CONFIDENTIAL UNTIL PUBLISHED

External Assessment Group Report commissioned by the NIHR Evidence Synthesis Programme on behalf of NICE

Ruxolitinib for treating polycythaemia vera (ID5106)

Produced by	Southampton Health Technology Assessments Centre (SHTAC)
Authors	Joanne Lord, Professorial Research Fellow, Health Economics
	Lois Woods, Senior Research Assistant, Evidence Synthesis and Information Specialist
	Asyl Liyakat Hawa, Research Fellow, Health Economics
	David Alexander Scott, Principal Research Fellow, Statistics
	Geoff Frampton, Senior Research Fellow, Evidence Synthesis
Correspondence to	Dr Geoff Frampton
	Southampton Health Technology Assessments Centre (SHTAC)
	School of Healthcare Enterprise and Innovation
	Alpha House
	Enterprise Road, University of Southampton Science Park
	Southampton SO16 7NS
	www.southampton.ac.uk/shtac
Date completed	21st November 2022
	Updated following factual accuracy check 8th December 2022

We also thank Dr Joanna Picot, Southampton Health Technology Assessments Centre (SHTAC), for reading and commenting on a draft of this report for quality assurance.

Declared competing interests of the authors and advisors

- The authors declare none.
- Dr Innes received an honorarium from Novartis (manufacturer of ruxolitinib) to cover the registration fee for the European Haematology Association Annual Meeting, June 2022. He declares that this did not involve any consultancy work for Novartis.
- Dr Narayanan received an honorarium from Novartis (manufacturer of ruxolitinib) to attend a virtual educational masterclass (10-11 February 2022) covering mixed haematology topics. He declares that he did not attend any sessions relevant to ruxolitinib or polycythaemia vera. Dr Narayanan also received honoraria from Bristol-Myers Squibb (manufacturer of hydroxycarbamide) to attend advisory board meetings on myelodysplastic syndrome and multiple myeloma (ongoing from March 2022); to attend the European Haematology Association meeting (online; June 2022); and to attend a Myeloproliferative Neoplasms Preceptorship virtual meeting covering general topics in myeloproliferative disorders (6-7 October 2022). He is also an investigator on a study of low-risk myelodysplasia unrelated to polycythaemia vera (MIRAS project ID 313756) sponsored by Bristol-Myers Squibb.

Copyright is retained by Novartis for the following:

- EAG report tables 17, 23, 24
- Information in parts of ERG report tables 9, 10, 12-15, 22, 26, 25, Appendix 9.2, Appendix 9.3
- EAG report figures 3, 4

Rider on responsibility for report

The view expressed in this report are those of the authors and not necessarily those of the NIHR Evidence Synthesis Programme. Any errors are the responsibility of the authors.

This report should be referenced as follows:

Lord J, Woods L, Hawa AL, Scott DA, Frampton G. Ruxolitinib for treating polycythaemia vera: A Single Technology Appraisal. Southampton Health Technology Assessments Centre, 2022.

Contributions of authors

Joanne Lord critically appraised the health economic systematic review, critically appraised the economic evaluation, and drafted the report; Lois Woods critically appraised the clinical effectiveness systematic review and drafted the report; Asyl Liyakat Hawa critically appraised the health economic systematic review, critically appraised the economic evaluation, and drafted the report; David Alexander Scott critically appraised the indirect treatment comparison and the clinical effectiveness systematic review and drafted the report; Geoff Frampton critically appraised the clinical effectiveness systematic review, drafted the report and is the project co-ordinator and guarantor.

1 Contents

LIST OF TABLES

6

LIST of FIGURES

LIST OF ABBREVIATIONS

AE	Adverse event
AIC	Academic in confidence
AML	Acute myeloid leukaemia
BAT	Best available therapy
BNF	British National Formulary
BSH	British Society for Haematology
CHR	Complete haematological remission
CI	Confidence interval
CIC	Commercial in confidence
CRD	Centre for Reviews and Dissemination
CS	Company submission
CQ	Clarification question
CSR	Clinical study report
DSU	Decision Support Unit
EAG	External Assessment Group
EFS	Event-free survival
ELN	European LeukemiaNet
EMA	European Medicines Agency
EMC	Electronic Medicines Compendium
EORTC QLQ-C30	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire
EPAR	European Public Assessment Report
EQ-5D-3L	European Quality of Life Working Group Health Status Measure 3 Dimensions, 3 Levels
EQ-5D-5L	European Quality of Life Working Group Health Status Measure 5 Dimensions, 5 Levels
EQ-VAS	EuroQol Visual Analogue Scale
ESMO	European Society for Medical Oncology
GEMFIN	Grupo Español de Enfermedades Mieloproliferativas Filadelfia Negativas
HC	Hydroxycarbamide (this is synonymous with hydroxyurea)
HCT	Haematocrit
HMRN	Haematological Malignancy Research Network

8

HR	Hazard ratio
HRG	Healthcare Resource Group
HRQoL	Health-related quality of life
HTA	Health technology assessment
ICER	Incremental cost-effectiveness ratio
IFN	Interferon
IPD	Individual patient level data
IRR	Incidence-rate ratios
ITC	Indirect treatment comparison
ITT	Intent(ion) to treat
JAK	Janus-associated Kinase
KM	Kaplan-Meier
MAIC	Matched-adjusted indirect comparison
MDS	Myelodysplastic syndrome
MF	Myelofibrosis
MF-8D	Myelofibrosis 8 dimensions health outcome measure
MHRA	Medicines and Healthcare products Regulatory Agency
MPN(s)	Myeloproliferative neoplasm(s)
MPN-SAF TSS	Myeloproliferative Neoplasm Symptom Assessment Form Total Symptom Score
NA	Not applicable
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NMSC	Non-melanoma skin cancer
NR	Not reported
OR	Odds ratio
OS	Overall survival
PAS	Patient Access Scheme
PFS	Progression-free survival
PSA	Probabilistic sensitivity analysis
PSM	Partitioned survival model
PSS	Personal Social Services
PV	Polycythaemia vera
R/I	Resistant and/or intolerant
QALY	Quality-adjusted life year

QoL	Quality of life
RCT	Randomised controlled trial
RR	Relative risk/risk ratio
SAE	Serious adverse event
SD	Standard deviation
SE	Standard error
SLR	Systematic literature review
SmPC	Summary of product characteristics
STM	State transition model
TA	Technology appraisal
TE	Thromboembolic event
TEAE	Treatment-emergent adverse event
TFS	Transformation-free survival
TSD	Technical Support Document
TTD	Time to discontinuation
UK	United Kingdom
US	United States
VAS	Visual analogue scale
WBC	White blood cell(s)

EXECUTIVE SUMMARY

This summary provides a brief overview of the key issues identified by the external assessment group (EAG) as being potentially important for decision making. It also includes the EAG's preferred assumptions and the resulting incremental cost-effectiveness ratios (ICERs).

Section 1.1 provides an overview of the key issues. Section 1.2 provides an overview of key model outcomes and the modelling assumptions that have the greatest effect on the ICER. Sections 1.3 to 1.6 explain the key issues in more detail. Background information on the condition, health technology, evidence and information on the issues are in the main EAG report.

All issues identified represent the EAG's view, not the opinion of the National Institute for Health and Care Excellence (NICE).

1.1   Overview of the EAG's key issues

Table 1 Summary of key issues

Issue number	Headline description	EAG report section
1	Relevance of the trial populations for modelling UK practice	4.2.3
2	Modelling the relative treatment effect for overall survival	4.2.6.2.1
3	Waning of the treatment effect	4.2.6.2.1
4	Modelling approach: state-transition or partitioned-survival	4.2.2.3
5	Model structure: health states and events	4.2.2.3
6	Extrapolation of time to ruxolitinib discontinuation	4.2.6.1.1
7	Source for utility estimates: MF-8D or EQ-5D	4.2.7.2

The key differences between the company's preferred assumptions and the EAG's preferred assumptions are :

- Use of the general population mortality constraint for survival prior to discontinuation of ruxolitinib throughout the time horizon, rather than only post-trial.
- Partitioning of the best available treatment (BAT) state into substates for first BAT, second or subsequent BAT and no further BAT.
- Estimates for the hazard ratio (HR) for ruxolitinib compared with BAT from the MAJIC-PV trial, constant or time-varying HR.

11

- The distribution used for extrapolation of the time to ruxolitinib discontinuation.
- Source for estimates of utilities for ruxolitinib and BAT: EQ-5D values from RESPONSE-2 trial data or MF-8D values from the RESPONSE trial.

1.2 Overview of key model outcomes

NICE technology appraisals compare how much a new technology improves length (overall survival) and quality of life in a quality-adjusted life year (QALY). An ICER is the ratio of the extra cost for every QALY gained.

Overall, the technology is modelled to affect QALYs by:

- Lower mortality rates while patients are on ruxolitinib than with standard therapies.
- Better health-related quality of life (utility) while patients are on ruxolitinib than during treatment with standard therapies alone.
- Small overall increase in utility due to reduced incidence of myelofibrosis, thromboembolism, haemorrhage, adverse reactions and therapeutic phlebotomy.

Overall, the technology is modelled to affect costs by:

- The high cost of ruxolitinib compared with standard drug treatments.
- Savings due to reduced use of therapeutic phlebotomy and reduced follow-up and monitoring after the first six months of treatment with ruxolitinib.
- Savings due to reduced need for treatment of myelofibrosis, haemorrhage, thromboembolism and adverse reactions.
- Some additional costs for treatment of non-melanoma skin cancer, acute myeloid leukaemia and myelodysplastic syndrome.

The modelling assumptions that have the greatest effect on the ICER are:

- The hazard ratio for overall survival with ruxolitinib compared with best available therapy.
- Assumptions about waning of the treatment effect for overall survival.
- The distribution used for extrapolation of time to discontinuation of ruxolitinib.
- Use of EQ-5D or MF-8D utility estimates for ruxolitinib and best available therapy.

1.3 The decision problem: summary of the EAG's key issues

The EAG have not identified any key issues with the decision problem. Other issues relating to the decision problem are discussed in section 1.6 below.

1.4 The clinical effectiveness evidence: summary of the EAG's key issues

The EAG have not identified any key issues with the clinical effectiveness evidence. Other issues relating to the clinical effectiveness evidence are discussed in section 1.6 below.
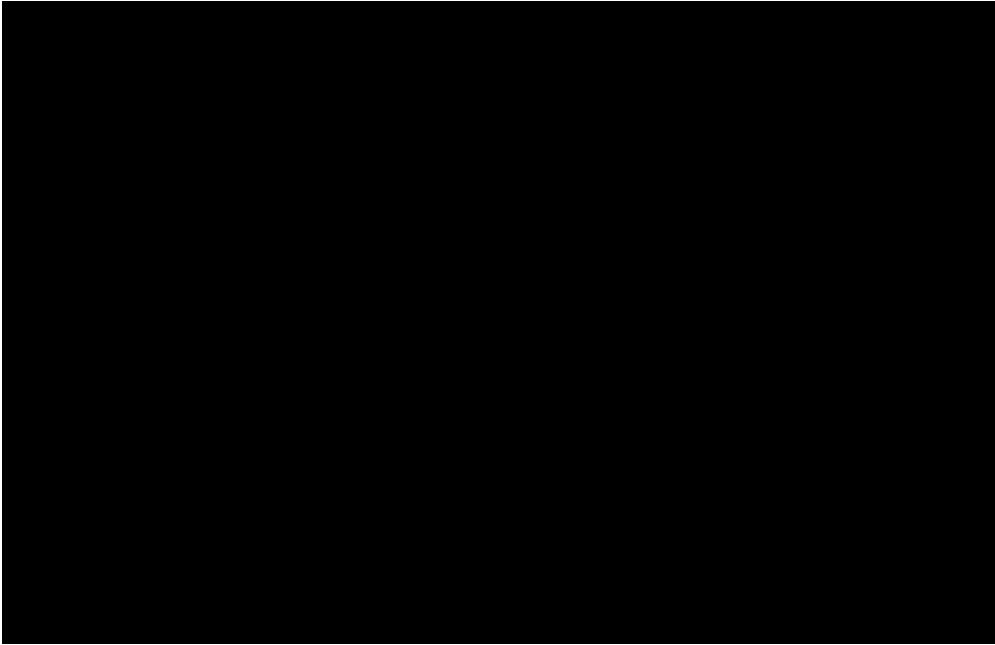
1.5 The cost-effectiveness evidence: summary of the EAG's key issues

Issue 1 Relevance of the trial populations for modelling UK practice

Report section	4.2.3
Description of issue and why the EAG has identified it as important	There is some uncertainty over whether the MAJIC-PV trial or the company's RESPONSE and RESPONSE-2 trials provide a better basis for modelling survival for the relevant patient population in UK practice. This issue is important because cost-effectiveness estimates differ for versions of the model based on the three trial populations. The EAG considers that, as MAJIC-PV was a wholly UK based trial, it is more obviously relevant for the UK PV population and clinical context. This reflects the view of clinical experts consulted by the EAG. The company have put forward the view that the patients recruited to MAJIC-PV represent a 'high-risk' subgroup of the licensed indication for ruxolitinib. In their 'primary' model, the company use survival extrapolations fitted to RESPONSE and RESPONSE-2 data. Alongside this, they report a 'subgroup model' with extrapolations fitted to MAJIC-PV data.
What alternative approach has the EAG suggested?	We consider that the MAJIC-PV trial population is likely to provide a more appropriate basis for modelling outcomes in UK practice. But we also report cost-effectiveness results based on the RESPONSE and RESPONSE-2 populations, as these provide a comparison for the subgroups with and without splenomegaly.
What is the expected effect on the cost-effectiveness estimates?	ICER estimates are lower for the MAJIC-PV population. With the company's base case assumptions, the ICERs are ██████, ██████ and ██████ per QALY for the RESPONSE, RESPONSE-2 and MAJIC-PV populations, respectively. With the EAG preferred assumptions, these ICERs are ██████, ██████ and ██████ respectively.

