFORT-I and II was not explored. Survival data was requested from the manufacturer in the ERG's clarification letter. The manufacturer provided survival HRs from the COMFORT-I and II trials using updated survival data to beyond 100 weeks. The impact of these new data will be explored by the ERG and presented in Section 6.

Figure 5.2 Survival by level of response from the MDACC cohort of the Phase I/II trial (MS, Figure 27, pg. 142)



*Comparison of <25% Reduction to ≥50% Reduction.

The median survival for non-responders, derived from the Cervantes study[7], was combined with the derived HR from the Phase I/II trial (0.3) to calculate the 12-week probability of mortality by IPSS class (high-risk or intermediate-2) applied to the responder state, see Table 5.6.

The ERG identified a number of potential concerns regarding this basis for deriving the estimates of overall survival which were used in the model. In particular, the current approach to extrapolation (non-responders overall survival) assumes that overall survival follows an exponential distribution. No supporting evidence is provided to justify or test this assumption nor did the MS include the testing of different distribution assumptions. Also, the HR calculated by the manufacturer and used in estimating overall survival among responders versus non-responders in the model remained constant with respect to time (i.e. the same mortality rate is applied for 35 years). These issues are important, since the disparity between the model cost-effectiveness results for shorter time horizons (such as at 144 weeks) and 35 years suggests that the main cost-effectiveness advantage is conferred in the period of extrapolation. The ERG believes that these assumptions may be too bold and will undertake further analysis, see Section 6.

Table 5.6 12-week probability of mortality applied to patients in the responder health state (MS, Table B17, pg. 144)

	IPSS classification
High-risk	2.13%
Intermediate-2	1.21%

Adverse events and complications

The MS model considers AEs and complications in different ways. AEs could be experienced by both responders and non-responders. The MS states that the model incorporates the most frequent grade 3 and 4 AEs reported in the COMFORT-II trial. There are no utility decrements assigned to these AEs as it is assumed that they are short-lived and managed through treatment titration and interruption. The trial report for COMFORT-II states that 'grade 3-4 AEs were more frequently observed in the ruxolitinib arm compared to the BAT arm (ruxolitinib: 41.8%; BAT: 24.7%)'. The model used the 48 week AE rates (see Table 5.7) from the COMFORT-II trial to derive 12 week probabilities. The costs associated with these AEs are presented in Section 5.2.7.

Table 5.7 COMFORT-II 48 week grade 3/4 adverse event rates (MS, Table B18, pg 145)

AE	**Ruxolitinib**	**BAT**
Neutropenia	0.0%	0.0%
Fatigue	0.68%	0.0%
Nausea/vomiting	0.7%	0.0%
Diarrhoea	1.4%	0.0%
Anaemia	11.0%	4.1%
Thrombocytopenia	7.5%	4.1%
Pneumonia	1.4%	4.1%
Asthenia	1.4%	1.4%
Stomatitis/gastroenteritis	1.4%	0.0%

The MS provided no clear rationale for what appears to the ERG to be an ad-hoc selection of AEs to include. The MS provides a full list of all AEs and Grade 3-4 AEs in Table B12, pg.111 to 112 in the MS, which the ERG has presented here. As can be seen from the table there may be justification for including some of the excluded AEs, given the rates at which they were experienced and the differential between the ruxolitinib and BAT groups. Whilst it is plausible that the inclusion of all

AEs or all Grade 3-4 AEs may not have a huge impact on the overall result of the model, it is clear that AEs incur costs within the model. Some sensitivity analysis on the impact of incorporating all AEs may be warranted. Some of the excluded AEs, such as pyrexia may have little impact on costs but further justification is warranted.

Table 5.8 Summary of Table B12, MS. Pg111 48 week adverse event rates from COMFORT-II

Non-haematological adverse events (≥ 10% of ruxolitinib-treated patients)	Ruxolitinib % Any grade (grade 3/4)	BAT % Any grade (grade 3/4)
Diarrhoea	23(1)	12(0)
Peripheral oedema	22(0)	26(0)
Asthenia	18(1)	10(1)
Dyspnoea	16(1)	18(4)
Nasopharyngitis	16(0)	14(0)
Pyrexia	14(2)	10(0)
Cough	14(0)	15(1)
Nausea	13(1)	7(0)
Arthralgia	12(1)	7(0)
Fatigue	12(1)	8(0)
Pain in extremity	12(1)	4(0)
Abdominal pain	11(3)	14(3)
Back pain	10(2)	11(0)
Headache	10(1)	4(0)
Pruritus	5(0)	12(0)
Anaemia (grade 3 or 4)	42	31
Thrombocytopenia (grade 3 or 4)	8	7

Unlike AEs, complications were assigned both costs and utility decrements within the MS base case model. However, complications of MF were assigned only to non-responders. The manufacturer justifies this assumption by stating that complications are due to splenomegaly. This justification is not entirely true as infection and sepsis are related to blood cell counts, which is independent of splenomegaly. Table 5.9 presents those complications considered under the 'other complications' umbrella in the manufacturer's model. The total rate of other complications in Table 5.9 was summed and converted into a 12-weekly probability of 13.5% (MS, pg. 146) for use in the model. These other complication rates were taken from the ongoing HMRN audit.[29]

Table 5.9 Other complications from MF (Signpost to MF, pg.146, Table B19)

Complications	Frequency	Complication rate (per person per year)
Infection	41	0.26
Sepsis	3	0.02
Hyposplenism/infarction	15	0.1
Oesophageal varices	4	0.03
Gastric varices	2	0.01
Pleural effusion	15	0.1
Portal hypertension	3	0.02
Melaena	4	0.03

In addition to the complications presented in Table 5.9, the model evaluates the common interventions which it states 'are undertaken to resolve the complications of massive splenomegaly' such as splenectomy, and splenic irradiation. Estimates of splenectomy, and splenic irradiation were derived from an unpublished US chart review.[31] Annual splenectomy and splenic irradiation rates from this data source were converted to 12-week probabilities of 0.75% and 0.43%, respectively, for use in the model. The 12-week probability was assumed constant over the duration of the model (35 years in the base case). These rates may not be valid for the UK population; these interventions are not commonly used.[15]

It should also be noted that splenectomy requires a very specific set of conditions that a patient must meet to undergo the procedure. The ERG feels that it is unlikely that patients will have the same probability of splenectomy as they age. Assuming constant rates of complications through 35 years seems bold and likely to be inappropriate. The ERG feels that it would have been more appropriate to model splenectomy using a separate state in the model. Those patients undergoing splenectomy can do so only once (a fact which has not been accounted for in the model), and may achieve greater relief of symptoms and improved quality of life, not just temporary utility decrements and costs. Splenectomy, if successful, has been shown to reduce symptom burden by eliminating the spleen.[32-33]

Splenic irradiation has been shown to reduce spleen size and alleviate symptom burden in MF patients[34-35]and so, like splenectomy cannot be assumed to incur only decrements and costs, without benefits. In addition, given the nature of these interventions it is possible that a patient's survival post-splenectomy or splenic irradiation may differ from those patients not receiving these interventions.

In summary, the rationale behind the MS inclusion of both splenectomy and splenic irradiation as complications is not fully justified. The BCSH guidelines indicate that splenectomy and splenic irradiation are treatments for splenomegaly, not complications of the condition.[3] At present the ERG believes that the model does not fully capture the impacts of these treatments in terms of costs and outcomes. The ERG feels that the rationale for not considering these as alternative treatment comparators in the model has not been fully justified and the method of including these options has been oversimplified.

Leukaemic transformation

LT was not included in the MS base case, but was presented as a sensitivity analysis. To include LT the manufacturer employed an additional health state in the model. Annual LT probabilities were derived from the Phase I/II study[36] This study reported annual LT rates of 3.6% for MF patients treated with ruxolitinib and 3.8% for MF patients in the historical control. The manufacturer stated that since their own review did not identify any studies that analysed LT rates by spleen size they therefore assumed these rates in the model (i.e. probabilities of 3.6% and 3.8% for responders and non-responders, respectively). The survival estimate following LT was 3.9 months for responders and non-responders alike. [37] Following LT a utility decrement was applied. Patients entering this state can remain in the LT state as a responder or a non-responder until death.

Although the final scope stated that "at later stages of the disease, up to 23% of patients may transform to AML", the MS justified the exclusion of LT from the main analysis by referencing a discussion with clinical experts that indicated that it was uncertain how ruxolitinib would affect the incidence of LT. The ERG feel that this exclusion was not justified, and further, the ERG clinical expert indicated that the conservative assumption is equivalent rates of LT rather than the manufacturer's assumption of different rates presented in their base case. Results assuming equivalent rates of LT will be presented for all of the additional analyses (Section 6) undertaken by the ERG.

5.2.6 Health related quality of life

In the MS the main health benefit assessed was quality-adjusted life-years (QALYs). These were calculated using preference scores for the two main health states, responder and non-responder, with

utility decrements assigned for patients experiencing MF related complications. As previously stated no decrements were assigned to AEs. The MS considered that the impact that AEs would have on HRQoL would be small, that these AEs were expected given the mechanism of ruxolitinib, and generally that they would decline over time through dose reductions, treatment interruptions and blood transfusions (MS pg. 160).

Due to a lack of a preference-based measure in the COMFORT trials and a lack of published studies specifically in MF populations which included relevant data, HRQoL data for other similar populations was sought from published literature. Potential papers were found using a systematic search (MS, Appendix 10.10, pg. 262). As justification for selecting utilities from other disease populations the manufacturers provided an abstract by Kiladjian et al.[38] that compares EORTC scores from COMFORT-II to studies that gathered EORTC scores from other forms of cancer. Kiladjian et al.[38] compared mean EORTC-QLQ-30 scores at baseline (in terms of the functional scales, symptom scales and global health status/QOL) in COMFORT-II patients with reported baseline EORTC-QLC-30 scores for AML [39] CML[40] and breast cancer.[41] In this study it was reported that mean EORTC QLC-30 functional, global health status/QoL, and symptom scores for patients with MF in COMFORT-II were comparable to or no worse than have been reported for patients with AML, CML and breast cancer. Table 5.10 below summarises the results from the comparative study by Kiladjian et al.[38] which underpinned the selection of utilities used for the base case analysis.

Table 5.10 Mean EORTC QLQ-C30 scores at baseline in COMFORT-II and for patients with AML, CML, and breast cancer (reproduced from Kiladjian et al. [38]

Mean (SD) EORTC QLQ-C30 score	MF (N=219) (COMFORT-II)	AML (N=155) [39]	CML (N=73) [40]	Metastatic breast cancer (N=225)[41]
Functional status				
Physical functioning	68.0 (22.9)	-	78.0 (21.0)	86 (17)
Role functioning	66.2 (31.0)	-	78.1 (36.3)	85 (27)
Cognitive functioning	78.6 (23.5)	82.2 (18.9)	86.1 (22.1)	83 (20)
Emotional functioning	73.6 (23.6)	86.1 (18.5)	78.8 (22.6)	67 (21)
Social functioning	78.1 (27.0)	66.1 (31.0)	84.3 (25.7)	84 (21)
Symptom scales				
Fatigue	46.6 (28.1)	36.2 (22.7)	29.8 (25.4)	28 (21)
Pain	25.9 (29.9)	13.7 (20.4)	10.1 (19.6)	23 (22)
Nausea/vomiting	5.9 (14.3)	9.0 (18.3)	5.0 (12.9)	6 (12)
Dyspnea	35.8 (32.8)	11.3 (17.1)	15.5 (20.9)	15 (21)
Insommnia	33.7 (34.1)	20.4 (26.1)	26.9 (32.7)	27 (27)
Appetite loss	18.2 (28.2)	18.0 (30.5)	13.7 (28.8)	13 (22)
Constipation	9.3 (20.4)	7.9 (19.1)	9.6 (19.6)	11 (22)
Diarrhea	16.8 (26.6)	12.6 (25.1)	7.3 (19.4)	6 (14)
Global health status/QOL	53.7 (21.8)	-	70.2 (21.5)	45 (30)

Based on this study the manufacturer selected metastatic breast cancer utility data to inform the base case. This was justified based on the fact that at baseline the two populations (MF and breast cancer) appeared to have similar quality of life (MS, pg.157). However, the difference in quality of life between a breast cancer population and a metastatic breast cancer population was not discussed. The MS subsequently elected to use the metastatic breast cancer utility values from a previous NICE appraisal of eribulin.[1] Alternative utility values were explored in sensitivity analysis for CML and NHL. Whilst the MS provides a clear rationale for why they chose breast cancer as a representative population, they provide no rationale for the move to a metastatic population or the selection of these specific utility values.