13

What additional evidence or analyses might help to resolve this key issue?	Further expert opinion and evidence on the relevance of the three trial populations to UK practice.

Issue 2 Modelling the relative treatment effect for overall survival

Report section	4.2.6.2.1, Table 22, Table 27 and
	Figure 5 and Figure 6 below show the KM data with the company's choice of distribution for TTD for ruxolitinib due to reasons other than death in comparison with the selected scenario distributions from Table 27 above for the licensed population with and without splenomegaly.
	Figure 5 Comparison of selected scenario distributions for TTD for ruxolitinib for the licensed population with splenomegaly
	Abbreviations: KM: Kaplan-Meier; TTD: time to treatment discontinuation; OS: overall survival. Source: Reproduced from CS Appendix N Figure 18 using selected distributions.

14

Figure 6 Comparison of selected scenario distributions for TTD for ruxolitinib for the licensed population without splenomegaly

Abbreviations: KM: Kaplan-Meier; TTD: time to treatment discontinuation; OS: overall survival.
Source: Reproduced from CS Appendix N Figure 19 using selected distributions.

Table 28 below shows cost-effectiveness results for selected company scenarios for the MAJIC-PV population analysis. Again, from the many scenarios conducted by the company, we have selected scenarios that relate to key uncertainties and that have an impact on the ICERs.

Table 28

Description of issue and why the EAG has identified it as important	Cost-effectiveness is highly sensitive to the relative treatment effect on overall survival.
	The company use results from the MAJIC-PV trial to inform estimates for their base case analyses. We agree with this decision as cross-over within the RESPONSE and RESPONSE-2 trials means that estimates of treatment effects from these trials are highly confounded. The EAG are not aware of any other data that would provide a more robust analysis. Other sources of evidence regarding the effect of ruxolitinib on survival, including the company's ITC and an analysis of Spanish registry data are less robust. The currently unpublished manuscript for the MAJIC-PV reports a hazard ratio for overall survival (ruxolitinib compared with best available treatment) of ███████████████████████).
	However, the company use a time-varying estimate of the hazard ratio, which they estimated with a piecewise Cox proportional hazards model using reconstructed Kaplan-Meier data from MAJIC-PV. This includes a bigger treatment effect (lower HR) from year █ onwards:

15

	███. The company justifies this approach based on expert advice and visual inspection and analysis of the MAJIC-PV KM results.
What alternative approach has the EAG suggested?	The EAG prefer the constant HR estimate from MAJIC-PV due to uncertainty over the statistical validity of the company's post hoc analysis. However, we report a scenario results with the company's time-varying HR, as this may be considered clinically plausible.
What is the expected effect on the cost-effectiveness estimates?	The HR for OS has a large impact on the ICER. The company's base case estimates increase from ████, ████ and ████ (RESPONSE, RESPONSE-2 and MAJIC-PV populations respectively), to ████, ████ and ████
What additional evidence or analyses might help to resolve this key issue?	Further expert opinion on the plausibility of an increasing relative effect on survival over time. The economic analyses for subgroups with and without splenomegaly currently use the same estimates of treatment effects, estimated form the MAJIC-PV trial. Further analysis should be conducted to update these analyses if subgroup analysis of MAJIC-PV data by splenomegaly status.

Issue 3 Waning of the treatment effect

Report section	4.2.6.2.1
Description of issue and why the EAG has identified it as important	In their base case analyses, the company assume that the treatment effect diminishes linearly from the end of trial follow-up (5 years) and stops at 20 years (HR=1). This was based on clinical expert judgement that approximately twice the number of patients would be alive at 20 years with ruxolitinib compared with current treatment (see CS section B.3.3.4). The company note uncertainty over these assumptions, and report scenario analysis with the period of waning varied from 5 to 50 years.
What alternative approach has the EAG suggested?	We have not changed the company's waning assumptions in EAG preferred analysis, as the assumption of waning might be seen to mitigate against uncertainty over the treatment effect. However, we note that it might be appropriate to use a longer waning period, or to remove waning from the model, when used in combination with the more conservative fixed HR estimate.
What is the expected effect on the cost-effectiveness estimates?	The duration of waning has a big impact on the ICER. For example, the company's base case ICER for the MAJIC-PV population is ████ with a loss of effect at 10 years, and ████ with loss of effect at 30 years.
What additional evidence or analyses might help to resolve this key issue?	Further expert opinion on the plausibility of waning from a biological and clinical perspective.

Issue 4 Modelling approach: state-transition or partitioned-survival

Report section	4.2.2.3
Description of issue and why the EAG has identified it as important	It is not clear if different results from the company's state-transition model (STM) for the RESPONSE and RESPONSE-2 populations and their partitioned-survival model (PSM) for the MAJIC-PV population relate to differences in the modelling technique or to the different populations. This adds structural uncertainty to the interpretation of the economic evaluation results. NICE Decision Support Unit Technical Support Document 19 reports that STM and PSM models can give very different results, and that it is not clear which approach is more reliable. TSD19 therefore recommends parallel development of STM and PSM models to verify the plausibility of PSM extrapolations.
What alternative	Comparison of alternative modelling approaches (STM and

approach has the EAG suggested?	PSM) within the same dataset.
What is the expected effect on the cost-effectiveness estimates?	Unknown
What additional evidence or analyses might help to resolve this key issue?	Development of a PSM for the RESPONSE and RESPONSE-2 populations to enable comparison with results from the STM model. It is not possible for the company to develop an STM for the MAJIC-PV population, as they do not have access individual patient data. However, we would encourage the MAJIC-PV investigators to consider appropriate economic evaluation based on the trial data, or to make the data available for such an analysis.

Issue 5 Model structure: health states and events

Report section	4.2.2.3
Description of issue and why the EAG has identified it as important	The EAG also has concerns over the structure of the company's models, as they do not reflect the natural history of PV, and therefore may not reflect long-term impacts of the condition on survival and quality of life. The model 'health states' are based on treatment phases (before and after discontinuation of ruxolitinib) rather than on stages of disease. Although discontinuation of ruxolitinib is likely to be related to long-term survival, other intermediate outcomes such as progression-free survival or event-free survival are likely to be more strongly prognostic. Another problem with the current structure, is that the best available therapy (BAT) arm is modelled with a single health state, with three substates for first-line, second and subsequent line, and discontinuation of all BAT. EAG clinical advisors have suggested that this progression between lines of therapy does not reflect current practice. Furthermore, the decrements in utility for the latter two substates are based on assumption, rather than evidence. We also have concerns that the company's model structure does not reflect increasing risks of key complications of PV, such as myelofibrosis, and major thromboembolic or haemorrhagic events with age. The use of fixed incidence annual rates for these events is not realistic.

18

What alternative approach has the EAG suggested?	Consideration of an alternative model structure based on a measure of disease progression and a simplified approach to modelling the subsequent types of event.
What is the expected effect on the cost-effectiveness estimates?	Unknown
What additional evidence or analyses might help to resolve this key issue?	Exploration of an alternative model structure to better reflect the natural history of PV.

Issue 6 Extrapolation of time to ruxolitinib discontinuation

Report section	4.2.6.1.1
Description of issue and why the EAG has identified it as important	The results for the company's primary analysis based on the RESPONSE and RESPONSE-2 trials were moderately sensitive to the distribution used for the time to treatment discontinuation. The company used an odd spline model with one knot for the extrapolation of TTD for ruxolitinib due to reasons other than death in the primary analysis. The same distribution was used for both RESPONSE and RESPONSE-2 trial data. The EAG note that, in the primary analysis, pre- and post-discontinuation survival for ruxolitinib make use of pooled RESPONSE and RESPONSE-2 data, as few deaths were observed in the trial, whereas data from the two trials are used separately for TTD for ruxolitinib due to reasons other than death.
What alternative approach has the EAG suggested?	The EAG have selected the Weibull distribution as a preferred assumption for TTD for ruxolitinib, a parametric distribution which has a better fit the RESPONSE trial data more appropriately. The Weibull distribution has a similar fit for the RESPONSE-2 trial data.
What is the expected effect on the cost-effectiveness estimates?	Implementing a Weibull distribution in place of an odds spline model in the company base case reduces the ICER for the licensed population with splenomegaly to ▮▮▮▮ per QALY and increases the ICER for the licensed population without splenomegaly to ▮▮▮▮ per QALY.
What additional evidence or analyses might help to resolve this key issue?	Additional scenario using pooled IPD from RESPONSE and RESPONSE-2 trials for TTD for ruxolitinib due to reasons other than death.

Issue 7 Source for utility estimates: MF-8D or EQ-5D

Report section	4.2.7.2
Description of issue and why the EAG has identified it as important	There is uncertainty over the most appropriate instrument to estimate utilities for the economic model. This has a large impact on the ICER. Utilities are available from two sources: EQ-5D-5L data from the RESPONSE-2 trial, and estimates from data collected in the RESPONSE trial and valued using the MF-8D, which is a disease-specific utility measure developed for myelofibrosis. The company argue that the EQ-5D is not appropriate for PV, based on psychometric evidence and precedent for myelofibrosis (TA386 and TA756), and the similar nature of symptoms for PV and MF. They also report an exploratory psychometric analysis comparing RESPONSE-2 data for the EQ-5D and a PV symptom score (the MPN-SAF). This provides some evidence in favour of the MF-8D, including greater responsiveness and lower susceptibility to ceiling effects. However, the MF-8D was not developed for use in PV, and the company had to make assumptions to substitute the PV symptom score for the myelofibrosis symptom score used in the MF-8D. There is also a lack of direct evidence validating the EQ-5D and MF-8D in a PV population.
What alternative approach has the EAG suggested?	We use EQ-5D utilities in the EAG preferred analysis. This follows the NICE preference for use of the EQ-5D when available from relevant clinical trials and improves consistency across NICE appraisals. There is some evidence in favour of the MF-8D measure, but also uncertainty about its transferability from MF to PV.
What is the expected effect on the cost-effectiveness estimates?	Replacing MF-8D with EQ-5D utilities in the company's base case increases the ICER for the MAJIC-PV population ███████ to ███████ per QALY. Increases are similar in the RESPONSE and RESPONSE-2 populations.
What additional evidence or analyses might help to resolve this key issue?	Further evidence that the EQ-5D is not appropriate for people with PV. Comparative evidence for the psychometric performance of MF-8D and EQ-5D utilities for a population with PV

1.6 Other issues

The company have excluded radioactive phosphorus from their decision problem although this is stated as a relevant comparator in the NICE scope. As explained in section 2.3.2

below, we believe the exclusion of radioactive phosphorus is appropriate and unlikely to influence validity of the cost-effectiveness results.

The results of the company's indirect treatment comparison (ITC) for overall survival are highly uncertain, primarily due to limited adjustment for imbalances in prognostic factors between the treatment groups (section 3.4). However, the EAG are not aware of alternative data sources that would enable a more robust ITC analysis to be conducted. Overall survival estimates from the ITC are not used in the company's economic analysis base case but do inform scenario analyses (section 4.2.6.2.1).

All three randomised controlled trials included by the company are at high risk of bias, due to the open-label nature of the trials, confounding of long-term outcomes by crossover in the RESPONSE and RESPONSE-2 trials, selective reporting of HRQoL outcomes, and the handling of missing data for HRQoL outcomes in all three trials. For MAJIC-PV there is additionally a lack of clarity around the randomisation process and there are some differences in patient characteristics between the treatment arms (section 3.2.3). Limitations of the existing data and reporting mean that the clinical efficacy outcomes are subject to uncertainty that would be difficult to resolve unless new evidence (and clearer reporting of studies) becomes available. The high risk of bias means that variance estimates from the three RCTs such as 95% confidence intervals would underestimate the uncertainty present.

The survival extrapolations used in the company's base case incorporate a constraint to ensure that the mortality rate cannot be less than that in the general population (adjusted for age and gender). This constraint is applied through the time horizon, except for survival prior to discontinuation of ruxolitinib in the company's primary model, for which the general population mortality constraint was only applied after the trial period (5 years). In response to clarification question B4, the company provided a scenario analysis including the mortality constraint throughout the time horizon and a revised version of their model with an option to apply this scenario. We consider this to be a correction to the company's model and have applied it in EAG preferred analyses.

Other issues that have a limited impact on ICERs are: the EAG adjustment to the cost of managing grade 1 and 2 thromboembolic events; and use of the partition of the BAT state to model first line BAT, second and subsequent line BAT and no further BAT substates.

1.7 Summary of EAG's preferred assumptions and resulting ICER

We made the following changes to the company's base case analyse in the EAG preferred analysis:

- Correction to apply the general population mortality constraint for survival prior to discontinuation of ruxolitinib throughout the time horizon
- The partition of the BAT health state was not used
- Constant HR for overall survival from the MAJIC-PV trial
- Weibull extrapolation for time to ruxolitinib discontinuation in the primary model
- EQ-5D utility values estimated from the RESPONSE-2 trial
- Additional costs for management of Grade 1-2 thromboembolic events

Table 2 Summary of cost-effectiveness results

Scenario	Incremental cost	Incremental QALYs	ICER (change from company base case)
Company's base case			
RESPONSE trial population (with splenomegaly)	■	■	■
RESPONSE-2 trial population (without splenomegaly)	■	■	■
MAJIC-PV trial population	■	■	■
EAG's preferred base case			
RESPONSE trial population (with splenomegaly)	■	■	■
RESPONSE-2 trial population (without splenomegaly)	■	■	■
MAJIC-PV trial population	■	■	■

Modelling errors identified and corrected by the EAG are described in section 5.3.3. For further details of the exploratory and sensitivity analyses done by the EAG, see section 6.1.

22

2 INTRODUCTION AND BACKGROUND

2.1 Introduction

This report is a critique of the company's submission (CS) to NICE from Novartis on the clinical effectiveness and cost effectiveness of ruxolitinib for treating polycythaemia vera (PV). It identifies the strengths and weaknesses of the CS. Clinical experts were consulted to advise the external assessment group (EAG) and to help inform this report.

Clarification on some aspects of the CS was requested from the company by the EAG via NICE on 10th October 2022. A response from the company via NICE was received by the EAG on 27th October 2022 and this can be seen in the NICE committee papers for this appraisal.

2.2 Background

2.2.1 Background information on polycythaemia vera

Polycythaemia vera (PV) is a type of myeloproliferative neoplasm (MPN) characterised by overproduction of blood cells and platelets in the bone marrow, particularly red blood cells (erythrocytosis).[1][2] The uncontrolled nature of the proliferation of blood cells defines PV as a cancer.[3]

CS section B.3.1.1 provides a clear overview of the disease including: a brief description; epidemiology; relevance of the Janus-associated Kinase (JAK) 2 mutation; diagnosis (blood cell counts, and the haematocrit which is the proportion of red blood cells in a volume of blood, usually expressed as a percentage); symptoms (the most significant being splenomegaly, pruritus (itching), and fatigue); mortality associated with thromboembolic events, cardiovascular events and disease progression; and a discussion of the definitions of high-risk disease and resistance or intolerance to hydroxycarbamide (also discussed in section 3.2.1 of this report).

CS section B.1.3.1 notes the association of increased haematocrit (HCT) levels, i.e. an increased red blood cell mass with vascular complications. This is consistent with the British Society for Haematology (BSH) guidelines that show an increased HCT of >0.45 is a risk factor for thrombosis which in turn is a risk factor for overall survival, hence HCT control is a key goal of therapy.[4]

The EAG note that up to date incidence and prevalence data for PV specifically for England are not available. Data for the UK are available from the Haematological Malignancy Research Network (HMRN) which gives a crude estimate for incidence as 1.7 per 100,000, a prevalence of 1.9 per 100,000, and 1130 expected UK cases per year.[5] These figures sit within the ranges estimated from European registry data and other sources provided in the CS (CS section B.1.3.1).

The current treatment pathway is discussed in CS section B.1.3.2 and covers treatment goals, the course of disease progression, first-, second- and third-line treatments, unmet need, and the safety profile of other cytoreductive therapies.

- As stated in the CS, the main goals of treatment are to reduce the incidence of thrombotic and haemorrhagic complication and the long-term risk of transformation to myelofibrosis (MF) or acute myeloid leukaemia (AML).[4][6]

- European and UK guidelines exist: the European Society for Medical Oncology (ESMO) clinical practice guidelines for Philadelphia chromosome-negative chronic MPNs (which includes PV)[7]; the Pan-London Haemato-Oncology Clinical Guidelines for MPNs;[6] and the British Society for Haematology (BSH) guideline for the diagnosis and management of PV.[4] These guidelines are generally similar and have authors in common, the CS refers to the BSH guideline only which is appropriate as it is recent (2019) and applies to the whole of England.

- Cytoreductive therapy is appropriate in certain low-risk patients, for example if white blood cell (WBC) levels are high or if phlebotomy tolerability is poor. This means that such otherwise low-risk patients would join the high-risk pathway shown in CS Figure 3 (although this reason for joining the high-risk pathway is not shown fully in CS Figure 3). Therefore, not all patients who receive hydroxycarbamide may have necessarily met the criteria for high-risk based on their age or prior thrombosis.

- CS Figure 3 accurately represents the BSH recommendations for management options in high-risk patients, that is first-line treatment with either hydroxycarbamide or interferon-alfa, and second-line treatment switching to whichever of hydroxycarbamide or interferon-alfa they did not receive first-line.[4] The EAG's clinical experts agree that for first- and second-line treatments this is a good representation of clinical practice except that two of the treatments listed for third-line, pipobroman

24

and radioactive phosphorus, are no longer used (see decision problem section 2.3.2 below). One clinical expert noted that the diagram does not show that in clinical practice patients often cycle on and off hydroxycarbamide, or between hydroxycarbamide and interferon-alfa, to manage side-effects.

- CS Figure 3 refers to interferon-alfa but we note that, according to the BNF[8] and British PV guidelines, interferon-alfa has been superseded by peginterferon-alfa,[6] or is recommended in preference to interferon-alfa.[4] One of the EAG's clinical experts commented that pegylated interferon-alfa may be offered to patients who cannot tolerate interferon-alfa or hydroxycarbamide, but tolerance remains relatively poor so extensive monitoring is still required. The company's economic analysis uses costs for peginterferon-alfa (section 4.2.8.2) which the EAG agree is appropriate.

- Not all patients respond to or can tolerate hydroxycarbamide, hence the population group for the licensed indication. The CS refers to the updated ELN consensus criteria for resistance to or intolerance of hydroxycarbamide for use in clinical trials published in 2022 (CS Table 4),[9] and also states that these criteria are not always used in clinical practice, confirmed by the EAG's clinical experts (see the decision problem discussion for the population in section 2.3.1). However, the original ELN consensus criteria for resistance to or intolerance of hydroxycarbamide are relevant here as they applied at the time the studies included in the CS were conducted. Those criteria are published in Barosi et al. 2010 and duplicated in Table 3 below.[10]

25

Table 3 ELN definition of resistance/intolerance to hydroxycarbamide in patients with PV from Barosi et al. 2010[10]

Definition of resistance/intolerance to hydroxycarbamide in patients with polycythaemia vera	
1	Need for phlebotomy to keep haematocrit <45% after 3 months of at least 2 g/day of Hydroxycarbamide, OR
2 a	Uncontrolled myeloproliferation, i.e. platelet count >400 x 10^9/l AND white blood cell count >10 x 10^9/l after 3 months of at least 2 g/day of Hydroxycarbamide, OR
3	Failure to reduce massive[a] splenomegaly by more than 50% as measured by palpation, OR failure to completely relieve symptoms related to splenomegaly, after 3 months of at least 2 g/day of Hydroxycarbamide, OR
4	Absolute neutrophil count <1.0 x 10^9/l OR platelet count <100 x 10^9/l or haemoglobin <100 g/l at the lowest dose of Hydroxycarbamide required to achieve a complete or partial clinico-haematological response[b], OR
5	Presence of leg ulcers or other unacceptable Hydroxycarbamide-related non-haematological toxicities, such as mucocutaneous manifestations, gastrointestinal symptoms, pneumonitis or fever at any dose of Hydroxycarbamide
[a] Organ extending by more than 10 cm from the costal margin. [b] Complete response was defined as: haematocrit <45% without phlebotomy, platelet count ≤400 x 109/l, white blood cell count ≤10 x 109/l, and no disease related symptoms. Partial response was defined as: haematocrit <45% without phlebotomy, or response in three or more of the other criteria (Barosi et al, 2009). Table sourced directly from: Barosi et al. 2010[10]	

2.2.2 Background information on ruxolitinib

A description of ruxolitinib, brand name Jakavi®, is provided in CS section B.1.2. Ruxolitinib is a JAK1 and JAK2 protein kinase inhibitor that inhibits dysfunctional signalling pathways caused by JAK gene mutations, reducing the excessive production of red blood cells which is characteristic of PV. Ruxolitinib aims to reduce symptoms and control HCT levels in order to reduce the risk of thromboembolic events and the associated complications which can lead to death.

Ruxolitinib is licensed for the treatment of adult patients with PV who are resistant to or intolerant of hydroxycarbamide. European Medicines Agency (EMA) marketing authorisation was granted in January 2015 and UK marketing authorisation was granted in January 2021.[11] Ruxolitinib is also licensed for use in myelofibrosis and graft versus host disease.

A summary of product characteristics (SmPC) for the 10 mg tablet of ruxolitinib is provided in CS Appendix C. Ruxolitinib is taken orally in tablet form with a starting dose for PV of 10 mg twice daily. The SmPC provided in CS Appendix C specifies a 10 mg tablet only, but dosage information in CS Table 2 outlines 5 mg increments for titration based on safety and efficacy up to a maximum of 25 mg twice daily. The MHRA website lists all SmPCs for each of the 5,

10, 15 and 20 mg tablets.[11-14] Doses may be increased if efficacy is insufficient and blood counts are adequate, and they may be decreased or discontinued if blood counts fall below specified thresholds.[11] Therefore, complete blood cell counts should be evaluated prior to treatment with ruxolitinib and regularly thereafter as advised in the SmPC.[11]

2.2.3 The position of ruxolitinib in the treatment pathway

CS section B.3.1.2 proposes ruxolitinib as an alternative cytoreductive therapy as a treatment option for patients with PV who are resistant to or intolerant of hydroxycarbamide which they may have received either first-line or second-line. This is in line with positioning in the scope of this appraisal and as recommended by the BSH.[4]

One of the EAG's two clinical experts suggested that ruxolitinib might be used second-line after interferon-alfa because some patients receive interferon-alfa as their first cytoreductive therapy due to hydroxycarbamide not being suitable (e.g. younger age/family planning). However, those reasons (younger age/family planning) are not part of the definition of resistance to or intolerance of hydroxycarbamide so those patients would not be in the licensed indication. The other clinical expert said there are no data to support ruxolitinib use after interferon-alfa as first line therapy. They explained that as patients often cycle between hydroxycarbamide and interferon-alfa therapies that could create a circumstance for use of ruxolitinib third-line according to CS Figure 3.