Whilst the ERG acknowledges that the MS has selected the most recent NICE evaluations in these disease areas, some further justification should have been presented. Table 5.11 summarises the utility values applied in the MS. On further examination of the NICE appraisal for eribulin,[1] the base case utility estimates are apparently related to the revised utility estimates calculated by the ERG in the NICE appraisal for patients in the "responder" health state (0.823) and progression health state (0.446). For the utility values based on CML, the values represented patients in the "chronic phase – imatinib treatment" (0.854) or "chronic phase – imatinib treatment after loss of cytogenetic response" (0.854) and the "accelerated" phase of CML (0.5952). For the utility values based on NHL the utility values related to patients who were in the states "disease free" (0.88) and "active disease – relapsed" (0.62) in the appraisal of rituximab. The ERG feel that there is a high level of uncertainty surrounding the appropriateness of the utility estimates assumed to represent MF.

Table 5.11 Utility values used in the de-novo model (MS, Table B22, pg. 161)

Health state	Model base case	SA (CML)	SA (NHL)
Baseline (all patients start in the non-responder state)	0.446 (Eribulin appraisal) [1]	0.595 (Imatinib)[42]	0.62 (Rituximab appraisal) [43]
Responders	0.823 [1]	0.854 (Imatinib appraisal)[42]	0.88 (Rituximab appraisal) [43]
Non-responders	0.446 (Eribulin appraisa) [1]	0.854 (Imatinib appraisal)[42]	0.88 (Rituximab appraisal) [43]
Complications of MF	0.1 (decrement, manufacturer's assumption)	0.1 (decrement manufacturer's assumption)	0.1 (decrement manufacturer's assumption)

Further justification for the specific choices of utility data used in the base case and sensitivity analysis was sought from the manufacturer, but no information beyond what was already available was provided by the manufacturer. The MS states that it had been planned that utilities were to be derived from a mapping exercise using data collected from the pivotal trials, but in the model the utilities used were eventually derived from the literature instead. In the MS the reasons given by the manufacturer for not using estimates derived from the pivotal trials included:

- The fact that there were large numbers of missing HRQoL scores over time (61.5% of ruxolitinib and 42.4% of BAT patients had scores at baseline and week 48);

- Clinicians involved in the trial indicated that because of crossover the utility of people left in the BAT arm at week 48 could not be considered representative of people with MF because those left in the trial after crossover and dropouts were the least unwell patients;

- The mapping considered responders and non-responders and as such the utility of non-responders therefore reflect the quality of life of a large proportion of people treated with ruxolitinib, plus everyone treated with BAT, and is not representative of the quality of life of people treated with MF;

- Finally, the MS stated that clinicians and patients report that the key driver of quality of life in MF is the burden of symptoms, rather than spleen size (and measured using the MFSAF questionnaire for instance). The MFSAF was not employed in COMFORT-II and only partially employed in COMFORT-I, as a result MFSAF data were not incorporated in the mapping, and so a constellation of symptoms associated with MF was excluded from the analysis.

Two mapping exercises were identified by the ERG, both of which appear to have been undertaken by the research group that built the manufacturer's model using EORTC-QLQ-30 data from COMFORT-II to map to the EQ-5D and EORTC-8D preference based measures, presented as a poster, and published as an abstract concurrent with the ISPOR meeting in Chicago in June 2012 (Roskell et al,). [2] The mapping exercise results were not referred to or presented in the MS. In fact, the manufacturer claimed the mapping was unviable. The ERG sought clarification on why values from the mapping exercise were omitted. The answer from the manufacturer conveyed that the mapping exercise did not separate responders and non-responders. However, the conference abstract would suggest that this is not the case. The results presented in the abstract gave EQ-5D values of 0.754 for responders and 0.670 for non-responders. The differential between responder and non-responder utilities is a key driver of the model. The utilities used in the base case have the largest difference between the two groups (responders and non-responders). Further analysis of the model using alternative utility values will be undertaken by the ERG and presented in Section 6.

The manufacturer's model also assumes that utility gains are constant over time. This assumption may not adequately capture the progressive nature of MF, worsening of symptoms and continuous reduction in HRQoL, natural aging, and development of complications experienced by MF patients over the longer term. As a consequence, the ERG has concerns regarding the appropriateness of using the health utility values applied in the model for the calculation of total QALYs.

5.2.7 Resources and costs

The economic model included the following direct cost categories:

- drug acquisition costs (excluding any administration or monitoring related costs)
- the costs of managing MF responders and non-responders
- the costs of blood transfusions
- the costs of palliative care
- the costs of the main complications of splenomegaly (splenectomy, splenic irradiation, and "other")
- the costs of treating AEs

Resource use is based on data from the ongoing HMRN audit, COMFORT-II and clinical opinion (with experience in managing MF). The HMRN audit involved two adjacent UK Cancer Networks, a total population 3.6 million and collected information about patients diagnosed with a haematological malignancy since 2004.[29]As stated in the MS, this was the only source available of UK-specific resource use on management of MF. Data on resource use was also obtained from COMFORT-II.

Ruxolitinib treatment costs

The 12-week drug costs for patients treated with ruxolitinib were estimated based on the starting doses as defined in the SPC and the actual dose usage in COMFORT-II. In the trial, dose increases were in increments of 5 mg twice daily (maximum 5 mg increments), not exceeding 25 mg twice daily. Ruxolitinib could be taken by patients according to the following tablet dosage: 5 mg twice daily, 10 mg twice daily, 15mg twice daily, 20 mg twice daily or 25 mg twice daily. These dosage schedules are in line with the licensed indication. The unit costs for each dose (i.e. price per bottle then the calculated price per tablet) are the UK list price for ruxolitinib. The MS stated that data on the proportions of patients receiving different doses was not available. Only 5 mg tablets were dispensed in COMFORT-II and the mean daily dose was 30.3 mg/day. In the base case, the cost per day for ruxolitinib was calculated assuming that treating physicians prescribed tablets in the form of two 15 mg or 20 mg tablets, at a cost of £120 in the model. The cost used by the manufacturer accurately reflects the likely costs of delivering ruxolitinib under the licensed indication.

As currently constructed, the model assumes no drug wastage. This assumption may not accurately reflect drug usage in practice. The ERG has some concern about drug wastage considering that the shelf-life of the drug is only 30 days.[9] Given that most AEs are managed by dose reduction or interruption it is possible that drugs would expire before all were used, leading to additional costs. There is no evidence to support what sort of impact drug expiry might have on overall costs.

Table 5.12 Ruxolitinib costs (MS, Table B25, pg. 169)

Dose required/ twice daily	Tablet prescribed	Number of tablets/day	Total dose/day	Cost/day
5 mg	5 mg	2	10mg	£60
10 mg	5 mg	4	20 mg	£120
15 mg	15 mg	2	30 mg	£120
20 mg	20 mg	2	40 mg	£120
25 mg	20 mg and 5 mg	2*2	50 mg	£180

BAT treatment costs

The 12-week BAT costs were based on the proportion of patients receiving different types of treatment in COMFORT-II based on data collected at baseline, week 12, week 24, week 36 and week 48. The average proportion of patients receiving each medication or no treatment from week 0 to week 48 was then calculated (Table 5.13, from MS Table B26, pg. 170) and then divided by the number of patients. In the MS it was assumed that patients received treatment for the full 12 weeks on the basis of no available data from COMFORT-II on intensity and duration of BAT treatment. It is unclear to the ERG if this assumption is realistic. If it is not, then the implication is that resource use and hence the costs of BAT treatment could be overestimated in the model. The doses applied to BAT treatments in the model were based on those recommended in the BNF with the exception of lenalidomide (administered at 10 mg rather than 25 mg daily), thalidomide (50 mg rather than 250 mg daily) and interferon alfa 2a (135 micrograms weekly) which were based on clinical advice. The 28-day BAT costs were calculated and then converted to 12-week costs and assigned in the model. All unit costs for BAT were based on the BNF. The MS noted that the use of certain BAT treatments may be higher in practice than observed in COMFORT-II based on HMRN data and a survey by Qureshi et al.[15, 29] which showed that more patients were using medication before the trial than during; suggesting that on enrollment into the trial some patients stopped taking BAT. The manufacturer presented a sensitivity analysis that compared ruxolitinib treatment to two therapy scenarios: hydroxycarbamide (HU), and a combination of thalidomide and lenalidomide.

The manufacturer states in the MS,that the BSCH guidelines[3] reinforce that all current treatments are only temporarily effective. In light of this, the ERG considers it inappropriate that non-responders are considered to take active BAT until death. This potentially grossly overestimates the cost of BAT treatment. In addition, the inclusion of lenalidomide was considered inappropriate by the ERG clinical expert as the treatment is not routinely prescribed in the UK.[29] If the cost of lenalidomide is omitted from the BAT, the overall cost of BAT treatment falls from £702.03 to £402.03; if lenalidomide is replaced with hydroxycarbamide the cost falls to £402.65. These alternative assumptions are briefly explored in Section 6.

Table 5.13 BAT therapy costs (MS, Table B26, pg. 170)

Treatment	Proportion of patients (%)	28-day cost
Hydroxyurea	50.4%	£10.26
Anagrelide	6.0%	£377.60
Prednisone	9.9%	£119.62
Methylprednisolone	2.8%	£34.34
Prednisolone	0.8%	£9.69
Epoetin alfa	8.3%	£641.43
Thalidomide	3.2%	£298.48
Lenalidomide	2.0%	£5,040.00
Mercaptopurine	3.6%	£88.36
Thioguanine	2.0%	£411.94
Danazol	5.6%	£85.80
Peginterferon alfa 2a (Pegasys)	1.2%	£431.04
Melphalan	2.4%	£7.70
Acetylsalicylic acid	2.8%	£0.47
Cytarabine	0.8%	£11.03
Colchicine	0.8%	£4.39
Deferasirox (Exjade)	0.4%	£1,411.20
Folic acid	1.2%	£0.13
Lysine acetylsalicylate	0.8%	£1.65
Interferon alfa 2a (Roferon-A)	1.2%	£283.76
Average 28-day pharmacy cost for best available therapy		**£234.01**
Average 12-week pharmacy cost for best available therapy		**£702.03**

Drug initiation and switching costs

The MS assumed that there was no additional resource use (e.g. consultations, additional tests) associated with administering treatment, monitoring or switching (BAT patients) to ruxolitinib. Both treatment options had the same level of monitoring. Given that the drug is given in tablet form and patients may actually experience lower rates of outpatient visits while on ruxolitinib, the ERG considers this assumption conservative.