The EAG's clinical experts indicated that they are familiar with using ruxolitinib, at higher doses, in myelofibrosis (MF) patients for whom the drug was recommended in 2016 according to NICE guideline TA386.[15] Ruxolitinib was also used in 38 UK centres as part of the MAJIC-PV randomised controlled trial (RCT) between 2012 and 2022 for PV.[16] Therefore, the NHS has experience of using ruxolitinib to treat myeloproliferative diseases.

> **EAG conclusions**
>
> The company's description of the care pathway appears appropriate, although in relation to the positioning of ruxolitinib in the pathway, there was a difference of opinion between the EAG's clinical experts about whether treatment with ruxolitinib might follow treatment with first-line interferon-alfa.

2.3 Critique of the company's definition of the decision problem

Table 4 compares the company's decision problem to the final scope for this appraisal issued by NICE. The ERG considers that the decision problem adheres to the NICE scope but with the following caveats relating to the population and comparators.

2.3.1 Population

The populations stated in the NICE scope and company decision problem are consistent. However, the EAG's clinical experts commented that definitions of hydroxycarbamide resistance and intolerance are not standardised in clinical practice so there is some uncertainty as to how well the definitions used in the clinical trials would match those used in clinical practice. The definition of intolerance can be somewhat subjective (e.g. reliant on judging the tolerability of a skin rash, leg ulcer or fatigue). One expert commented that the criteria defining hydroxycarbamide resistance and intolerance are more stringent than would be used in clinical practice. Note that the European LeukemiaNet (ELN) have recently published a consensus-based update of the definition of hydroxycarbamide resistance and intolerance (CS Table 4) (Marchetti et al. 2022[9]) but the clinical trials were completed prior to this definition being approved (clarification response A1).

2.3.2 Comparators

The EAG's clinical experts concurred that hydroxycarbamide and interferon-alfa are the most relevant comparators, with anagrelide, busulfan and radioactive phosphorus used rarely if at all:

- Radioactive phosphorus is specified in the NICE scope but excluded from the company's decision problem as the company argue that it is no longer used in practice (CS Table 1). One of the EAG's clinical experts commented that radioactive phosphorus has highly variable availability and is used very rarely. It is a one-off treatment that covers 6 months so may be of benefit for elderly frail patients unable to tolerate frequent treatments. However, it does increase the risk of leukaemia. The other expert stated that radioactive phosphorus is generally unavailable and not used. British PV guidelines suggest that radioactive phosphorus is only suitable for people with limited life expectancy.[4][6] The company have not included radioactive phosphorus among the best available therapy (BAT) treatments in their economic analysis (section 4.2.8.2) which the EAG believe is appropriate.
- Anagrelide / busulfan: Both clinical experts said they would rarely use these therapies. One commented that anagrelide increases the risk of transformation to myelofibrosis or acute myeloid leukaemia (AML) and has a poor side-effects profile especially for elderly

28

people. British PV guidelines suggest that anagrelide is rarely used as it is relatively platelet-specific, but it may be used in combination with hydroxycarbamide for people with difficult platelet control.[4][6] Busulfan increases the risk of transformation to leukaemia and is only used for people with limited life expectancy.[4][6]

- The NICE scope and company decision problem refer to interferon-alfa. As noted in section 2.2.1 above, interferon-alfa has largely been replaced in practice by peginterferon-alfa which has a relatively better tolerability.

Table 4 Summary of the decision problem

	Final scope issued by NICE	Company's decision problem	Rationale if different from the final NICE scope	EAG comments
Population	Adults with PV that is resistant or intolerant to hydroxycarbamide	In line with final scope	Not applicable	The scope and decision problem are consistent. However, the EAG's clinical experts noted that there is no single standard definition of hydroxycarbamide resistance or intolerance in clinical practice and definitions of intolerance may be subjective (section 2.3.1).
Intervention	Ruxolitinib with established clinical management	In line with final scope	Not applicable	The scope and decision problem are consistent.
Comparators	Established clinical practice without ruxolitinib, comprising of treatment with phlebotomy and aspirin, and: • hydroxycarbamide • IFN-alfa • anagrelide • busulfan • radioactive phosphorus	Established clinical practice defined as treatment with phlebotomy and aspirin, and BAT, including: • hydroxycarbamide • IFN-alfa • anagrelide • busulfan	Radioactive phosphorus was listed in the final scope but excluded in the submission as clinical feedback indicated that this is no longer used in the UK (CS Table 1)	The EAG's clinical experts commented that hydroxycarbamide and IFN-alfa (or pegylated IFN-alfa) are the main comparators; the other therapies are used rarely if at all. The EAG agree with the exclusion of radioactive phosphorus (section 2.3.2)
Outcomes	The outcome measures to be considered include: • CHR (including reporting of HCT, WBC count and platelet count	Key outcomes are: • CHR including reporting of HCT, WBC count and platelet count separately • TTD	Not applicable	The company's outcomes are consistent with those specified in the NICE scope (NB the scope does not explicitly mention overall survival but it's

	separately) • TTD • mortality • symptom relief (including a reduction in spleen size, itching, fatigue and phlebotomy) • thrombosis • progression to AML or MF • adverse effects of treatment • HRQoL	• OS • symptom relief (including a reduction in spleen size, itching, fatigue and phlebotomy) • thrombosis • safety (including transformation to AML/MF and adverse events) • HRQoL		inclusion in the decision problem is appropriate). Mortality is not listed in the decision problem but is reported by the company trials and CS. Note that itching and fatigue are assessed by HRQoL instruments whilst thrombosis is reported as an adverse event.
Subgroups	People with and without splenomegaly	In line with final scope	Additional subgroup based on MAJIC-PV population (high-risk PV)	Each subgroup (with splenomegaly, without splenomegaly, and high-risk patients) is represented by a separate clinical trial.

Source: CS Table 1 with modifications. AML: acute myeloid leukaemia; BAT: best available therapy; CHR: complete haematological remission; HCT: haematocrit; HRQoL: health-related quality of life; IFN: interferon; MF: myelofibrosis; OS: overall survival; TTD: time to discontinuation; WBC: white blood cells

3 CLINICAL EFFECTIVENESS

3.1 Critique of the methods of review(s)

The company carried out a systematic literature review (SLR) that aimed to identify RCTs on the clinical efficacy and safety of any treatment in PV patients. The SLR was generally well-conducted and the EAG believe all relevant RCTs have been identified.

CS Appendix D.1.3 states the SLR identified eight unique clinical studies but only four are included in the submission. The four excluded studies had been identified according to the SLR eligibility criteria (CS Appendix D Table 8) which includes any intervention and any comparator and so the criteria are broader than both the NICE scope and the company decision problem. The reasons for exclusion are not given, but the EAG believe the studies were excluded appropriately:

- ARD12042:[17] a randomised phase 2 dose-finding study of fedratinib. This treatment is not a comparator.
- NCT00928707 (UCT1):[18] a randomised phase 2 dosing study of givinostat. This treatment is not a comparator.
- NCT00726232:[19] a randomised phase 2 dose-finding study of ruxolitinib. There was no best available therapy (BAT) arm therefore the evidence is inferior to the pivotal trials. Discussed in a footnote in CS section B.2.2.
- RELIEF (NCT01632904):[20 21] RCT for ruxolitinib versus hydroxycarbamide plus placebo. Discussed in CS section B.2.2 and excluded as the population was not resistant to or intolerant of hydroxycarbamide according to the modified ELN criteria. Study details are in CS Appendix D.1.3. The EAG note that the population "had been receiving a stable dose of hydroxycarbamide and were generally well controlled but still reported disease-associated symptoms". The EAG's clinical experts both agreed that the population in RELIEF is not reflective of patients resistant to or intolerant of hydroxycarbamide in the NHS PV population.

The SLR only searched for RCTs and indirect comparisons (referred to as matched-adjusted indirect comparisons, MAICs) but not observational studies or real-world evidence due to the use of an RCT study design filter in the searches. An indirect comparison comparing the ruxolitinib arm of RESPONSE against BAT data from a real-world registry (GEMFIN) was included and is used in the company's ITC (section 3.3). However, it is not transparent whether the GEMFIN registry is the only source of relevant comparator evidence suitable for

use in the company's ITC analysis (see section 3.3.2 for the critique of studies included in the ITC).

The three RCTs included by the company have been completed and are summarised below (section 3.2). Details of the EAG's full appraisal of the company SLR are provided in Appendix 9.1.

> **ERG conclusions on the methods of review**
>
> The company SLR appropriately identified all relevant RCTs. However, the way in which the GEMFIN registry study that informed the ITC was identified and selected is unclear, and no systematic search was conducted for other relevant observational studies.

3.2   Included studies

The three RCTs included in the CS are:

- RESPONSE: [22] a randomised comparison of ruxolitinib against BAT among patients with PV resistant or intolerant to hydroxycarbamide who had splenomegaly. Crossover from BAT to ruxolitinib occurred from week 32.
- RESPONSE-2:[23] a randomised comparison of ruxolitinib against BAT among patients with PV resistant or intolerant to hydroxycarbamide without palpable splenomegaly. Crossover from BAT to ruxolitinib occurred from week 28.
- MAJIC-PV[16] a randomised comparison of ruxolitinib against BAT among "high risk" patients with PV resistant or intolerant to hydroxycarbamide either with or without splenomegaly. Crossover was only permitted to the BAT arm (Table 5 below).

Primary clinical effectiveness analyses were conducted at weeks 32, 28 and 52 in the RESPONSE, RESPONSE-2 and MAJIC-PV trials respectively. In MAJIC-PV overall survival was reported up to 5 years after randomisation. Due to substantial crossover in RESPONSE and RESPONSE 2, long-term outcomes for these trials were reported primarily for the ruxolitinib arm excluding crossovers, up to 5 years. Extensive information on RESPONSE and RESPONSE-2 is available in the CS and in a series of clinical study reports (CSRs) provided for each main assessment timepoint in each trial (except the week 32 CSR which was not provided by the company). In contrast, relatively limited information on the MAJIC-PV trial is available, provided in the CS and an unpublished manuscript.[16]

33

3.2.1 Characteristics of the included studies

Details of the RCTs are reported for RESPONSE and RESPONSE-2 in CS section B.2.3.1, and for MAJIC-PV in CS section B.2.11.1, with further methodological details, including outcomes for all the trials in CS Appendix Table 11, CONSORT flow diagrams in CS Appendix D.2, and eligibility criteria in CS Appendix M.1. The main trial characteristics are summarised in Table 5 below.

Table 5 Summary characteristics of the included RCTs

Study characteristic	RESPONSE[22]	RESPONSE-2[23]	MAJIC-PV[16]
Funding	Company-sponsored	Company-sponsored	Investigator-led; funded by Leukaemia & Lymphoma Research (UK)
Study design	Open label phase 3 RCT: ruxolitinib vs BAT	Open label phase 3 RCT: ruxolitinib vs BAT	Open label phase 2 RCT: ruxolitinib vs BAT
Country	International, multi-centre 3 UK sites,[24] unknown number of UK patients	International, multi-centre No UK sites	UK-wide, multi-centre 38 UK sites
Population	Patients with polycythaemia vera R/I to HC [a] with splenomegaly	Patients with polycythaemia vera R/I to HC [a] without palpable splenomegaly	Patients with high-risk [b] polycythaemia vera R/I to HC [a] (with or without splenomegaly)
Randomisation	1:1; stratified according to resistance versus intolerance to HC	1:1; stratified according to resistance versus intolerance to HC	1:1; stratified according to gender
Number of participants	Ruxolitinib arm: n=110 BAT arm: n=112	Ruxolitinib arm: n=74 BAT arm: n=75	Ruxolitinib arm: n=93 BAT arm: n=87
Crossover	BAT arm only: patients failing to meet the primary outcome at week 32 were eligible to crossover to receive ruxolitinib	BAT arm only: patients failing to meet the primary outcome at week 28 were eligible to crossover to receive ruxolitinib	No crossover to the ruxolitinib arm was allowed. Ruxolitinib arm: if no response was observed at year 1 (primary outcome) patients changed to receive BAT
Duration	2010-2018; study is complete; data cut-off represent all patients who completed week 256 or discontinued according to protocol	2014-2020; study is complete; data cut-off represent all patients who completed week 260 or discontinued according to protocol	2012-2022; study is complete; data represent all 5 years of follow-up

BAT: best available therapy; HC: hydroxycarbamide; RCT: randomised controlled trial: R/I: resistant or intolerant; UK: United Kingdom.
[a] R/I to HC defined according to ELN consensus criteria,[10] described above in section 2.2.1.

The company trials RESPONSE and RESPONSE-2 are open label RCTs providing evidence for the indicated population split across two trials: one for patients with splenomegaly and one for patients without splenomegaly. However, crossover to the ruxolitinib arm was introduced early, after 32 weeks in RESPONSE and after 28 weeks in RESPONSE-2, which confounds longer-term results after the primary outcome analyses. Therefore, evidence from the MAJIC-PV trial, also an open label RCT, is used to inform hazard ratios for overall survival, overall survival in the BAT population, and several subgroup analyses. Data used in the economic model are outlined in CS section B.3.3 Table 21 and in section 3.2.4 of this report.

The RESPONSE and RESPONSE-2 trials followed the criteria for resistance/intolerance outlined in Table 3 above, with a minor exception relating to hydroxycarbamide dose (explained in clarification response A1). MAJIC-PV followed different "modified criteria" for resistance/intolerance (not separated) which are clearly listed in Table S1 of the trial manuscript[16] but lack an explanation for their source or selection. The MAJIC-PV criteria for resistance/intolerance appear to be stricter than the current (2022) guideline criteria reported in CS Table 4. However, as noted in section 2.3.1 above, definitions of hydroxycarbamide resistance/intolerance are not standardised in clinical trials or clinical practice.

The population in the MAJIC-PV trial is a broadly defined high-risk population compared to high-risk as defined in the BSH guidelines ($\geq$65 and/or prior thrombosis – as outlined in CS Figure 3 of the treatment pathway)[4]. In MAJIC-PV the age threshold is lowered to $\geq$60 and additional criteria can also indicate high-risk including significant or symptomatic splenomegaly, platelet count >1000 x 10^9/L, diabetes or hypertension requiring pharmacological therapy for >six months.[16] It is not obvious from the trials' baseline characteristics (Appendix 9.2 of this report) that the MAJIC-PV population is higher-risk than those included in the RESPONSE and RESPONSE-2 trials, as there is overlap of median age, % with prior thrombosis, median platelet counts and other characteristics between trials. However, the mortality rate was substantially higher in MAJIC-PV than the other trials (section 3.2.8 below), which is consistent with the population being at higher risk.

As MAJIC-PV includes patients with and without splenomegaly it covers more of the population in the licensed indication than either of the RESPONSE or RESPONSE-2 trials individually. Additionally, the MAJIC-PV trial contributes a wholly UK population, and with more stringent outcomes (outcomes assessment section 3.2.4), that is relevant to NHS

clinical practice compared to the company trials where only the RESPONSE trial has three UK sites and an unknown number of UK participants. CS section B.2.11.1 argues that the MAJIC-PV trial population is anticipated to represent the majority of patients with PV who are resistant to or intolerant of hydroxycarbamide which the EAG and our clinical experts agree is reasonable.

Limitations

The three included RCTs are limited by being open label (discussed in the risk of bias section of this report, section 3.2.3). The RESPONSE and RESPONSE-2 trials are limited by early crossover, however the MAJIC-PV trial should provide sufficient unconfounded evidence for longer-term outcomes. There is limited data available for the MAJIC-PV trial as it has only recently completed. There is no clinical study report or statistical analysis plan available for verification of study details or results in MAJIC-PV (clarification response A5), and individual level patient data could not be made available to the company because it was an investigator-led trial.

3.2.2   Patients' baseline characteristics in the included RCTs

Patients' baseline characteristics for RESPONSE and RESPONSE-2 are reported together in CS Table 7, and for MAJIC-PV in CS Appendix M.2.1. The EAG have combined key patient baseline characteristics from all three trials in Appendix 9.2 of this report.

Patient characteristics are similar for the RESPONSE and RESPONSE-2 trials, with the exception that participants in RESPONSE-2 did not have splenomegaly according to the trial eligibility criteria.

MAJIC-PV participants are slightly older on average than those in the company trials, but the age range is the same. The proportion of males, ECOG performance status, and percentage haematocrit (HCT) level, are similar. The MAJIC-PV BAT arm had more participants who had a prior thromboembolic event than in the company trials although the proportion of prior thromboembolic events in the ruxolitinib arm is similar to the company trials. Some characteristics in the MAJIC-PV trial are reported differently to the way in which they are reported in the two company trials, such as for white blood cell and platelet counts, JAK2 mutation status, including an extra category for patients who are both resistant *and* intolerant, and spleen size is measured differently, which makes it difficult to compare them with the characteristics in the company trials.

The EAG's clinical experts agreed that the patients' baseline characteristics in all the included trials are generally reflective of patients with PV who are resistant to or intolerant of hydroxycarbamide in the UK. However, the experts noted the following exceptions:

- The median age in MAJIC-PV is slightly higher than in the RESPONSE and RESPONSE-2 trials and is probably more reflective of that seen in clinical practice, although there is heterogeneity both in the trials and in practice.
- One clinical expert expected 15- 20% of patients would have had a prior PV-related thromboembolic event whereas the frequencies in the trials were higher than this (Appendix 9.2). There is also an imbalance within the MAJIC-PV trial for one of the indicators of high-risk for PV (proportion of patients who had a prior thromboembolic event) where the BAT arm is ▇ at risk than the ruxolitinib arm.

EAG conclusions on the included RCTs

All relevant RCTs (n=3) are included in the CS, with each containing up to five years of data from relevant populations, and all are complete. The trials reflect different subgroups of the licensed indication (patients with or without splenomegaly, or a combination). The MAJIC-PV trial is most likely to reflect UK clinical practice and is not confounded by crossover to the ruxolitinib arm, although the data available from the trial are limited.

3.2.3 Risk of bias assessment

Company and EAG risk of bias assessments for the RESPONSE, RESPONSE-2 and MAJIC-PV trials are shown in Appendix 9.3.

All three trials were judged by both the company and EAG to be at high risk of one or more types of bias.

Patient care, recording of outcomes, especially patient reported outcomes which involve subjective judgements, and analysis of outcomes could have been influenced by patients' and investigators' knowledge of the treatment allocation groups, due to the open-label designs of the trials. Additionally, some HRQoL outcomes including the MPN-SAF TSS were reported without any indication of sample sizes and variances. Analyses of HRQoL outcomes excluded missing data but did not specify the amount of missing data and/or reasons for data being missing.

In MAJIC-PV the randomisation process is unclear and the open-label trial design may have ███████████████████████████████████ (Figure S2 in the draft trial manuscript[16] shows that some patients ████████████████████████ after randomisation). In all trials there appears to have been selective reporting of HRQoL outcomes (including protocol-specified EQ-5D results ████████████ for MAJIC-PV). For further details see Appendix 9.3.

After weeks 32 and 28 respectively, outcomes in the RESPONSE and RESPONSE-2 trials would be confounded by crossover if analysed according to the originally randomised ruxolitinib and BAT groups. This confounding is acknowledged by the company: following crossover, the trial results are generally reported in the CS as single cohorts (the originally-randomised ruxolitinib arm, and the crossover cohort), rather than parallel randomised arms, which is appropriate. The comparative evidence for ruxolitinib versus BAT is limited to 32 and 28 weeks respectively in these trials.

Longer-term comparative evidence is available from the MAJIC-PV trial (52 weeks) which was not subject to crossover from BAT to ruxolitinib, although crossover from ruxolitinib to BAT was permitted for patients who did not achieve a complete or partial response of the primary outcome after 1 year. However, crossovers are not reported transparently: (i) The timing of crossovers from ruxolitinib to BAT is not reported (the EAG assume all occurred after 1 year as per the trial protocol, but reasons for crossover in Figure S2 of the draft manuscript included non-compliance, and hydroxycarbamide resistance and toxicity, which would seem unlikely to obey a 1-year assessment timescale. (ii) The draft trial manuscript states that

███
████████ (Table S2 of the draft manuscript). The CS and draft trial manuscript do not discuss the implications of the crossovers to the BAT arm or the receipt of ruxolitinib on the BAT arm. It is unclear whether the patients in question would have had a better or worse prognosis than the other patients in each arm and hence the risk of bias associated with these two aspects of participant flow is unclear. The draft trial manuscript[16] states that supporting analyses were performed censoring at the time the BAT patients began ruxolitinib and these analyses did not affect the conclusions from the modified ITT analysis. However, results of these analyses are not reported.

A consequence of all three trials being at high risk of bias is that uncertainty around the outcomes is not fully captured in the variance measures such as 95% confidence intervals, where reported.

EAG conclusions on risk of bias assessment

Overall, the EAG consider the trials to be at high risk of bias due to the open-label nature of all three trials, potential imbalances between groups in the RESPONSE and RESPONSE-2 trials after crossover at 32 and 28 weeks respectively, selective reporting of HRQoL outcomes, and the handling of missing data for HRQoL outcomes in all trials. For MAJIC-PV there is additionally a lack of clarity around the randomisation process, there are some differences in patient characteristics between the treatment arms, and the implications of crossovers from ruxolitinib to BAT, and of receipt of ruxolitinib by some patients in the BAT arm, are not fully clear.

3.2.4 Outcomes assessment

A large number of outcomes was assessed in the included trials (listed in CS Appendix Table 11), and these are reported in various degrees of detail in the CS, CS Appendices, trial publications and, for the RESPONSE and RESPONSE-2 trials, also in several CSRs provided by the company for different assessment timepoints. We have prioritised those outcomes relevant to the NICE scope and decision problem as summarised in Table 6. The outcomes are briefly explained in the sections below.

Table 6 Summary of the outcomes presented in this report

Outcome type	Summary	Where results reported
Primary trial outcomes (see section 3.2.4.1 below)	RESPONSE trial: HCT control & spleen size reduction (composite outcome) at week 32	Section 3.2.6.1
	RESPONSE-2 trial: HCT control (assessed as absence of phlebotomy ineligibility) at week 28	Section 3.2.6.2
	MAJIC-PV trial: Complete haematological remission (ELN criteria) (composite outcome) at 1 year	Section 3.2.6.3
Key secondary trial outcomes (see section 3.2.4.2 below)	Two "key" secondary outcomes were specified by the company: complete haematological remission in RESPONSE and RESPONSE-2; and durability of the primary outcome of RESPONSE beyond week 32	Section 3.2.6.4
Individual components of the primary outcomes	HCT level	Section **Error! Reference source not found.**
	Phlebotomy ineligibility	Section 3.2.6.6
	Spleen size	Section 3.2.6.7

39

Survival outcomes	Overall survival is a key outcome for the economic analysis (other survival outcomes are also presented where reported)	Section 3.2.6.8
HRQoL outcomes	Numerous measures are reported in the trials; we have prioritised the EQ-5D, MPN-SAF, EORTC QLQ-C30 and PSIS as explained in section 3.2.4.3 below	Section 3.2.6.9
Safety outcomes	Safety outcomes specified in the decision problem and identified as important by the EAG's clinical experts are presented where reported (section 3.2.4.4 below)	Section 3.2.8
ELN: EuropeanLeukemiaNet; HCT: haematocrit. Abbreviations for HRQoL instruments are explained in **Error! Reference source not found.** below.		

3.2.4.1 Primary efficacy outcomes

The primary efficacy outcomes do not inform the economic model but are important to demonstrate clinical efficacy.

HCT control. This is a key target of therapy for PV. HCT control can be measured directly as the haematocrit per volume of blood (target <45%) or indirectly via measures of phlebotomy, such as phlebotomy ineligibility (or absence of phlebotomy eligibility) which are indicative of adequate HCT control. The primary outcomes of the trials either assessed HCT control alone (RESPONSE-2) or included HCT control as a part of broader composite outcomes (RESPONSE, MAJIC-PV). HCT control was also included as a separate secondary outcome in RESPONSE-2 and MAJIC-PV.

The primary outcome of RESPONSE-2 was the proportion of patients achieving HCT control at 28 weeks, measured (according to ELN criteria) as absence of phlebotomy eligibility, where phlebotomy eligibility is defined as HCT of >45% that was at least three percentage points higher than baseline, or an HCT of >48%, whichever was lower.