Cost of managing MF responders and non-responders

Patients with intermediate-2 and high-risk MF in the responder and non-responder states in the model were assigned a 12-week cost for ongoing management of MF consisting of GP visits, outpatient attendances, and hospital inpatient length of stay. Resource use was based on the recent HMRN audit for outpatient attendances and hospitalisations.[29] No data on GP visits was available from the audit. Therefore, the MS assumed that non-responders were likely to be symptomatic and so were assumed to visit their GP every 2 weeks (six visits every 12 weeks). The MS stated that no corresponding resource use data was available for responders and they therefore assumed a cost for intermediate-2 and high-risk patients of 30% of the costs for non-responders. The ERG clinical expert did not find these assumptions unreasonable.

The manufacturer stated that this assumption regarding GP visits may be an underestimate (if complications of splenomegaly and bothersome symptoms of MF are alleviated by ruxolitinib). As this assumption was not evidence based, the ERG will undertake some sensitivity analysis to ascertain the impact of varying the ratio (Section 6). The unit costs for GP visits were from the PSSRU.[44] Outpatient attendances used NHS reference costs.[45] These were appropriate sources from which to derive these costs.

Cost of blood transfusions

The cost associated with blood transfusions was included as a sensitivity analysis based on the proportion of patients in COMFORT-II defined as transfusion dependent. The MS assumed that 50% to 60% of patients needed three blood transfusions of 2 units every 12 weeks receiving on average 13 to 15.16 units. However the MS noted that based on HMRN data the average number of units received by intermediate-2 and high-risk MF patients was higher than the number used in the model, with a median of 18.95 and 12.75 units respectively per person per year. The HMRN data was skewed in intermediate-2 patients by one patient who was hospitalised for 181 days. The unit costs for blood transfusions were taken from NHS reference costs and NHS Blood Transplant.[45-46] The ERG felt that this was an appropriate estimation of blood transfusion costs.

Cost of managing the main complications of splenomegaly

The costs associated with managing the main complications of splenomegaly were included in the model, these were splenectomy, splenic irradiation and 'other complications'; however these were only applied to non-responders. The unit costs assigned to splenectomy and splenic irradiations, consisting of a period of inpatient hospitalisation, were based on NHS reference costs; which was appropriate. Further, it is feasible that these complications which are directly related to splenomegaly may not be experienced by responders. However 'other complications' which include infection and sepsis are likely to be experienced by both non-responders and responders. The MS stated that it was not possible to obtain an accurate estimate for 'other complications' in the model, and they assume a one-off cost of £2,500 per complication, which they later tested in sensitivity analysis across a range of £1,000 to £4,000. The ERG has some concerns relating to the assumption that responders experienced no costs of complications in the model.

Cost of palliative care

A one-off cost of £5274.28 associated with palliative care was assigned to patients before moving to the terminal dead state in the model. The cost of palliative care was based on the end-of-life cost in a hospice from Coyle et al[47] Coyle's figures were updated to 2011 using PSSRU indexes. These costs appear comparable to other cancer end-of-life costs.

Cost of treating adverse events

The MS stated that the most frequent grade 3 and 4 AEs reported in COMFORT-II were included in the base case analysis (see Table 5.9). The choice of the types of AEs included in the model appear to be reasonable but, as mentioned earlier, they may not adequately reflect all types of AEs based on those observed in COMFORT-II. The proportion of AEs requiring hospitalisation and a period of (short stay) inpatient care as well as the average length of stay in hospital were based on the manufacturer's own assumptions. The unit costs were obtained from the NHS reference costs. The choice of health resource use groups used (for length of stay and cost per day assumptions) and hospitalisation rates for each AE event appear to be plausible but the ERG were unable to fully validate these assumptions.

Cost of leukaemic transformation

The average cost for LT has been estimated from US cost data from Katz et al.[48]and converted into UK pounds but the currency conversion exchange rate used was not presented. As this originates in a US setting with different resource costs and organisational contexts, it is unlikely to be generalisable to the UK setting.

Price adjustment

There was no specific price year stated for all costs. It is also not explicitly mentioned if reflation was required for any of the unit costs applied in the model. However, from choice of reference cost data (NHS reference costs 2010/11 (reference 109 in MS)), haematological costs (NHS Blood and Transplant 2010/11 (reference 107 in MS)), and the PSSRU (reference 104 in MS)), reflation (Curtis 2011) applied to Coyle's[47] palliative care costs, it can be surmised that the price year is 2010/11.

Table 5.14 Adverse event costs (MS, Table B29, pp. 177-8)

Adverse Event	Cost per day	Hospitalisation rate	Length of stay (days)	Cost	Source
Neutropenia	£416.79	100%	7.0	£2,917.56	Average (weighted by activity levels) of: WA02W: disorders of immunity without HIV/AIDS without CC WA02Y: disorders of immunity without HIV/AIDS without CC National Schedule of Reference Costs, 2010 NHS Trusts and PCTs combined Non-Elective Inpatient (Short Stay) HRG Data
Fatigue	£425.57	50%	1.5	£319.18	Average (weighted by activity levels) of: WA18V: Admission for unexplained symptoms with Major CC WA18X: Admission for unexplained symptoms with Intermediate CC WA18Y: Admission for unexplained symptoms without CC National Schedule of Reference Costs, 2010 NHS Trusts and PCTs combined Non-Elective Inpatient (Short Stay) HRG Data
Nausea/ vomiting	£407.00	40%	4.0	£651.20	FZ43C: Non-Malignant Stomach or Duodenum Disorders with length of stay 1 day or less National Schedule of Reference Costs, 2010 NHS Trusts and PCTs combined Non-Elective Inpatient (Short Stay) HRG Data
Diarrhoea	£437.00	40%	4.0	£699.20	FZ36F: Intestinal Infectious Disorders with length of stay 1 day or less National Schedule of Reference Costs, 2010 NHS Trusts and PCTs combined Non-Elective Inpatient (Short Stay) HRG Data
Anaemia	£490.35	40%	6.5	£1,234.02	Average (weighted by activity levels) of: SA04D: Iron Deficiency Anaemia with CC SA04F: Iron Deficiency Anaemia without CC National Schedule of Reference Costs, 2010 NHS Trusts and PCTs combined Non-Elective Inpatient (Short Stay) HRG Data
Thrombocytopenia	£474.62	50%	6.0	£1,471.05	Average (weighted by activity levels) of: SA12D: Thrombocytopenia with CC

Adverse Event	Cost per day	Hospitalisation rate	Length of stay (days)	Cost	Source
					SA12F: Thrombocytopenia without CC National Schedule of Reference Costs, 2010 NHS Trusts and PCTs combined Non-Elective Inpatient (Short Stay) HRG Data
Pneumonia	£456.65	100%	5.0	£2,283.23	Average (weighted by activity levels) of: DZ11A: Lobar, Atypical or Viral Pneumonia with Major CC DZ11B: Lobar, Atypical or Viral Pneumonia with CC DZ11C: Lobar, Atypical or Viral Pneumonia without CC National Schedule of Reference Costs, 2010 NHS Trusts and PCTs combined Non-Elective Inpatient (Short Stay) HRG Data
Asthenia	£420.01	50%	1.5	£315.01	Average (weighted by activity levels) of: AA31A: Headache, Migraine or Cerebrospinal Fluid Leak with CC AA31B: Headache, Migraine or Cerebrospinal Fluid Leak without CC DZ38Z: Oxygen Assessment and Monitoring National Schedule of Reference Costs, 2010 NHS Trusts and PCTs combined Non-Elective Inpatient (Short Stay) HRG Data
Stomatitis/ gastroenteritis	£463.82	40%	4.0	£742.12	Average (weighted by activity levels) of: CZ23W: Major Head, Neck and Ear Disorders 19 years and over with Major CC CZ23X: Major Head, Neck and Ear Disorders 19 years and over with Intermediate CC CZ23Y: Major Head, Neck and Ear Disorders 19 years and over without CC National Schedule of Reference Costs, 2010 NHS Trusts and PCTs combined Non-Elective Inpatient (Short Stay) HRG Data

AE, adverse event; CC, chief complaint; HIV, human immunodeficiency virus; HRG, healthcare resource group

5.2.8 Cost effectiveness base case results

The base case cost-effectiveness analysis presented in the MS compared ruxolitinib with BAT for a population of intermediate-2 and high-risk MF patients.

The total incremental QALYs associated with ruxolitinib patients compared to BAT patients was 1.15 largely due to the gain in QALYs generated by responders (1.1452 QALYs) versus non-responders (0.303 QALYs). The total incremental costs with ruxolitinib compared to BAT were £85,027. These additional costs were driven by differences in drug costs (102.65%) rather than differences due to any other cost components (less than 3% between ruxolitinib and BAT groups).

A full incremental cost-effectiveness analysis was conducted for the total QALYs and costs accruing over the 35-year time horizon of the model. The results are presented in Table 5.15. In the MS base case ruxolitinib had an ICER of £73,980 per QALY when compared with BAT.

Table 5.15 MS base case results

Technologies	Total costs (£)	Total LY	Total QALYs	Incremental costs (£)	Incremental LY	Incremental QALYs	ICER (£) (QALYs)
BAT	£51,908	3.99	1.67				
Ruxolitinib	£136,935	5.03	2.82	£85,027	1.04	1.15	£73,980

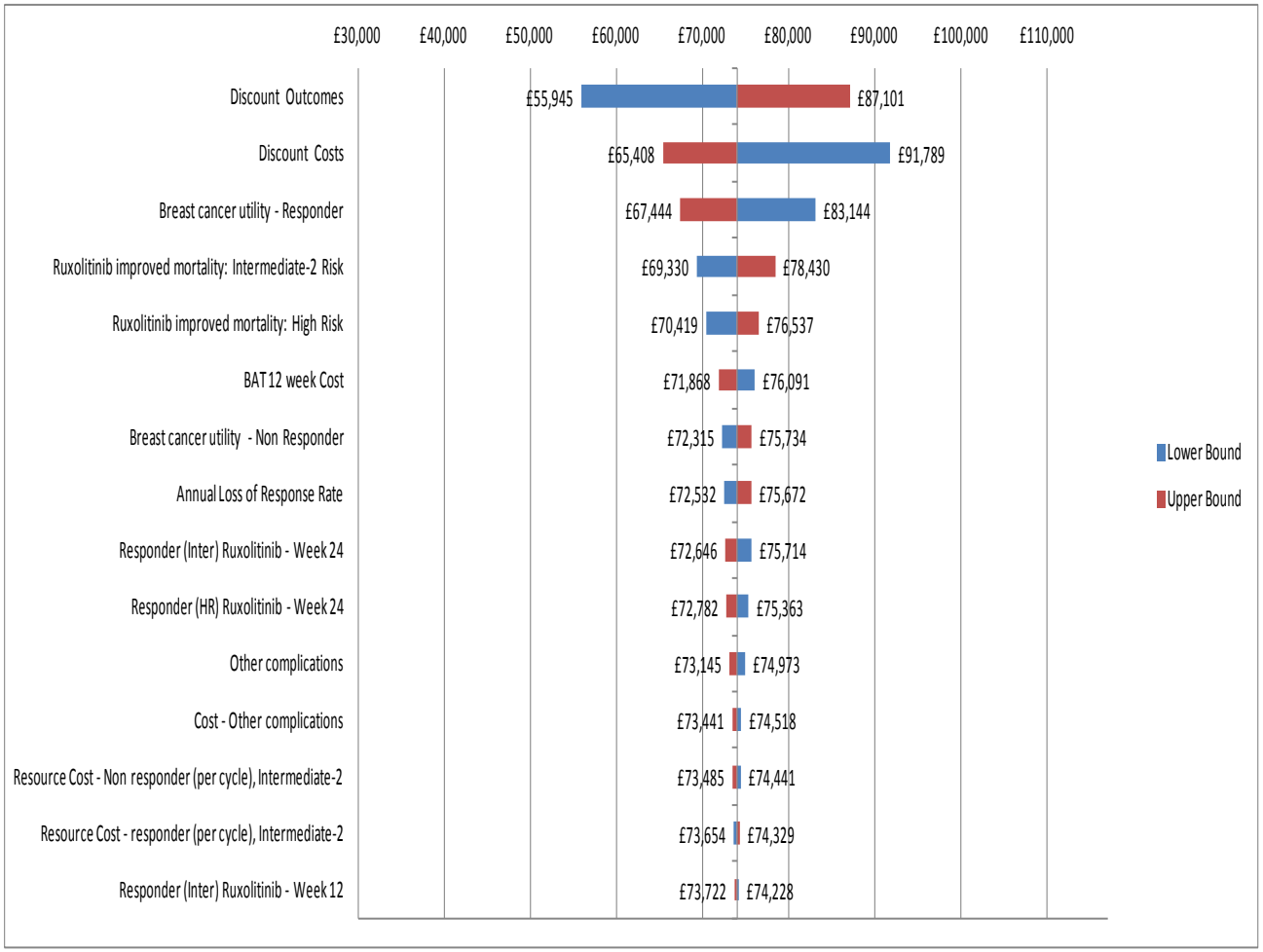
5.2.9 Sensitivity analyses

To evaluate uncertainty the manufacturer undertook one-way, scenario deterministic analysis, and probabilistic sensitivity analysis. One way analysis allowed the testing of extreme values and diagnosing the drivers of the model results; scenario analysis allowed the varying of structural assumptions in the model; and the probabilistic sensitivity analysis allowed the measurement of the effect of parameter uncertainty on the results of the model by varying all parameters simultaneously for a large number of repetitions and then summarising the results of the simulations.

The manufacturer conducted a suite of one-way sensitivity analysis around parameters in the model. For a list of the parameters included in the one way sensitivity analysis and their deterministic values and lower and upper bounds, see MS Section 7.6.2 (Table B30, pg. 181). The results of these sensitivity analyses were presented as a tornado diagram which is presented here, see Figure 5.3.

The one-way sensitivity analysis is described as varying parameters within the 95% CI or a reasonable range, defined in the model submission as +/- 20% of the mean.

Figure 5.3 Results of one-way sensitivity analysis



The ERG felt that it was unclear if the use of +/-20% was sufficient variance to represent uncertainty surrounding a number of of the parameters outlined in Figure 5.3. Some of the parameters varied had very little impact, but given how the original estimates were derived (i.e. based on assumptions by the manufacturer) a wider variance may have been warranted. As might be expected, the impact of increasing discount rates to 6% has the greatest effect, but there is no justification that this higher rate would be appropriate. As highlighted earlier in the report, the ERG feel that there is a high level of uncertainty surrounding survival and utilities. The results of the one-way sensitivity analysis show that varying these parameters – even when changed whilst holding all other parameters constant, has a significant impact on the ICER results. Further, the cost of BAT which has also been highlighted as being uncertain by the ERG, is also demonstrated to be an influential parameter.

Table 5.16 Impact on the ICER of deterministic sensitivity analyses conducted by the manufacturer (See Table 5.1 for signposts to MS)

	ICER Rux versus BAT		**Range of values used in sensitivity analysis**	
	Lower Bound	**Upper Bound**	**Lower Bound**	**Upper Bound**
Base case	£73,980	—	—	—
One-way sensitivity analyses				
Breast cancer utility values for responders	£83,144	£67,444	0.751	0.886
Ruxolitinib improved mortality hazard ratio for intermediate-2 risk MF patients	£69,330	£78,430	0.1	0.5
Ruxolitinib improved mortality hazard ratio for high-risk MF patients	£70,419	£76,537	0.1	0.5
Cost of BAT therapy	£76,091	£71,868	£562	£842
Scenario analyses				
CML utilities	£77,092	—	—	—
NHL utilities	£75,123	—	—	—
Transfusion dependence	£75,887	—	—	—
Leukaemic transformation	£79,184	—	—	—

Scenario analyses were also conducted for the following:

- utility values for CML and NHL were used instead of metastatic breast cancer utilities;
- LT was added as an additional state in the model;
- transfusion dependence was incorporated into the model;
- ‘subgroup analyses’ were presented for different BAT treatment regimens;
- alternative stopping rules (12 weeks) and response criteria (25% spleen volume reduction) were investigated for initial treatment

The manufacturer presents results using CML utilities and NHL utilities instead of utilities for metastatic breast cancer.

- When CML utilities were used the ICER increased to £77,092 from the base case ICER of £73,980 (MS, Table B37, pg. 195).

- When NHL utilities were substituted, the ICER increased to £75,123 (MS,Section 10.14.2, Table D25, pg. 286).

These results reinforce the ERG's belief that the most important factor in utilities is the difference between the responder and non-responder states. Metastatic breast cancer utilities have the highest difference between responders and non-responders; as a consequence they produce the lowest ICER for ruxolitinib, while the opposite is true for CML. A presentation done by Roskell and colleagues[2] using COMFORT-II data and mapping EORTC-QLQ30 scores to EQ5D found an even smaller difference between responder and non-responder utilities than the difference between the upper and lower utilities in CML. Even with alternative assumptions, there is still considerable doubt around the appropriateness of the utilities used within the model.

When transfusion dependence was added to the model, the ICER for ruxolitinib increased to £75,887 (MS, Section 7.7.7, Table B39, pg. 196). When an additional state for LT was added the ICER increased to £79,184 (MS, Section 7.7.7, Table B38, pg. 196). The manufacturer states that approximately 20% of MF patients die due to LT, therefore, it is appropriate for the costs entailed in LT to be included in the evaluation of cost-effectiveness. It was also shown that patients taking ruxolitinib were more likely to become transfusion dependent than BAT patients (COMFORT-II CSR, pp. 124-5).

The manufacturer presents an analysis comparing ruxolitinib versus hydroxycarbamide, and an analysis comparing ruxolitinib versus a combination of thalidomide and lenalidomide. The ERG considers the results of both these analyses irrelevant, as, while hydroxycarbamide is the most often used treatment in MF, it is by no means the only one, as shown by the many other treatments included in the base case BAT and in the HMRN audit.[29] A comparison against all of the BAT options simultaneously would be required to allow an appropriate full incremental analysis. The scenario in which lenalidomide is evaluated is considered by the ERG to be inappropriate because the ERG clinical expert and the HMRN audit[29] indicate that lenalidomide is very rarely used to treat MF patients in the UK (see Section 5.3.2). Beyond this, the total number of patients who received lenalidomide at any point during the COMFORT-II trial is three.[22]

The sensitivity analysis, presented as an alternative base case in the MS, involving the changing of the response criteria to 25% reduction in spleen volume and changing the stopping rule to 12 weeks resulted in an ICER of £66,453 (MS, Section 10.14.1, Table D21, pg. 283). This assumption was justified by the manufacturer on the basis that the majority of patients in COMFORT-II had demonstrated a response by 12 weeks and clinicians consulted by the manufacturer reported that most patients would achieve a clinically meaningful response before 12 weeks. Given that the IWG-MRT definition of clinical improvement does not contain any reference to a spleen reduction commensurate to the 25% reduction in spleen volume rule,[17] it is not clear to the ERG that this analysis is clinically meaningful.

The probabilistic sensitivity analysis found that, at a willingness-to-pay for a QALY of £30,000, ruxolitinib had a 0% chance of being cost-effective. Even at a threshold of £50,000/QALY, approaching which some treatments have been recommended by NICE (utilising the End of Life consideration), there is still a 0% probability of cost-effectiveness. Even under the second more favourable base case analysis, ruxolitinib still has a 0% probability of cost-effectiveness at a threshold of £50,000/QALY. Figures 5.4 and 5.5 show the CEACs for ruxolitinib versus BAT.

Figure 5.4 CEAC for ruxolitinib versus BAT base case scenario (MS Figure 30, pg. 198)

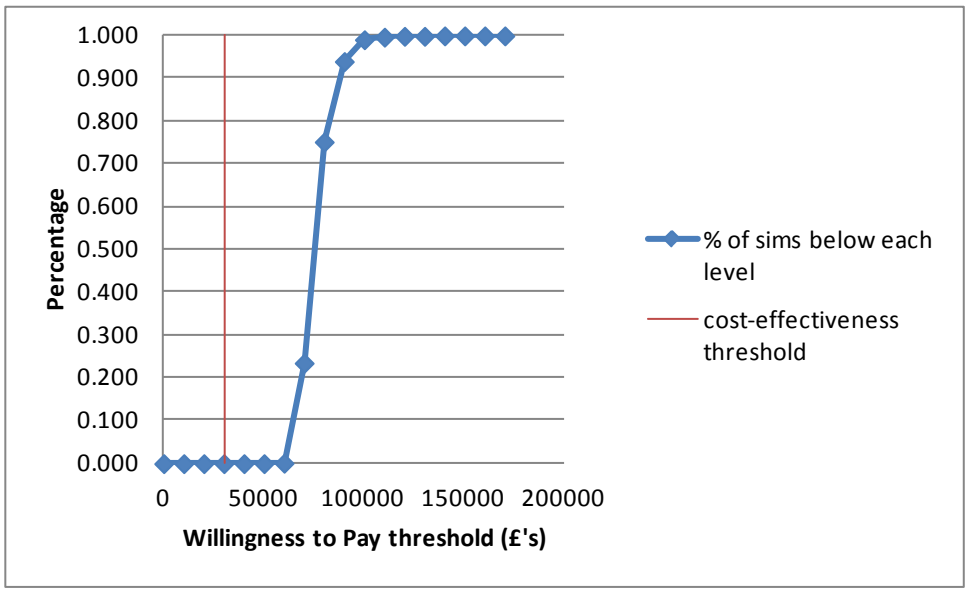
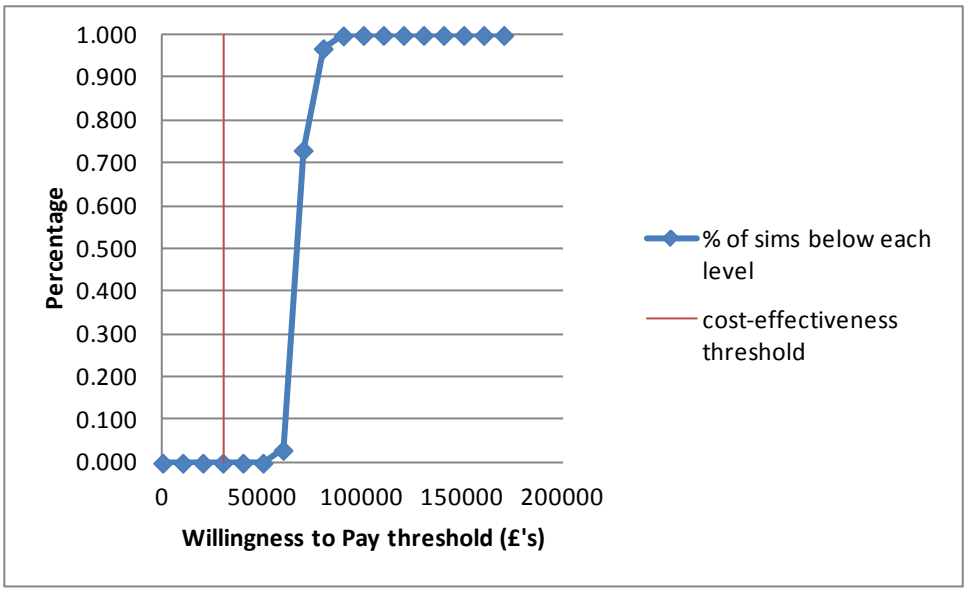


Figure 5.5 CEAC for ruxolitinib versus BAT - 25% threshold and 12 week initial treatment scenario (MS Figure 44, pg. 288)



5.2.10 Model validation and face validity check

The MS included a Markov trace illustrating the results of fifty 12 week cycles of the model with regard to proportion of patients in the four model health states described in Section 5.2.4 (MS Table B32, pg. 185) and the number of QALYs (MS Table B33, pg.191) for ruxolitinib and BAT treated patients. The Markov trace was sufficient to indicate that the model was working as expected.