HCT control and spleen size reduction. This was the composite primary outcome of RESPONSE, assessed at 32 weeks and defined as the proportion of patients achieving HCT control according to modified ELN response criteria (as above for RESPONSE-2) and a $\geq$35% reduction in spleen size. HCT control and spleen size were also reported as separate secondary outcomes. The EAG's clinical experts noted that assessment of spleen volume (i.e. using imaging techniques rather than palpation) is not very practical and not always assessed in practice.

40

Complete haematological remission (CHR) according to ELN criteria. This was the composite primary outcome of MAJIC-PV, assessed at one year and defined as the proportion of patients achieving all of the following: HCT <45% without phlebotomy for 3 months; platelets ≤400 × 10^9/L; WBC count ≤10 × 10^9/L, and normal spleen size. It requires fulfilment of all the ELN criteria for complete clinico-haematological response (CLHR) except for resolution of disease-related symptoms[25] and is therefore the most stringent primary outcome reported across the trials. CHR is clinically meaningful to report but it is not used in the economic model.

There is little evidence that stringent achievement of the ELN criteria contributes to improved outcomes apart from the HCT target,[4 26] and one of the EAG's clinical experts said that absence of phlebotomy, by aiming to maintain HCT levels below 45%, is the most critical outcome. Therefore, although the RESPONSE and RESPONSE-2 trials use less stringent combinations of criteria than MAJIC-PV, each primary outcome fulfils the most important aspect of the minimum reported criteria for response, i.e. HCT control.

3.2.4.2 Secondary efficacy outcomes

Complete haematological remission (CHR) is another composite outcome, considered a key secondary outcome in the RESPONSE and RESPONSE-2 trials. It comprises the modified ELN HCT control criteria, platelet counts and WBC counts. NB the definition of CHR in the RESPONSE and RESPONSE-2 trials differs from the CHR definition for the primary outcome in the MAJIC-PV trial mentioned above (which uses original ELN criteria for HCT control and includes spleen size).

The NICE scope indicates that WBC and platelet counts should be considered for reporting separately. These are included as haematological events in CS Appendix F and are taken into account in the summary of safety (section 3.2.8).

Survival outcomes. Overall survival at 5 years, reported in all three trials, is a secondary outcome informing the economic analysis. Transformation-free survival was also reported in RESPONSE and RESPONSE-2. Other survival outcomes, including progression-free survival and event-free survival, were reported for MAJIC-PV, but as hazard ratios for the ruxolitinib comparison rather than median point estimates.

3.2.4.3 HRQoL outcomes

The wide range of HRQoL measures used in the trials is summarised in Table 7 below. Results are reported in section 3.2.6.9 of this report for those measures highlighted in bold: EQ-5D (from RESPONSE-2), MPN-SAF (from all trials), EORTC-QLQ-C30 (from RESPONSE) and PSIS (from RESPONSE and RESPONSE-2). These HRQoL measures have been prioritised by the EAG as they inform the economic analysis and/or were considered clinically relevant by the EAG's experts. Full names of these instruments are given in Table 7 below.

EQ-5D data from RESPONSE-2 are used in a scenario analysis in the economic model (discussed further in section 4.2.7.2 below).

MPN-SAF and **EORTC QLQ-C30** results from RESPONSE are used in the economic model base case (see section 4.2.7.2 below), mapped to MF-SAF using assumptions validated by clinical experts advising the company, to form MF-8D utility values (a preference-based measure for myelofibrosis) (CS section B.3.4).

MPN-SAF is a myeloproliferative disease-specific instrument which has three versions reported in the trials (Table 7): MPN-SAF, MPN-SAF TSS (total symptom score) and MPN-10 (10 item version). These instruments have all been validated for mixed populations with myeloproliferative diseases that include PV.[27][28] The EAG's clinical experts confirmed that the MPN-10 is the version most used in clinical practice, and it includes dimensions for fatigue and itching. All trials measured the proportion of patients achieving ≥50% reduction in total symptom score which the EAG's clinical experts confirmed is a clinically meaningful change.

PSIS: This symptom-specific instrument assesses itching which is a bothersome symptom for many patients with PV. PSIS does not inform the economic analysis. The EAG have reported this outcome alongside the other HRQoL instruments to illustrate the effect of ruxolitinib at controlling PV symptoms. However, the company do not explain whether the PSIS has been validated or what the minimum clinically important change is for this instrument.

Table 7 HRQoL outcomes for the RESPONSE, RESPONSE-2, and MAJIC-PV trials

Source of PROs	RESPONSE[29]	RESPONSE-2[30]	MAJIC-PV[31]
PROs reported in the CS	**MPN-SAF** [a] at Week 32 **EORTC QLQ-C30** [a] at Week 32 and Week 80 and Week 256 PSIS at Week 32 and Week 256 PGIC at Week 4 and Week 32	Change from baseline to Week 28 for **MPN-SAF TSS**, **EQ-5D-5L**, PSIS and PGIC	**MPN-SAF TSS** over 5 years
PROs specified in the protocol	As above, plus MPN-PAF (RESPONSE Protocol section 6.2.4.1)	As above, plus WPAI (RESPONSE-2 Protocol section 10.5.5)	MPN-SAF, MDASI and EQ-5D (MAJIC Protocol section 8)
PROs listed in CS Appendix Table 11	As above for 'PROs reported in the CS', plus ECOG score.	As above for 'PROs reported in the CS', plus WPAI.	As above for 'PRO specified in the protocol', with different terminology: MPN10, MDASI and EQ-5D

Sources: CS section B.2.7; CS section B.2.11.2; CS Appendix Table 11; RESPONSE protocol; RESPONSE-2 protocol; MAJIC protocol.
ECOG: Eastern Cooperative Oncology Group Performance Status Scale; EORTC QLQ-C30: European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire; MDASI: MD Anderson Symptom Inventory; MPN-PAF: Myeloproliferative Neoplasm Pruritus Assessment Form; MPN-SAF: Myeloproliferative Neoplasm Symptom Assessment Form; MPN-SAF TSS: Myeloproliferative Neoplasm Symptom Assessment Form Total Symptom Score (abridged MPN-SAF with 11 factors); MPN-10: abridged MPN-SAF TSS with 10 factors; PGIC: Patient Global Impression of Change; PRO: patient reported outcome; PSIS: Pruritis Symptom Impact Scale; WPAI: Work Productivity And Impairment.
[a] Three dimensions from EORTC QLQ-C30 and five dimensions from MPN-SAF (mapped to MF-SAF) were combined to form MF-8D utility values; MF-8D was not measured in the trials.

As noted in the risk of bias section (section 3.2.3), there appears to be selective reporting among the HRQoL outcomes:

- There are several HRQoL outcomes specified in the trial protocols for which results are not reported in the CS, Appendices, or trial publications (MPN-PAF and WPAI in RESPONSE and RESPONSE-2, and EQ-5D, MDASI and MPN-10 in MAJIC-PV) (Table 7). This might reflect selective reporting, particularly the lack of EQ-5D results for MAJIC-PV (though the remaining outcomes were considered less important by the EAG's clinical experts).

- It is unclear which MPN-SAF tool the MAJIC-PV trial used or if the terminology (MPN-SAF/MPN-SAF TSS/MPN-10) has been used interchangeably in MAJIC-PV.

43

3.2.4.4 Safety outcomes

The range of adverse events reported by the company (CS sections B.2.10 and B.2.11.3, and CS Appendix F) is appropriate. Adverse events of special interest are reported and relevant to PV (thromboembolic events, second malignancies, non-melanoma skin cancer, transformation to MF, and transformation to AML) (CS Table 16). Transformation to MF and transformation AML are outcomes in the NICE scope and are also reported as efficacy outcomes in CS sections B.2.7.1 and B.2.7.2 as transformation-free survival. The EAG's clinical experts agreed that malignancies, particularly non-melanoma skin cancer (NMSC) are important. One expert commented that there may also be risk of lymphoma from ruxolitinib treatment. Another expert emphasised that infections, particularly herpes zoster reactivation, are important due to the immunosuppressive characteristics of ruxolitinib.

The trials use different frequency thresholds making it difficult to compare the rates between trials: RESPONSE reports adverse events occurring at a rate of $\geq$5 per 100 patient-years; RESPONSE-2 reports adverse events occurring in $\geq$3% of patients adjusted for patient-year exposure; and MAJIC-PV reports descriptive proportional statistics (n, %) for adverse events occurring in $\geq$10% of patients. The trials report the number of adverse events occurring at different CTCAE (Common Terminology Criteria for Adverse Events)) grades differently: MAJIC-PV reports adverse events (except for infections and malignancies) for all grades for the ruxolitinib and BAT arms combined, and Grades 3, 4 and 5 are reported separately, whereas the RESPONSE and RESPONSE-2 trials report adverse events for any grade for each arm, and Grades 3-4 are combined.

> **EAG conclusions on outcomes assessment**
>
> All reported outcomes are relevant to the disease, particularly HCT control for clinical effectiveness and the reporting of relevant adverse events of specific interest. Some outcomes are reported inconsistently across the trials, e.g. different complete haematological response outcomes, and thresholds for reporting of adverse events differed between trials. A wide range of HRQoL measures were used but reporting appears to be selective.

3.2.5 Statistical methods of the included studies

The CS reports statistical methods only for the primary outcomes. A summary of the EAG's assessment of statistical methods in the trials is provided in Table 8, with information for

44

secondary and other outcomes sourced from the trial protocols, CSRs and publications. The full assessment is provided in Appendix 9.4.

Table 8 Statistical methods of the RESPONSE, RESPONSE-2 and MAJIC-PV trials

	RESPONSE	RESPONSE-2	MAJIC-PV
Analysis populations	Appropriate for the primary and two key secondary outcomes (full analysis set), and safety outcomes (safety set). Unclear for the remaining secondary outcomes and HRQoL measures.	Appropriate for the primary and key secondary outcomes (full analysis set), and safety outcomes (safety set). Unclear for the remaining secondary outcomes and HRQoL measures.	Limited details of the analysis populations are reported; analysis populations for HRQoL outcomes are unclear. Potential for bias due to unaccounted for missing data (see Appendix 9.3).
Sample size and power calculations	Trial appears to be adequately powered for the primary outcome and probably also the two key secondary outcomes. Adequacy of the sample size for detecting treatment effects in the remaining secondary outcomes is uncertain.	Trial appears to be adequately powered for the primary outcome and key secondary outcome. Adequacy of the sample size for detecting treatment effects in the remaining secondary outcomes is uncertain.	Trial appears to be adequately powered for the primary outcome. Adequacy of the sample size for detecting treatment effects in the remaining secondary outcomes is uncertain.
Methods to account for multiplicity	The type I error control procedure is appropriate but only three outcomes are included. The likelihood of type I error in testing the remaining secondary outcomes is uncertain.	The type I error control procedure is appropriate but only two outcomes are included. The likelihood of type I error in testing the remaining secondary outcomes is uncertain.	No information available. The likelihood of nonsignificant treatment effects being declared significant is uncertain. Reliance on the statistical test results alone for inference is therefore inadvisable.
Analysis of outcomes	The statistical methods appear generally appropriate. The CS does not state whether the analyses were checked or validated.	The statistical methods appear generally appropriate. The CS does not state whether the analyses were checked or validated.	The statistical methods appear generally appropriate. NB alpha ▉ and ▉ confidence intervals are applied for the primary outcome (stated in the trial protocol) giving a ▉▉▉▉ chance of nonsignificant findings being declared significant. No information on whether analyses were checked.
Handling of missing data	Appropriate for primary and secondary	Appropriate for primary and secondary	Overall missing data were not accounted for,

	outcomes. Missing data were not accounted for in analyses of HRQoL and other exploratory outcomes. Number and reasons for missing data not fully reported.	outcomes. Missing data were not accounted for in analyses of HRQoL and other exploratory outcomes. Number and reasons for missing data not fully reported.	and the amount of missing data and reasons for data being missing were not reported.
Subgroup analyses	The pre-specified subgroup analysis method is appropriate. A post-hoc subgroup analysis of patients who received interferon-alfa, pooled from RESPONSE and RESPONSE-2, had small sample sizes ranging from 13 to 30 participants.	The pre-specified subgroup analysis method is appropriate. A post-hoc subgroup analysis of patients who received interferon-alfa, pooled from RESPONSE and RESPONSE-2, had small sample sizes ranging from 13 to 30 participants.	No subgroup analysis method or results are reported.