The submitted economic model has been checked by the ERG for internal consistency, and external consistency.

Internal consistency

The MS states that the model has been quality assured having been reviewed by a person other than the model developer and that parameters were varied within extreme values beyond what would be considered "reasonable" to ascertain whether the change in the simulated costs and utilities was consistent with a priori expectation. However, no further details of the review process are provided.

The ERG undertook a review of the manufacturer's base case and sensitivity analysis. Parameter inputs have been checked for corrective predictive validity (i.e. independent sensitivity analyses have been undertaken and results were mostly consistent with those expected, e.g. increasing the response rate of ruxolitinib increases its effectiveness). However, a number of errors/potential inconsistencies were found:

- The utility decrement applied to LT for responders appeared to incorrectly refer back to empty cells in the model cycle calculations worksheets, specifically for worksheets:

- "INCB018424 Calcs - Inter Risk" and "INCB018424 Calcs - High Risk", For the column stating in cell AX49 (LT) =AE49*($AE$8+$Z$18)+AF49*($AD$8+$Z$18)+(AG49-AF49+AH49-AE49)*($AD$13+$Z$18)

- => $AD$13 should be $AE$8

- "Comparator Calcs - Inter Risk", and Comparator Calcs – High Risk" (LT) column starting in cell AS49

- =AB49*($AB$8+$W$18)+AC49*($AA$8+$W$18)+(AD49-AC49+AE49-AB49)*($AA$13+$W$18)

- => $AA$13 should be $AB$8

Sensitivity analysis results in Section 6 have been corrected for this error.

- Running the model with equivalent response rates for ruxolitinib and BAT, equal HRs for survival, no utility decrement for any health states (though unrealistic), no treatment discontinuation, and no complications should produce the same results. Instead, slightly different results between the treatments occur, as the "months as a responder" cells show below:

Table 5.17 Demonstration of unidentified error in the manufacturer model

Per-patient outcomes: lifetime	Ruxolitinib	BAT	difference
Months as a Responder	16.90	18.71	-1.80
Number of Splenectomies	0.00	0.00	0.00
Number of Splenic Irradiation	0.00	0.00	0.00
Number of Other Complications	0.00	0.00	0.00
Overall Survival (months)	47.85	47.85	0.00
Leukemia-free Survival (months)	47.85	47.85	0.00
Quality-Adjusted Life-Years	3.99	3.99	0.00

External consistency

The manufacturer stated that it was not possible to validate their model results against any external data as it was not possible to compare directly the results of the model against the clinical trials or other sources as there are no other models developed in the same disease area. Given the lack of similar studies the ERG were also unable to externally validate the manufacturer's model.

6 IMPACT ON THE ICER OF ADDITIONAL CLINICAL AND ECONOMIC ANALYSES UNDERTAKEN BY THE ERG

The ERG has undertaken additional analyses focusing on the issues and uncertainties highlighted in Section 5. It is the opinion of the ERG that the model presented in the MS is an over simplification of MF and does not appropriately account for disease progression over time. The ERG acknowledges that the lack of long-term data on many key parameters has driven these simplifications and has undertaken its analysis with a view to highlighting the implications of some of these assumptions.

The high level of parameter uncertainty and disease complexity makes the derivation of an alternative base case scenario difficult. However, the ERG felt that there was some uncertainty surrounding the exclusion of LT from the base case model. The ERG feel that the exclusion may not be fully justified, as over 20% of intermediate-2 or high-risk MF patients make this transition.[6] Also assuming different rates of LT for responders (3.6%) and non-responders (3.8%) in the sensitivity analysis presented in the MS may not be warranted. The ERG clinical expert indicated that the conservative assumption is to assume equivalent rates of LT. To reflect the uncertainty and allow some comparisons to be made, the ERG have presented all further analyses results for varying parameter/assumption in two modelling scenarios: scenario one - the manufacturer's base case; and scenario two – the manufacturer's base case including LT, with equivalent rates for responders and non-responders.

Detailed below are the assumptions and parameters explored by the ERG under these two modelling scenarios:

- survival data
    - distributional assumptions for base case survival data
    - HR for mortality
- treatment response
    - response rates
    - definition of responder
- discontinuation rates
- utility data
- resource use and costing
    - GP visits
    - lenalidomide
    - responders ratio

These analyses are undertaken with the view of highlighting the high level of uncertainty in the data and assumptions made by the manufacturer. The results should be considered with the major structural limitations of the model in mind. The ERG also presents a final scenario which incorporates an alternative set of assumptions and parameters. This scenario is as plausible as the scenario presented as the manufacturer's base case, but it is also as uncertain. The scenario assumptions include:

- reduced the time-horizon of the model to 15 years;
- allowed equivalent rates (3.6%) of LT between ruxolitinib and BAT;
- allowed transfusion dependence;
- mapped utilities from Roskell et al.[2]in a MF population;
- and used the survival HR from long term COMFORT-II data (manufacturer's response document).

6.1 Overall survival

Weibull regression for survival data

In the MS, the manufacturer derived the non-responder 12-week survival probabilities from the median survival in intermediate-2 and high-risk patients from the Cervantes et al.[7] study assuming an exponential distribution. The impact of other distributional assumptions was not explored. The ERG undertook an analysis assuming a Weibull distribution (shape parameter 0.9) to estimate 12-week survival probabilities which generated estimates for non-responders and subsequently can be combined with the HR to derive a probability for responders. The results of this analysis are presented in Table 6.3.

Table 6.2 Mortality probabilities for exponential and Weibull distributions based on Cervantes et al. [7]

Regression used	Intermediate-2 risk		High-risk	
	Non-responder	Responder	Non-responder	Responder
Exponential (base case)	3.92%	1.21%	6.86%	2.13%
Weibull	5.18%	1.55%	8.54%	2.56%

Alternative hazard ratios (COMFORT-I and II data)

In addition to exploring alternative distributional assumptions the ERG have highlighted several limitations with the use of the Phase I/II trial[30]survival data to estimate the HR. These issues have been discussed in Sections 4 and 5. In the ERG's points for clarification letter, the ERG requested long-term survival data for a sub-group of this trial: the intermediate-2 and high-risk groups in order to have a HR from a population representative of the one being modelled. The manufacturer was unable to provide data from the phase I/II trial because the data are owned by the MDACC trial centre.

In addition, the ERG requested and the manufacturer provided updated long term survival data from the COMFORT-I and II trials. These data were provided in some detail, along with new estimates of HR values derived assuming an exponential distribution. However, these data were not presented in a format that would allow the ERG to test alternative distributional assumptions, but the graph provided for COMFORT-I suggests that in this instance an exponential assumption may have been reasonable. The HRs have been calculated assuming that there is no difference between intermediate-2 and high-risk patients, which is in line with the manufacturer's original assumptions.

The ERG have undertaken analyses using both the base case survival data using an alternative Weibull distribution and using the new survival data from COMFORT-I and II, provided in response to the ERG's clarification letter. These analyses are presented in Table 6.3 and 6.4.

Table 6.3 Incremental cost-effectiveness ratios results using Weibull regression for survival data

Model scenario	ICER (£) base case results (exponential)	ICER (£) alternative distribution (Weibull)
MS base case (no LT)	**73,980**	74,274
ERG alternative (MS base case plus equal LT)	78,642	79,303

Figures in bold are the results of the manufacturer's base case

Table 6.4 Incremental cost-effectiveness ratios results using alternative hazard rations (COMFORT-1 and II data)

Model scenario	Analysis	ICER(£)
MS base case (no LT)	MS base case	**73,980**
	COMFORT-I	86,086
	COMFORT-II	83,129
ERG alternative (MS base case plus equal LT)	Alternative ERG	78,642
	COMFORT-I	90,557
	COMFORT-II	88,278

Figures in bold are the results of the manufacturer's base case

The ERG has presented alternative analyses changing the distributional assumption used to evaluate the base case survival data to a Weibull regression, and undertaking a re-analysis using the HRs from the longer-term survival data in COMFORT-I (0.58 HR) and COMFORT-II (0.52 HR) trials (manufacturer's response document). These analyses highlight uncertainty in the survival data, the limitations of which are discussed in Section 4. All of the analyses undertaken increased ruxolitinib's ICER (range: £74,274 to £90,557). Combined with the lack of disease progression in the model, these additional analyses suggest that uncertainty inherent in the survival data make any ICER results uncertain, with the results presented in the MS probably representing the best case scenario.

6.3 Treatment response

Response rates

As reported in Section 4, trial analyses were conducted using an ITT approach; the assumption was made for COMFORT-II that if data were not available for any reason at week 48 those missing patients would be assumed non-responders. The data from which the base case percentages were calculated were not presented in the MS, but were provided in an associated reference.[49] The ERG have used these data to estimate percentages assuming that any patient with missing data is a non-responder, in line with what has been done in the clinical trial analysis (see Table 6.5). The results of this analysis are presented in Table 6.6

Table 6.5 Percentage of patients with ≥ 35% reduction in spleen volume, assuming missing patients are non-responders COMFORT-II trial, IPSS classification

Treatment up to week	High-risk ruxolitinib	High-risk BAT	Intermediate-2-risk ruxolitinib	Intermediate-2-risk BAT
0	[redacted]	[redacted]	[redacted]	[redacted]
12	[redacted]	[redacted]	[redacted]	[redacted]
24	[redacted]	[redacted]	[redacted]	[redacted]

Table 6.6 Incremental cost-effectiveness ratios with missing patients as non-responders COMFORT-II trial, IPSS classification

Model scenario	Response rate	ICER (£) (QALYs)
MS base case (no LT)	COMFORT-II trial	**73,980**
	assuming missing patients are non-responders COMFORT-II trial, IPSS classification	75,571
ERG alternative (MS base case plus equal LT)	COMFORT-II trial	78,642
	assuming missing patients are non-responders COMFORT-II trial, IPSS classification	80,677

Figures in bold are the results of the manufacturer's base case

The ICERs using this alternative approach for response increased the ICERs slightly.

Definition of response

Two different definitions of response are explored:

- a strict adherence to IWG-MRT guidelines on splenomegaly related clinical improvement [17]
- a definition that defines response rates by 50% reduction in Total Symptom Score (TSS) and Patient's Global Impression of Change (PGCI)

IWG-MRT Definition of splenomegaly related clinical response

The measure used to define response in the MS model was at least a 35% reduction in spleen volume as measured by MRI. There are important distinctions between the MS definition of response and the IWG-MRT definition: the IWG-MRT defines response by palpable spleen length, which is less precise than spleen volume measurement, and patients with a palpable spleen length under 10 cm must have a 100% reduction in palpable spleen length not the 50% reduction (35% by volume) presented by the manufacturer.

In COMFORT-II, 32.3% of patients had a palpable spleen length of under 10 cm at baseline,[9] additionally, no patients achieved a 100% spleen volume reduction in the COMFORT-II trial (waterfall plots in COMFORT-CSR, Figure 11-1, pg. 117; and MS, Figure 10, pg. 82), therefore, omitting the response of 32.3% of responders is justified. Additionally, in COMFORT-I, it was demonstrated that MRIs detected a 41.9% response rate at 24 weeks, while palpation detected a 39% response rate, which equates to a risk ratio of 0.93 in favour of MRI. If we apply the risk ratio of palpation and MRI splenomegaly response detection, then response rates are lowered. The ERG assumes that the IWG-MRT response criteria affect intermediate-2 and high-risk patients equally. The results of these analyses are presented in Table 6.6.

Table 6.6 Incremental cost-effectiveness ratios with IWG-MRT definition of splenomegaly related clinical response

Model scenario	Patient group	ICER (£)	Source
MS base case (no LT)	Base case	**73,980**	MS
	IWG-MRT	81,110	COMFORT-II, EMA
ERG alternative (MS base case plus equal LT)	ERG alternative	78,642	MS
	IWG-MRT	82,525	COMFORT-II, EMA

Figures in bold are the results of the manufacturer's base case

Response defined by symptom reduction

The ERG clinical expert indicated that in addition to spleen reduction an important response for patients was reduction in symptoms, which may occur in conjunction with or without a 35% reduction in spleen volume. The ERG analysis presented here uses a definition of response, which is based on

improvement in symptoms and clinical improvement rthaer than a reduction in spleen volume: a 50% reduction in TSS, combineded with the percentage of patients who viewed their condition as much improved or very much improved using PGIC at 24 weeks from the COMFORT-I trial.[21]

The data available on symptom improvement in COMFORT- I was not delineated into intermediate-2 and high-risk groups as the model's data are, so equivalence in TSS response has been assumed for intermediate-2 and high-risk patients. This assumption is a simplifying assumption made due to the absence of data.

Response rates are defined by ruxolitinib treated patients and placebo treated patients. It is likely that the placebo response rates underestimate what would be observed with BAT treatment; greater response rates in BAT would increase the ICER for ruxolitinib.

No tabular presentation of data for each time-point used in the model was available, so the ERG has approximated 12 week response rates from the MS (Figure 6.1, reproduced from MS Figure 16, pg.89). The approximate response rate by TSS for ruxolitinib at 12 weeks was 45%. Placebo had an approximate response rate of 9.8% at 12 weeks. These response rates at 12 weeks were modified to approximate the number of responders and non-responders who would rate themselves as much improved or very much improved by using the PGIC figures from Table 6.7 (manufacturer's response document, Table 1, pg. 5). Perception of change at 12 weeks was assumed to be identical to perception of change at 24 weeks; which may be a strong assumption. For response rates from 24 to 36 weeks it was assumed that the number of responders was equivalent to the number of individuals in each treatment group who stated their perception of change as much improved or very much improved.

Figure 6.1 Response by 50% reduction in TSS (MS, Figure 16, pg. 89)

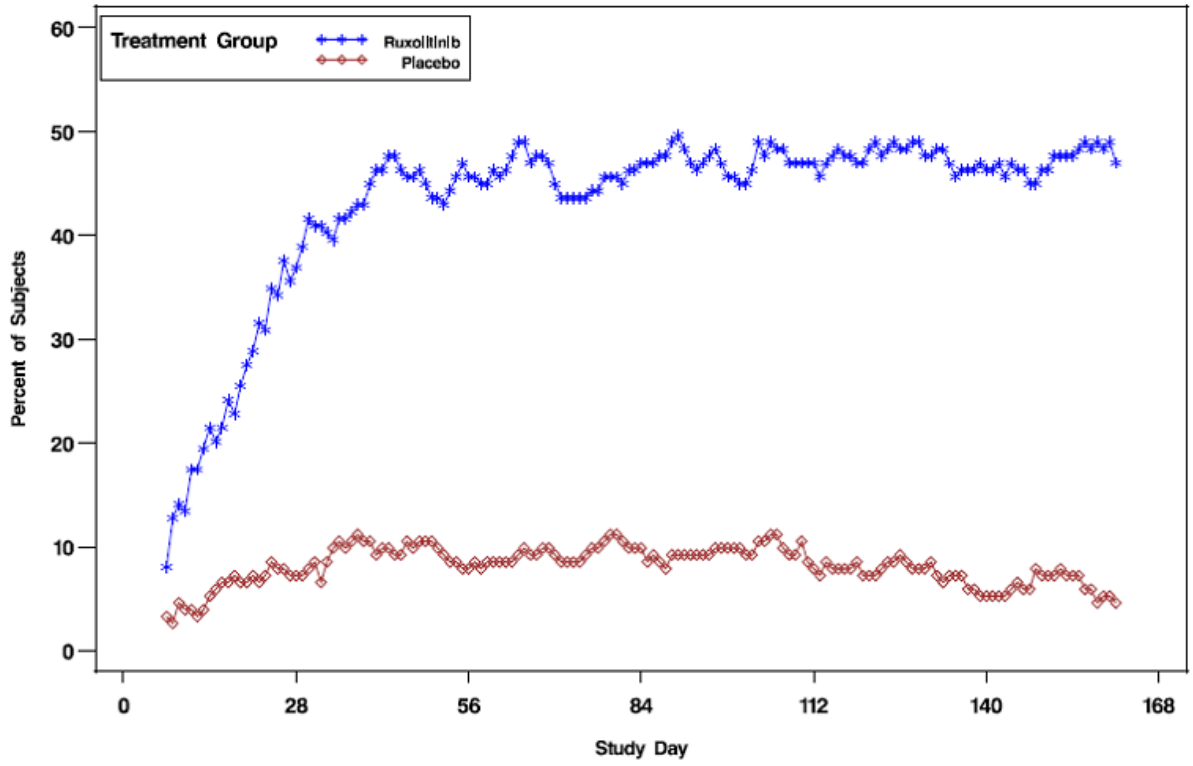


Table 6.7 Patients classified as responders by TSS with PGIC categories (COMFORT-I; manufacturer's response document, Table 1, pg. 5)

	Total Sample (n=227) [1]				Ruxolitinib (n=127) [2]				Placebo (n=100) [3]			
PGIC	Responder[4] (n=77)		Non-Responder[5] (n=150)		Responder[4] (n=68)		Non-Responder[5] (n=59)		Responder[4] (n=9)		Non-Responder[5] (n=91)	
	n	%	n	%	n	%	n	%	n	%	n	%
Very much improved	36	46.8	8	5.3	35	51.5	8	13.6	1	11.1	0	0.0
Much improved	30	39.0	26	17.3	27	39.7	19	32.2	3	33.3	7	7.7
Minimally improved	6	7.8	39	26.0	3	4.4	22	37.3	3	33.3	17	18.7
No change	2	2.6	36	24.0	0	0.0	6	10.2	2	22.2	30	33.0
Minimally worse	2	2.6	22	14.7	2	2.9	3	5.1	0	0.0	19	20.9
Much worse	1	1.3	14	9.3	1	1.5	0	0.0	0	0.0	14	15.4
Very much worse	0	0.0	5	3.3	0	0.0	1	1.7	0	0.0	4	4.4

[1] 227 (of the 237 subjects with paired Baseline Total Symptom and Week 24 Total Symptom Scores) have a valid PGIC response; [2] 127 (of the 133 subjects with paired Baseline Total Symptom and Week 24 Total Symptom Scores) have a valid PGIC response; [3] 100 (of the 104 subjects with paired Baseline Total Symptom and Week 24 Total Symptom Scores) have a valid PGIC response; [4] Number (%) of patients who achieved a

50% or greater improvement in modified MFSAF v2.0 diary Total Symptom Score (Baseline Total Symptom Score – Week 24 Total Symptom Score); [5] Number (%) of patients who did not achieve a 50% or greater reduction in modified MFSAF v2.0 diary Total Symptom Score (Baseline Total Symptom Score – Week 24 Total Symptom Score)

Table 6.8 combines data from Figure 16 (MS, pg. 89)6.1 and Table 6.7 to give response rates based on symptomatic burden.

Table 6.8 Response rate by week from combination of TSS and PGIC

Week	Ruxilitinib (%)	Placebo (%)
0	0.0	0.0
12	63.7	11.3
24	70.1	11.0
36	70.1	11.0
48	70.1	11.0

By inputting the TSS + PGIC data with and without LT the results in Table 6.9 are produced.

Table 6.9 Incremental cost-effectiveness ratios with TSS data with and without leukaemic transformation

Model scenario	Response rate	ICER(£)	Source
MS base case (no LT)	Base case	**73,980**	MS
	TSS + PGIC	79,536	COMFORT-I, manufacturer's response document
ERG alternative (MS base case plus equal LT)	Base case	78,642	MS
	TSS + PGIC	82,282	COMFORT-I, manufacturer's response document

Figures in bold are the results of the manufacturer's base case

The use of the alternative definition of response had a small impact on the ICER, increasing it slightly. The changes in response definition were analysed independently of the other modelling limitations. Due to the structural assumptions employed within the model and the lack of disease progression the impact of varying the response definition was not huge, ICERs increased ranging

from £79,536 to £82,282. This is not to say that the definition of response is not important or not highly uncertain, the results emphasise the limitations in the underlying modelling assumptions. Because one of the main drivers of cost-effectiveness in the model is the cost of ruxolitinib, increasing response increases undiscounted costs more than it increases undiscounted QALYs as most QALYs are gained in the extrapolation to 35 years. The alternative analyses undertaken by the ERG fail to fully characterise the impact of variation in the definition of response due to the limitation of the model structure.

6.4 Treatment discontinuation

In the model long term discontinuation rates for responders after initial treatment were estimated from the phase I/II study[23] and assumed an annual rate of 8.1% (or 12-week probability of 1.93%) and a range of 5.9% to 10.9% in their sensitivity analysis. The ERG considered it more appropriate to base long term discontinuation rates on the results of COMFORT-II (based on the rates published in the NEJM by Harrison et al.[22]) which, at the time of data cut-off (i.e. when the last patient had completed the 48 week visit) was 23.7% (32 of 146 patients had discontinued treatment), which equates to a 12-week probability of approximately 5.5%. The ERG conducted two analyses incorporating these rates for long term discontinuation.

(i) using a 12-week probability of 5.5% for long term discontinuation rate, while holding the initial rate the same (2.3).
(ii) using a 12-week probability of 5.5% for initial treatment phase discontinuation, whilst holding the long term rate the same (1.93)

The results of these analyses increase the ICER modestly as presented in Table 6.10.

Table 6.10 Incremental cost-effectiveness ratios with COMFORT-II discontinuation rates : (i) used over the long term in place of phase I/II study rates;(ii) used for initial discontinuation

Model scenario	Source (initial 12-week discontinuation rate) (long term 12-week discontinuation rate) (%)	ICER (£) (QALYs)
MS base case (no LT)	MS, Verstovsek et al. (initial 2.3) (long-term 1.93)	**73,980**
	(i) COMFORT-II (initial 2.3) (long-term 5.5)	83,680
	(ii) COMFORT-II (initial 5.5) (long-term 1.93)	74,616
ERG alternative (MS base case plus equal LT)	MS, Verstovsek et al. (initial 2.3) (long-term 1.93)	78,642
	(i) COMFORT-II (initial 2.3) (long-term 5.5)	88,622
	(ii) COMFORT-II (initial 5.5) (long-term 1.93)	79,436

Figures in bold are the results of the manufacturer's base case

6.5 Health-related quality of life

As the MS provided no utility scores from a population of MF patients, the ERG has undertaken a sensitivity analysis to examine the robustness of the results to alternative utility values. The ERG tested the impact of using EORTC scores mapped to EQ-5D scores using COMFORT-II data. Metastatic breast cancer utilities for responders (0.823) and non-responders (0.446) were replaced with values from the mapping exercise conducted by RTI, the firm that built the Excel model for the MS,[2] The utility values from the mapping exercise were 0.754 for responders and 0.670 for non-responders. These results are presented in Table 6.11. There is a considerable degree of uncertainty surrounding the health utility values. The ERG used alternative values for responder and non-responder in the model which increased the ICER for ruxolitinib versus BAT (by 31% to 39%).