EAG conclusions on study statistical methods

The primary and key secondary outcomes of RESPONSE and RESPONSE-2 were adequately powered and accounted for multiple testing; however, remaining outcomes were mainly summarised descriptively and could be subject to type I errors. Missing data and multiple testing were not adequately accounted for in the MAJIC-PV trial so the results should be interpreted with caution. Where reported (RESPONSE and RESPONSE-2), subgroup analyses were appropriate but in some cases subject to small sample sizes.

3.2.6 Efficacy results of the intervention studies

As noted in section 3.2.4, many outcomes were assessed in the included trials. We have prioritised the following outcomes in this report, as explained above (Table 6).

3.2.6.1 Primary outcome in RESPONSE (composite of phlebotomy ineligibility and spleen volume reduction)

HCT control as defined by phlebotomy ineligibility and reduction of ≥35% in spleen volume from baseline at week 32 was the primary outcome in the RESPONSE trial and is referred to as the "primary response". The odds of achieving the primary response at week 32 statistically favoured ruxolitinib over BAT (odds ratio >1.0). However, the majority of patients did not achieve a primary response (Table 9). Due to crossover, results after week 32 are reported for the randomised ruxolitinib arm of the trial, i.e. a single non-comparative cohort.

Of those originally randomised to ruxolitinib who achieved a primary response at week 32, nearly all had maintained the key secondary outcome of response at week 48. The estimated probability of maintaining the primary response from week 32 to week 256 in the ruxolitinib arm (a secondary outcome) was 74% but with a relatively wide 95% confidence interval (51% to 88%).

Table 9 Primary outcome in the RESPONSE trial

Outcome	Ruxolitinib	BAT	Difference	Source
Primary response at week 32 (primary outcome)	23/110; (20.9%) [a] 25/110 (22.7%) [b]	1/112 (0.9%)	20.02 (95% CI 12.22 to 27.82) p<0.001 OR 28.6 (95% CI 4.5−1206)	CS section B.2.7.1 and Table 11-5 in week 48 CSR
Secondary outcomes related to the primary outcome				
Durable primary response (response at week 32 maintained at week 48)	21/110 (19.1%)	1/112 (0.9%)	18.2 %−points; [c] p<0.001	CS section B.2.7.1
Probability of maintaining primary response for ≥1 year	94%	NA	NA	CS Figure 8
Probability of maintaining primary response for ≥80 weeks	92% (ITT) [d] 89% [d]	NA	NA	CS Appendix M.3.1
KM estimated probability of maintaining primary response at 208 weeks	73% (95% CI 49%−87%)	NA	NA	CS Appendix M.3.1
KM estimated probability of maintaining primary response from week 32 for 224 weeks	74% (95% CI 51% to 88%)	NA	NA	CS section B.2.7.1
Median duration of primary response	Not reached	Not reached	NA	CS section B.2.7.1
ITT: intention to treat population; KM: Kaplan-Meier; NA: Not applicable (due to patient crossover); OR: odds ratio. [a] Initial results reported by Vannucchi et al. 2015; [b] updated results from week 80 analysis reported in CS section B.2.7.1 which identified 2 further week 32 responders; [c] calculated by reviewer; [d] ITT population includes crossovers; 89% refers to patients randomised to ruxolitinib.				

3.2.6.2 Primary outcome in RESPONSE-2 (absence of phlebotomy eligibility)

HCT control as defined by phlebotomy ineligibility at week 28 was the primary outcome of the RESPONSE-2 trial. The trial did not include patients with palpable splenomegaly and so the primary outcome for RESPONSE-2 does not include spleen size. The odds of achieving HCT control at week 28 statistically favoured ruxolitinib over BAT (odds ratio >1.0). In the ruxolitinib arm 62% of patients achieved the primary outcome, compared to 19% in the BAT

47

arm. Due to crossover, results after week 28 are reported for the randomised ruxolitinib arm of the trial, i.e. a single non-comparative cohort (secondary outcomes). Among the patients randomised to the ruxolitinib arm, 21.6% had achieved durable HCT control to week 260 (Table 10).

HCT control as defined by the absence of phlebotomy was also assessed in the RESPONSE trial, as a secondary outcome, and shows a similar picture to that of RESPONSE-2: Of those who received ruxolitinib in RESPONSE, 60.0% achieved HCT control after 24 weeks' treatment (at the week 32 analysis) compared to 19.6% in the BAT arm.[22] The proportion in the ruxolitinib arm with durable HCT control was not reported for the RESPONSE trial, but the estimated probability of maintaining HCT control from week 32 to week 256 was 73% (95% CI 60% to 83%).[24] The median duration of HCT control was not reached in either trial (CS Appendix M.3.1 and M.3.2).

Table 10 Primary outcome in the RESPONSE-2 trial

Outcome	Ruxolitinib (N=74)	BAT (N=75)	Difference	Source
HCT control at week 28 (primary outcome)	46/74 (62%)	14/75 (19%)	OR 7.28 (95% CI 3.43 to 15.45); p<0.0001	CS section B.2.7.2
Secondary outcomes related to the primary outcome				
Proportion maintaining HCT control from week 28 to 52	▉	▉	OR ▉ (95% CI ▉) P<0.0001	Table 11-2 in week 80 CSR
Proportion maintaining HCT control from week 28 to 80	35/74 (47.3%)	2/75 (2.7%)	44.6 %−points [a] OR ▉ (95% CI ▉)	CS Appendix M.3.2 Week 80 CSR
Durable HCT control at week 156	30/74 (40.5%) [b]	NA	NA	CS Appendix M.3.2
Durable HCT control at 5 years (week 260)	16/74 (21.6%)	NA	NA	CS section B.2.7.2
NA: not applicable; OR: odds ratio. [a] calculated by reviewer; [b] patients originally randomised to ruxolitinib (i.e. excluding crossovers)				

3.2.6.3 Primary outcome in MAJIC-PV (composite of HCT control, WBC, platelet, and spleen volume thresholds by ELN criteria)

The primary outcome in MAJIC-PV, referred to as "complete haematological remission" according to ELN criteria[25] is a composite of HCT control [comprising HCT <45% with phlebotomy ineligibility], WBC counts, platelet counts, and spleen volume thresholds. The odds of achieving complete haematological remission at 1 year statistically favoured ruxolitinib over BAT (odds ratio >1.0), although fewer than half the patients receiving

48

ruxolitinib achieved a complete remission (Table). Nearly all of those who did not achieve a compete haematological remission at year 1 achieved a partial haematological remission, giving high overall response rates in both the ruxolitinib and BAT groups.

Table 11 Primary outcome in the MAJIC-PV trial (complete haematological remission)

Outcome	Ruxolitinib (N=93)	BAT (N=87)	Difference	Source
Proportion with complete haematological remission (ELN criteria) in year 1	▮/93 (▮)	▮/87 (▮)	▬▬▬▬▬▬▬▬	CS section B.2.11.2 and unpublished trial manuscript[16]
Secondary outcomes related to the primary outcome				
Proportion with partial haematological remission (ELN criteria) in year 1	▮/93 (▮)	▮/87 (▮)	▮ %−points [b]	
Overall response rate in year 1	▮	▮	▮%−points [b]	
OR: odds ratio; ELN: European EukemiaNet. [a] adjusted for gender. [b] calculated by reviewer.				

3.2.6.4 Key secondary outcomes

Complete haematological remission (composite of HCT control assessed as phlebotomy ineligibility; together with WBC and platelet count thresholds) was specified as a key secondary outcome in the RESPONSE and RESPONSE-2 trials. Note that this outcome differs from the complete haematological remission outcome of the MAJIC-PV trial reported above (which used ELN criteria that include a more stringent definition of HCT control [HCT <45% without phlebotomy] and a normal spleen size). In both trials the proportion achieving complete haematological remission statistically favoured the ruxolitinib arm after weeks 28 and 32, but was relatively low, not exceeding 24% (Table 12). Median duration of complete haematological remission was not reached in the RESPONSE trial (CS section B.2.7.1). In RESPONSE-2 the KM estimate of median duration of complete haematological remission from week 28 to week 260 (i.e. 5 years) was 34.0 weeks (95% CI 16 to 78 weeks) (CS section B.2.7.2).

Table 12 Complete haematological remission in the RESPONSE and RESPONSE-2 trials

Outcome	Ruxolitinib	BAT	Difference	Source
Proportion achieving CHR at week 32 in RESPONSE	26 [a] /110 (23.6%)	8 [a] /112 (8.9%)	14.7 %−points [a] p=0.003 [b]	CS section B.2.7.1

49

| Proportion achieving CHR at week 28 in RESPONSE-2 | 17/74 (23%) | 4/75 (5%) | OR 5.58 (95% CI 1.73 to 17.99); p<0.0019 | CS section B.2.7.2 Week 28 CSR |

CHR: complete haematological remission; OR: odds ratio; [a] calculated by reviewer; [b] Vannucchi et al. 2015[22] report p=0.003, CS reports p=0.0003

Durability of the primary outcome (HCT control and spleen volume reduction) at week 48 in the ruxolitinib arm was specified as a key secondary outcome in the RESPONSE trial. This is reported alongside the primary outcome in Table 9 above.

3.2.6.5 HCT measurements

HCT control is included as a component of the primary outcomes of all three included RCTs (sections 3.2.6.1 to 3.2.6.3). HCT levels are also reported separately in RESPONSE-2 and in MAJIC-PV.

In RESPONSE-2 the baseline and week 28 HCT levels were below the HCT control threshold of <45% for PV. At week 28 the HCT level had decreased in the ruxolitinib arm and increased in the BAT arm, confirming the cytoreductive action of ruxolitinib (Table 13).

Table 13 HCT levels in the RESPONSE-2 trial

Outcome	Ruxolitinib	BAT	Difference	Source
Baseline HCT, mean (SD)	42.8% (1.5%)	42.7% (1.4%)	0.1 %−points [a]	CS section B.2.7.2
Week 28 HCT, mean (SD)	40.2% (4.1%)	44.9% (3.8%)	−4.7 %−points [a]	
Change in HCT from baseline to week 28, mean (SD)	−2.6% [a]	2.2% [a]	4.8 %−points [a]	
[a] calculated by reviewer				

In MAJIC-PV, HCT levels in the ruxolitinib and BAT arms are shown visually in the supplement to the unpublished manuscript (Figure S4 in Harrison et al.[16]) over 54 weeks. Estimates of mean counts are not reported. Following randomisation, the mean HCT count in the ruxolitinib arm ████████████████████████████ whilst the HCT count in the BAT arm ██████████████████████, through the 54 weeks. These differences ███████████████████████████████████.

3.2.6.6 Phlebotomy rates

The trials reported the proportions of patients who underwent different numbers of phlebotomy procedures, as well at the proportions who had any or no phlebotomies. Here we summarise the proportions who had no phlebotomies as this is an indicator of HCT control.

50

The proportion of patients who had no phlebotomies in the RESPONSE and RESPONSE-2 trials (before crossover) and MAJIC-PV trial was consistently higher in the ruxolitinib arm of each trial than in the BAT arm (Table 14).

As the data in Table 14 show, 34% to 48% of patients in the BAT arms (prior to crossover) did not require phlebotomy. Overall, ruxolitinib increased the proportion who did not require phlebotomy by 23 to 41 percentage points relative to BAT, depending on the trial and assessment time.

Over the 5-year follow-up period, the proportion without phlebotomies in the ruxolitinib arm (excluding crossovers in RESPONSE and RESPONSE-2) was:

- 83% during weeks 80-256 in RESPONSE (CS Figure 15)
- 69% up to week 260 in RESPONSE-2 (CS Table 11)
- ■ in MAJIC-PV (Table 14 below).

Table 14 Proportion without phlebotomy in the RCTs

Outcome	Ruxolitinib	BAT	Difference	Source
Proportion with no phlebotomies in weeks 8-32 in RESPONSE [a]	80/110 (72.7%) [b]	38/112 (33.9%) [b]	38.8 %−points [b]	CS Figure 9 [c]
Proportion with no phlebotomies up to week 28 in RESPONSE-2	81.1%	40%	41.1 %−points [b]	CS Figure 19 [d]
Proportion with no phlebotomies up to 5 years in MAJIC-PV [e]	■/93 (■) [b]	■/87 (■) [b]	■ %−points [b]	Unpublished trial manuscript[16]
[a] patients who did not discontinue randomised therapy prior to week 8; [b] calculated by reviewer; [c] CS Figure 9 reports sample sizes less than the full analysis set, EAG calculations use the full analysis set (i.e. ITT analysis); [d] CS Figure 19 does not report the sample size, so unclear whether this is an ITT analysis; [e] assessment time not reported but EAG assume this was 5 years (since adjacent outcomes in the trial manuscript supplementary appendix were reported for 5 years)				

3.2.6.7 Spleen measurements

Spleen size is included as a component of the primary composite outcome of the RESPONSE trial (section 3.2.6.1 above). Spleen measurements are also reported separately for RESPONSE, and some limited information on spleen size is also available for RESPONSE-2 (spleen volume measurements are not reported for MAJIC-PV[16]).