Table 6.11 Incremental cost-effectiveness ratios with mean utility values from Roskell et al. 2012

Treatment	Total QALYs	Incremental QALYs	ICER (£) (QALYs)	Total QALYs	Incremental QALYs	ICER (£) (QALYs)
	MS base case (no LT)			ERG alternative (MS base case plus equal LT)		
BAT	**1.67**			1.49		
Ruxolitinib	**2.62**	**1.15**	**73,980**	2.41	0.92	78,642
	MS base case (no LT) (mapped utilities)			ERG alternative (MS base case plus equal LT) (mapped utilities)		
BAT	2.51			2.18		
Ruxolitinib	3.38	0.88	97,105	2.84	0.76	109,092

Figures in bold are the results of the manufacturer's base case

6.6 Resource use and costing

The ERG identified several issues and uncertainties with the approach taken to estimate resource use and costs in the MS. More details of these can be found in Section 5.2.7. The ERG has carried out several analyses altering resource cost assumptions.

In the first analysis, the ERG addresses concerns regarding the inclusion of lenalidomide. The ERG analysis substitutes lenalidomide with hydroxycarbamide, a conservative assumption designed to replace lenalidomide with the most commonly prescribed drug in MF treatment. If the cost of lenalidomide is substituted for hydroxycarbamide, the overall cost of BAT treatment falls from £702.03 to £402.65. The resulting ICERs with this substitution are presented in Table 6.12 below.

Table 6.12 Incremental cost-effectiveness ratios with hydroxycarbamide substituted for lenalidomide

Model scenario	BAT costs	ICER (£/ QALYs)
MS base case (no LT)	MS, COMFORT-II	**73,980**
	Hydroxycarbamide substituted for lenalidomide	75,141
ERG alternative (MS base case plus equal LT)	MS, COMFORT-II	78,642
	hydroxycarbamide substituted for lenalidomide	80,086

Figures in bold are the results of the manufacturer's base case

As a second analysis, the ERG has varied the number of GP visits for non-responders. In the base case, non-responders are treated with BAT, assumed to be symptomatic and assumed to visit their GP every 2 weeks (six visits every 12 weeks). Patients achieving a response are assumed, in the absence of evidence, to use 30% of the resources of non-responders, including GP visits, which results in approximately 1.8 visits every 12 weeks for responders. The ERG tested this assumption by increasing the number of GP visits from 6 visits every 12 weeks to 8 and 10 visits every 12 weeks. The resulting ICERs with variation in the number of GP visits are presented in Table 6.13.

Table 6.13 Incremental cost-effectiveness ratios with variation in the number of GP visits

Model scenario	GP visits	ICER (£) (QALYs)
MS base case (no LT)	6 (MS)	**73,980**
	8	73,902
	10	73,824
ERG alternative (MS base case plus equal LT)	6 (MS)	78,642
	8	78,504
	10	78,366

Figures in bold are the results of the manufacturer's base case

A third analysis tested the ratio applied to the other medical costs for responders compared to non-responders around the manufacturer's own assumption of 0.3. Using a value of 0.1 the ICER

decreased slightly and using a value of 0.5 the ICER increased slightly. These results are presented in Table 6.14.

Table 6.14 Incremental cost-effectiveness ratios with variation of the ratio applied to the other medical costs for responders compared to non-responders

Model scenario	Ratio applied to other medical costs for responders	ICER (£) (QALYs)
MS base case (no LT)	0.3 (MS)	73,980
	0.1	71,834
	0.5	76,125
ERG alternative (MS base case plus equal LT)	0.3 (MS)	78,642
	0.1	76,410
	0.5	80,874

Figures in bold are the results of the manufacturer's base case

The results of these sensitivity analyses suggest that any biases resulting from the costing and resource use issues identified by the ERG have minimal impact on the ICER. The ERG tested several alternative estimates for resource use and costs used in the model regarding lenalidomide treatment, GP visits and medical costs for non-responders; these analyses increased the ICER slightly.

6.7 Alternative plausible scenario

As a final sensitivity analysis the ERG undertook an analysis of another scenario using alternative plausible data and assumptions to combine several of the uncertainties used in the model. These changes to the MS base case involved: reducing the model time-horizon to 15 years; including LT with equivalent rates (3.6%) of LT between ruxolitinib and BAT; allowing transfusion dependence; substituting utilities from Roskell et al.[2] for metastatic breast cancer utilities; and using the survival HR from long term COMFORT-II data (manufacturer's response document). It is the opinion of the ERG that these data/assumptions are as plausible as those presented in the MS base case, although the ERG acknowledge that they may be no more certain.

When these combined variations were included in a single analysis the ICER was £148,867. The ERG feel that this analysis highlights the impact of the high level of uncertainty in the evidence base. The ERG feel that if disease progression had been incorporated appropriately into the model structure it is possible that this ICER may increase further, although a better characterisation of the uncertainty would have been possible.

6.8 Conclusions following additional work by the ERG

In addition to the structural limitations which the ERG were not able to correct, there is a high level of uncertainty for many of the key drivers of the model. The impact of varying some of the underlying assumptions and parameter values has been investigated by the ERG. However, due to the lack of robust data, the uncertainty surrounding any ICER estimate is huge. The ERG feel that the ICER presented in the manufacturer's base case is likely to represent a best case scenario, but without a more appropriate model structure, and more robust survival and utility data, no firm conclusions surrounding the value of the ICER can be drawn. What is clear from the analyses undertaken by the manufacturer and the ERG is that the value of the ICER is likely to increase from the result presented in the MS base case.

7 END OF LIFE

Life expectancy of patients with MF is associated with a range of risk factors and risk stratification scores have been developed (see Table 2.1 for a summary). The most recent and clinically relevant score (DIPSS-Plus) assigns high-risk patients a median survival of 1.3 years. However, patients eligible for ruxolitinib in clinical practice are a mix of intermediate-2 and high-risk (reflecting the trials), and intermediate-2 risk patients have a median survival of 2.9 years. In the trials which support the use of ruxolitinib in MF the risk classification reported was the IPSS, from which the mortality rates for high-risk and intermediate-2 risk patients are 2.3 years and 4 years respectively. This would suggest that the life expectancy of patients with MF does not meet the criterion set by NICE to meet the End of Life consideration. Furthermore, the manufacturer's model included a 35 year time horizon and, when the ERG challenged this as being potentially over-long, the manufacturer in their response defended it robustly. Also in the manufacturer's response to the ERG's query regarding the genralisability of the trial populations to the UK, they argued that ruxolitinib may be suitable for intermediate-1 risk patients, who according to DIPSS-Plus have a median survival of 6.5 years.

The RCTs do provide some evidence of a survival benefit with ruxolitinib compared with BAT, with borderline statistical significance; 86% versus 78% (HR 0.52, 95% CI 0.27 to 1.00) using updated survival data at a median of 112 weeks of follow-up. However at this time point only a small number of patients were included in the analysis and the result is uncertain.

The number of patients indicated for ruxolitinib treatment in the UK is small. The prevalence of primary MF has been estimated to be 2.7 per 100,000 population,[10] with MF secondary to PV or ET affecting 0.1 per 100,000 population.[11] The annual incidence of MF is estimated to be 0.34–0.76 per 100,000.[12-14] This equates to approximately 187– 420 individuals diagnosed with MF in England and Wales per year.

8 OVERALL CONCLUSIONS

Evidence from two good quality RCTs demonstrates that ruxolitinib is effective at reducing splenomegaly and its associated symptoms. However, important haematological symptoms of MF (in particular anaemia and thrombocytopenia) are worsened by ruxolitinib in some patients, at least in the short term, requiring dose interruptions and reductions, as well as blood transfusions. There is no evidence of an improvement in progression-free survival with ruxolitinib. There is some evidence that overall survival may be increased with ruxolitinib, although these data are less reliable.

The model presented in the MS does not fully capture disease progression. In addition to the structural issues, some of the underlying modelling assumptions are clinically dubious. The additional analyses undertaken by the ERG showed that the majority of plausible modifications to the model inputs resulted in an increase in the ICER. The alternative scenario presented demonstrates the effect of pooling a number of plausible modifications to undertake an alternative scenario. This scenario more than doubles the ICER presented by the manufacturer. The ERG feel that the lack of disease progression captured in the model and the lack of long term data make obtaining a more robust estimate of the ICER difficult. It is however very likely that the base case ICER presented by the manufacturer represents a best case scenario.

8.1 Implications for research

There are six ongoing trials of ruxolitinib for patients with myeloproliferative neoplasms, as described in Section 4.1.7, all of which are uncontrolled trials. Study NCT01317875 and study NCT01348490 (INCB 18424-258) will assess ruxolitinib in patients with low platelet counts ($< 100 \times 10^9$/L), which will fill an important gap in the evidence base. In addition, study NCT01340651 (INCB 18424-260) will evaluate the effects of a sustained release formulation of ruxolitinib on platelet count. Study NCT01558739 is a phase II open-label study to evaluate the efficacy of ruxolitinib in patients with PMF, PPV-MF or PET-MF and uses a composite endpoint of reduction in splenomegaly and/or reduction in total symptom score. It is anticipated that this UK study will generate local health resource utilisation data, which will enable a more reliable assessment of the cost effectiveness of ruxolitinib in the UK.

There is a general lack of RCTs of treatments for MF. Adequately controlled trials comparing different medical treatments, with long term efficacy and safety data, would help inform recommendations for the management of this disease. The suggestion of possible differential

effects of ruxolitinib by MF subtype may warrant exploration in adequately powered clinical trials.

References

1. National Institute for Health and Clinical Excellence (NICE). *Eribulin for the treatment of advanced metastatic breast cancer (TA250)*. London: National Institute for Health and Clinical Excellence; 2012. Available from: http://publications.nice.org.uk/eribulin-for-the-treatment-of-locally-advanced-or-metastatic-breast-cancer-ta250
2. Roskell NS, Mendelson ET, Whalley D, Knight C. Using a condition-specific measure of patient-reported outcomes to derive utilities in myelofibrosis. *Value Health* 2012;**15**:A224-A5.
3. Reilly JT, McMullin MF, Beer PA, Butt N, Conneally E, Duncombe A, et al. Guideline for the diagnosis and management of myelofibrosis. *Br J Haematol* 2012;**158**:453-71.
4. Scherber R, Dueck AC, Johansson P, Barbui T, Barosi G, Vannucchi AM, et al. The Myeloproliferative Neoplasm Symptom Assessment Form (MPN-SAF): international prospective validation and reliability trial in 402 patients. . *Blood* 2011;**118**:401-8.
5. Stein BL, Moliterno AR. Primary myelofibrosis and the myeloproliferative neoplasms: the role of individual variation. *JAMA* 2010;**303**:2513-8.
6. Passamonti F, Cervantes F, Vannucchi AM, Morra E, Rumi E, Pereira A, et al. A dynamic prognostic model to predict survival in primary myelofibrosis: a study by the IWG-MRT (International Working Group for Myeloproliferative Neoplasms Research and Treatment). *Blood* 2010;**115**:1703-8.
7. Cervantes F, Dupriez B, Pereira A, Passamonti F, Reilly JT, Morra E, et al. New prognostic scoring system for primary myelofibrosis based on a study of the International Working Group for Myelofibrosis Research and Treatment. *Blood* 2009;**113**:2895-901.
8. Gangat N, Caramazza D, Vaidya R, George G, Begna K, Schwager S, et al. DIPSS plus: a refined Dynamic International Prognostic Scoring System for primary myelofibrosis that incorporates prognostic information from karyotype, platelet count, and transfusion status. *J Clin Oncol* 2011;**29**:392-7.
9. Committee for Medicinal Products for Human Use (CHMP). *CHMP assessment report: Jakavi Ruxolitinib. EMA/465846/2012*. European Medicines Agency (EMA); 2012. [cited 2012]. Available from: http://www.emea.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/002464/WC500133226.pdf.
10. Orphanet. *Prevalence of rare diseases: Bibliographic data*. In: Orphanet Report Series; 2011. Available from: http://www.orpha.net/consor/cgi-bin/Education_Home.php
11. European Medicines Agency (EMA), Committee for Orphan Medicinal Products. *Public summary of opinion on orphan designation. (R)-3-(4-(7H-pyrrolo[2,3-d]pyrimidin-4-yl)-1H-pyrazol-1-yl)-3-cyclopentylpropanenitrile phosphate for the treatment of myelofibrosis secondary to polycythaemia vera or essential thrombocythaemia*. London: European Medicines Agency; 2011. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Orphan_designation/2009/10/WC500006067.pdf
12. McNally RJ, Rowland D, Roman E, Cartwright RA. Age and sex distributions of hematological malignancies in the U.K. *Hematol Oncol* 1997;**15**:173-89.
13. Phekoo KJ, Richards MA, Moller H, Schey SA. The incidence and outcome of myeloid malignancies in 2,112 adult patients in southeast England. *Haematologica* 2006;**91**:1400-4.
14. Girodon F, Bonicelli G, Schaeffer C, Mounier M, Carillo S, Lafon I, et al. Significant increase in the apparent incidence of essential thrombocythemia related to new WHO diagnostic criteria: a population-based study. *Haematologica* 2009;**94**:865-9.

15. Qureshi MR, MacLean C, McMullin MF, Harrison C. Management of myelofibrosis: a survey of current practice in the United Kingdom. *J Clin Pathol* 2012:e-pub: 2012/09/05, DOI: 10.1136/jclinpath-20120201056.
16. Novartis. Jakavi (ruxolitinib) Summary of product characteristics. [Personal
17. Tefferi A, Barosi G, Mesa RA, Cervantes F, Deeg HJ, Reilly JT, et al. International Working Group (IWG) consensus criteria for treatment response in myelofibrosis with myeloid metaplasia, for the IWG for Myelofibrosis Research and Treatment (IWG-MRT). *Blood* 2006;**108**:1497-503.
18. Abgrall J-F, Guibaud I, Bastie J-N, Flesch M, Rossi J-F, Lacotte-Thierry L, et al. Thalidomide versus placebo in myeloid metaplasia with myelofibrosis: a prospective, randomized, double-blind, multicenter study. *Haematologica* 2006;**91**:1027-32.
19. Tefferi A, Verstovsek S, Barosi G, Passamonti F, Roboz GJ, Gisslinger H, et al. Pomalidomide is active in the treatment of anemia associated with myelofibrosis. *J Clin Oncol* 2009;**27**:4563-9.
20. Bacigalupo A, Van Lint M, Olivieri A, Casini M, Angelucci E, Narni F, et al. A randomized trial of bone marrow (BM) versus peripheral blood (PB) allogeneic hemopoietic stem cell transplants (HSCT) in patients with myeloproliferative disorders. *Blood* 2006;**108**:850A-A.
21. Verstovsek S, Mesa RA, Gotlib J, Levy RS, Gupta V, DiPersio JF, et al. A double-blind, placebo-controlled trial of ruxolitinib for myelofibrosis. *N Engl J Med* 2012;**366**:799-807.
22. Harrison C, Kiladjian J-J, Al-Ali HK, Gisslinger H, Waltzman R, Stalbovskaya V, et al. JAK inhibition with ruxolitinib versus best available therapy for myelofibrosis. *N Engl J Med* 2012;**366**:787-98.
23. Verstovsek S, Kantarjian H, Mesa RA, Pardanani AD, Cortes-Franco J, Thomas DA, et al. Safety and efficacy of INCB018424, a JAK1 and JAK2 inhibitor, in myelofibrosis. *N Engl J Med* 2010;**363**:1117-27.
24. Harrison CN, Kiladjian J, Al-Ali HK, Gisslinger H, Waltzman RJ, Stalbovskaya V, et al. Results of a randomized study of the JAK inhibitor ruxolitinib (INC424) versus best available therapy (BAT) in primary myelofibrosis (PMF), post-polycythemia vera-myelofibrosis (PPV-MF) or post-essential thrombocythemia-myelofibrosis (PET-MF). *J Clin Oncol* 2011;**29**.
25. Verstovsek S, Mesa R, Gotlib J, Levy R, Gupta V, DiPersio J, et al. *Long-term outcome of Ruxolitinib treatment in patients with Myelofibrosis: durable reductions in spleen volume, improvements in quality of life, and overall survival advantage in COMFORT-I. Abstract Number 800*. In: 54th ASH Annual Meeting. Atlanta; 2012. Available from: https://ash.confex.com/ash/2012/webprogram/Paper48272.html
26. Cervantes F, Kiladjian J, Niederwieser D, Sirulnik A, Stalbovskaya V, McQuity M, et al. *Long-term safety, efficacy, and survival findings from COMFORT-II, a phase 3 study comparing ruxolitinib with best available therapy (BAT) for the treatment of myelofibrosis (MF). Abstract Number 801*. In: 54th ASH Annual Meeting. Atlanta; 2012. Available from: https://ash.confex.com/ash/2012/webprogram/Paper49101.html
27. European Medicines Agency (EMA). *Annex I: Summary of product characteristics* London: European Medicines Agency (EMA); 2012.
28. National Institute for Health and Clinical Excellence (NICE). *Final scope for the proposed appraisal of ruxolitinib for the treatment of myelofibrosis*. National Institute for Health and Clinical Excellence: London; 2012.
29. Haematological Malignancy Research Network (HMRN). *Clinical management, resource utilisation and outcome in primary and secondary myelofibrosis*. 2012. [cited 2012 November 30]. Available from: http://www.hmrn.org/Resources/Reports/Documents/HMRN_Myelofibrosis_Audit.pdf.

30. Verstovsek S, Kantarjian HM, Estrov Z, Cortes JE, Thomas DA, Kadia T, et al. Long-term outcomes of 107 patients with myelofibrosis receiving JAK1/JAK2 inhibitor ruxolitinib: survival advantage in comparison to matched historical controls. *Blood* 2012;**120**:1202-9.
31. RTIHealthSolutions. *Treatment patterns, clinical outcomes, and the associated burden of myelofibrosis in the United States. Novartis data on file JAK-001-12*; 2011 1 June 2011.
32. Barosi G, Ambrosetti A, Buratti A, Finelli C, Liberato NL, Quaglini S, et al. Splenectomy for patients with myelofibrosis with myeloid metaplasia: pretreatment variables and outcome prediction. *Leukemia* 1993;**7**:200-6.
33. Tefferi A. Myelofibrosis with myeloid metaplasia. *N Engl J Med* 2000;**342**:1255-65.
34. Elliott MA, Chen MG, Silverstein MN, Tefferi A. Splenic irradiation for symptomatic splenomegaly associated with myelofibrosis with myeloid metaplasia. *Br J Haematol* 1998;**103**:505-11.
35. Bouabdallah R, Coso D, Gonzague-Casabianca L, Alzieu C, Resbeut M, Gastaut JA. Safety and efficacy of splenic irradiation in the treatment of patients with idiopathic myelofibrosis: A report on 15 patients. *Leuk Res* 2000;**24 (6)**:491-5.
36. Verstovsek S, Estrov Z, Cortes JE, Thomas DA, Borthakur G, Kadia T, et al. The MD Anderson Cancer Center (MDACC) experience with ruxolitinib, an oral JAK1 and JAK2 inhibitor, in myelofibrosis: long-term follow-up outcomes of 107 patients from a phase I/II study. *Blood* 2011;**118**:1646.
37. Mesa RA, Li C-Y, Ketterling RP, Schroeder GS, Knudson RA, Tefferi A. Leukemic transformation in myelofibrosis with myeloid metaplasia: a single-institution experience with 91 cases. *Blood* 2005;**105**:973-7.
38. Kiladjian J, Gisslinger H, Passamonti F, Niederwieser D, Mendelson E. *Health-related quality of life and symptom burden in patients with myelofibrosis in the COMFORT-II study*. In: ASCO Annual Meeting. Chicago; 2012.
39. Scott NW, Fayers PM, Aaronson NK, Bottomley A, Graeff A, Groenvold M, et al. *EORTC QLC-30 reference values*. Brussels: European Organisation for Research and Treatment of Cancer (EORTC); 2008. Available from: http://groups.eortc.be/qol/sites/default/files/img/newsletter/reference_values_manual2008.pdf
40. Homewood J, Watson M, Richards SM, Halsey J, Shepherd PCA. Treatment of CML using IFN-alpha: impact on quality of life. *Hematology Journal* 2003;**4**:253-62.
41. Martinelli F, Quinten C, Maringwa JT, Coens C, Vercauteren J, Cleeland CS, et al. Examining the relationships among health-related quality-of-life indicators in cancer patients participating in clinical trials: a pooled study of baseline EORTC QLQ-C30 data. *Expert Review of Pharmacoeconomics & Outcomes Research* 2011;**11**:587-99.
42. Dalziel K, Round A, Stein K, Garside R, A P. *The effectiveness and cost effectiveness of imatinib for first line treatment of chronic myeloid leukaemia in chronic phase*. Exeter: Peninsula Technology Assessment Group, University of Exeter; 2003. Available from: http://www.nice.org.uk/nicemedia/live/11515/32751/32751.pdf
43. *Rituximab for the first line maintenance treatment of follicular non-Hodgkin's lymphoma: Roche submission to the National Institute for Health and Clinical Excellence, 10 August 2010*. [cited 2012 November 30]. Available from: http://www.nice.org.uk/nicemedia/live/11863/51922/51922.pdf.
44. Curtis L. *Unit costs of health and social care 2011*. Canterbury, Kent: Personal Social Services Research Unit (PSSRU); 2011. Available from: http://www.pssru.ac.uk/pdf/uc/uc2011/uc2011.pdf
45. *NHS Trusts and PCTs combined reference cost schedules 2010-11*. Department of Health; [cited 2012 November 30]. Available

from: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_131140.
46. NHS Blood and Transplant. *Annual review 2011-12: saving and improving lives*. Watford: NHS Blood and Transplant; 2011. Available from: http://www.nhsbt.nhs.uk/annualreview/pdf/nhsbt_annual_review_2011-2012.pdf
47. Coyle D, Small N, Ashworth A, Hennessy S, Jenkins-Clarke S, Mannion R, et al. Costs of palliative care in the community, in hospitals and in hospices in the UK. *Critical Reviews in Oncology-Hematology* 1999;**32**:71-85.
48. Katz LM, Howell JB, Doyle JJ, Stern LS, Rosenblatt LC, Piech CT, et al. Outcomes and charges of elderly patients with acute myeloid leukemia. *Am J Hematol* 2006;**81**:850-7.
49. Stalbovskaya V. *COMFORT II: Additional outputs for HE model. Novartis Internal Report, Commercial in Confidence*; 2012 26 January 2012.