In RESPONSE, 40% of patients in the ruxolitinib arm and 0.9% in the BAT arm achieved a ≥35% reduction in spleen volume after 24 weeks of treatment (week 32 analysis) according to CS section 2.7.1, but the trial publication[22] and week 48 CSR report 38.2% in the

51

ruxolitinib arm; the EAG are unclear which is correct. In the ruxolitinib arm, excluding crossovers, the estimated probability of maintaining a ≥35% reduction in spleen volume from week 32 to week 224 was 72% (95% CI 34% to 91%).[24]

In RESPONSE-2, according to the week 260 CSR, nine patients in the ruxolitinib arm had a palpable spleen, with the mean palpable spleen length at week 260 being 0.10 cm. In the BAT arm, nine patients had a palpable spleen but very few patients were assessed (n=5) at week 80, and the mean palpable spleen length was 0 cm (data for the remaining four patients are not reported). These findings suggest splenomegaly during long-term follow up was negligible in RESPONSE-2.

3.2.6.8 Survival outcomes

Survival outcomes reported in CS and trial publications are summarised below. The MAJIC-PV trial manuscript reports that 3-year overall survival ▨▨▨▨▨ between the trial arms: ▨▨▨▨▨▨▨ for BAT and ▨▨▨▨▨▨ for ruxolitinib. Hazard ratios comparing 5-year overall survival for ruxolitinib against BAT are also provided (see below); it is unclear why the 3-year and 5-year outcomes are not reported consistently (CS section B.2.11.2).

Overall survival at 5 years: KM estimates of OS at 5 years are reported for the ruxolitinib arm, excluding crossovers, in the RESPONSE trial (N=110) and RESPONSE-2 trial (N=74), and as a hazard ratio for the comparison of ruxolitinib (N=93) versus BAT (N=87) in MAJIC-PV:

- RESPONSE: 91.9% (95% CI 84.4% to 95.9%) (CS section B.2.7.1); median OS not reached (not reported in the CS, publications or CSRs - stated in the company's Factual Accuracy Check document)
- RESPONSE-2: 96% (95% CI 87% to 99%); median OS not reached (CS section B.2.7.2)
- MAJIC-PV: Median OS not reached;[16] OS hazard ratio, ruxolitinib versus BAT ▨▨▨▨▨▨▨▨▨ (CS section B.2.11.2).

Transformation-free survival at 5 years: KM estimates of TFS at 5 years for the ruxolitinib arm, excluding crossovers, for the RESPONSE trial (N=110) and RESPONSE-2 trial (N=74) were:

- RESPONSE: ▨▨ (95% CI ▨▨▨▨) (CS section B.2.7.1)
- RESPONSE-2: 94% (95% CI 85% to 98%) (CS section B.2.7.2).

52

Other survival outcomes at 5 years: The following hazard ratios based on KM estimates of median survival outcomes for the ruxolitinib arm (N=93) compared against the BAT arm (N=87) are reported for the MAJIC-PV trial in the unpublished trial manuscript:[16]

- Progression-free survival: HR ███████████████████
- Event-free survival: HR ████████████████
- Major thrombosis event-free survival: HR ██████████████████
- Haemorrhagic event-free survival: HR ██████████████████████.

3.2.6.9 HRQoL outcomes

The trials reported a range of HRQoL measures (with some evidence of selective reporting) (see Table 7 and section 3.2.4.3 above). However, the EAG's clinical experts commented that many of the HRQoL measures are not used in clinical practice nor widely in trials. Below we have prioritised those HRQoL measures that inform the economic analysis (EQ-5D, EORTC QLQ-C30, MPN-SAF), or are relevant to symptoms specified in the NICE scope (PSIS is an itching-specific instrument whilst MPN-SAF includes itching and fatigue among other symptoms). The EAG's clinical experts commented that the MPN-SAF and its derivatives such as MPN-10 are the HRQoL measures most used in clinical practice.

EQ-5D index score

The EQ-5D is specified as an outcome in the RESPONSE-2 and MAJIC-PV trials (Table 7 above) but is only reported for RESPONSE-2.

The company have presented EQ-5D scores from RESPONSE-2 in their submission (CS Figure 21 and the trial publication[23]) but these are difficult to interpret due to: (i) the scores are reported as percentage classes instead of their original scale; (ii) sample sizes are unclear since the numerators and denominators for the percentages are not provided; (iii) the use of percentages excludes any information on the variance of scores. The EAG have instead sourced the overall EQ-5D-5L scores from the RESPONSE-2 week 260 CSR, summarised in Table 15 below. These data suggest there is little difference in the change from baseline between the ruxolitinib and BAT arms, and within the ruxolitinib arm after crossover occurred.

The company note that a large proportion of patients reported no problems in all five EQ-5D domains at baseline, and they argue that EQ-5D is unsuitable for measuring HRQoL in PV (CS section B.3.4.1) (discussed below in section 4.2.7.2). However, point estimates of EQ-

5D scores from RESPONSE-2 were used in a scenario analysis in the company's economic model (CS section B.3.4.1).

Table 15 Changes in EQ-5D-5L health index score in the RESPONSE-2 trial

Mean (SD) change from baseline [a]	Ruxolitinib (N=74)	BAT (N=75)	Difference [b]	Source
Week 28	██████████	██████████	████	Table 14.2-2.6
Week 52	██████████	██████████ [c]	████	in week 260
Week 80	██████████	NA	NA	CSR
Week 104	██████████	NA	NA	
Week 156	██████████	NA	NA	
Week 247	██████████	NA	NA	
NA: not applicable. [a] Baseline mean varied with each assessment timepoint, presumably because not all patients had baseline measurements at all timepoints [b] calculated by reviewer. [c] patients who did not cross over to ruxolitinib				

MPN-SAF scores

MPN-SAF scores inform the company's economic analysis indirectly, via conversion to MF-8D scores (section 4.2.7.2). All three trials reported changes in MPN-SAF scores, although the reporting format is different for each trial, making comparisons across the trials difficult. This outcome also has some uncertainty relating to missing data.

- **RESPONSE** (CS section 2.7.1): At week 32, the proportion with ≥50% reduction in MPN-SAF total score (a clinically meaningful improvement) was 49% in the ruxolitinib arm (36/74) and 5% in the BAT arm (4/81) The reported sample sizes indicate that the ruxolitinib arm had 36/110 (33%) missing data and the BAT arm had 31/112 (28%) missing data compared to the full analysis set.

- **RESPONSE-2** (CS section B.2.7.2, CS Figure 20 and CS Appendix M.3.2): At week 28, the proportion with ≥50% reduction in MPN-SAF TSS was 45.3% in the ruxolitinib arm and 22.7% in the BAT arm. Sample sizes reported were 64 ruxolitinib patients and 22 BAT patients, meaning that the ruxolitinib arm had 46/110 (42%) missing data and the BAT arm had 90/112 (80%) missing data compared to the full analysis set.

- **MAJIC-PV** (unpublished manuscript[16]): Only the mean difference in the change from baseline in MPN-10* between the ruxolitinib and BAT arms is reported, for a range of timepoints from month 2 to month 60. The difference favoured ruxolitinib over BAT at all timepoints and was statistically significant up to around 24 months but statistical significance should be interpreted cautiously due to the large number of comparisons made (Table S8 in the draft trial manuscript[16]). The mean difference for ruxolitinib versus

54

BAT at 60 months was − ███████████████. Sample sizes are not reported so the extent of missing data is unclear. (*NB the source table refers to "MPN-10" but the wording in the manuscript implies that this is synonymous with the MPN-SAF).

EORTC QLQ-C30

The EORTC QLQ-C30 measure was utilised only in the RESPONSE trial. Improvements from baseline occurred across all of the six subscales for the ruxolitinib arm, both at week 32 and (excluding crossovers) at week 256, whilst scores worsened slightly for five of the six subscales at week 32 in the BAT arm (Table 16). The threshold for a clinically meaningful change (10 points) was reached for the ruxolitinib arm at week 32; the largest improvement with ruxolitinib and the largest worsening with BAT were both for the Global health status subscale. Sample sizes and variance measures are not reported for this outcome.

Table 16 EORTC QLQ-C30 questionnaire functional and QoL scales in the RESPONSE trial

Mean change in score from baseline	Ruxolitinib		BAT Week 32	Difference at week 32 [a]	Source
Scale	Week 32	Week 256			
Global health status/QoL	10.86	9.49	−4.82	15.68	Vannucci et al. 2015;[22] Supplementary Figure 8 in Kiladjian et al. 2020[24]
Physical functioning	6.44	7.05	−1.51	7.95	
Role functioning	5.3	2.08	−0.41	5.71	
Emotional functioning	7.92	7.55	1.04	6.88	
Cognitive functioning	4.17	6.08	−3.33	7.50	
Social functioning	7.66	5.73	−0.42	8.08	
[a] calculated by reviewer; minimal clinically important difference is 10 points. NB variance estimates and sample sizes are not reported; results are for patients with both baseline and week 32 / 256 data					

Pruritis Impact Symptom Scale (PSIS)

The company report changes from baseline in PSIS scores for RESPONSE (CS Figure 11) and RESPONSE-2 (trial publication[23]). The severity of PV-related itching, the extent to which the patient was bothered by itching, and the extent to which the itching interfered with daily life were improved to a greater extent in the ruxolitinib arm than the BAT arm at 32 weeks, both for 24 hour and 7-day recall periods, in both trials. However, the sample size and variance estimates for this outcome are not reported in the CS or trial publications. It is also unclear whether this tool has been validated and what the minimum clinically important difference would be.

Overall, there is evidence that ruxolitinib improves patients' symptoms relating to itching, but with some uncertainty around how variable and clinically significant these findings are.

3.2.7 Subgroup analyses

The NICE scope specifies two subgroups: patients with and without splenomegaly. These subgroups are covered by the different trial populations: in the RESPONSE trial all patients had splenomegaly (based on imaging measurements), whilst the RESPONSE-2 trial excluded patients with splenomegaly (based on splenic palpation) (CS Table 6). Note that the MAJIC-PV trial included high-risk PV patients irrespective of splenomegaly and thus provides evidence from a further relevant population reflecting the mix of patients seen in clinical practice.

3.2.7.1 Pre-specified subgroups in the trials

The following subgroup analyses were conducted for the primary outcome in each trial. The subgroup analysis results reported in the CS and trial publications are consistent with those specified in the trial protocols for RESPONSE and RESPONSE-2. For MAJIC-PV the trial protocol specifies exploratory subgroups, but these are not reported in the CS or the trial draft manuscript.[16]

RESPONSE

Pre-specified subgroup comparisons (trial protocol section 9.4.4) were: baseline palpable splenomegaly (<10cm versus ≥10cm below the costal margin), sex (male versus female), age group (≤60 years versus >60 years), hydroxycarbamide intolerance or resistance, region (US versus non-US), race (White or Caucasian versus other) and ethnicity (Hispanic or Latino versus other).

A forest plot showing the odds of achieving the composite primary response outcome at week 32 for each subgroup is provided in CS Figure 24 but is missing odds ratios for one subgroup in each pair so the EAG are unable to interpret this (the week 32 CSR was not provided by the company).

RESPONSE-2

Pre-specified subgroup comparisons (trial protocol section 10.4.4) were: hydroxycarbamide intolerance or resistance, sex (male versus female), age group (≤60 years versus >60 years), risk category (0 risk factors versus 1-2 risk factors including age >60 and/or previous thromboembolism).

56

CS Figure 25 shows the odds of achieving complete haematological remission at week 28 for each of these subgroups. All odds ratios are greater than 5.0 and have overlapping confidence intervals, suggesting that the odds of achieving the primary outcome did not differ between subgroups.

3.2.7.2 Post-hoc subgroup analyses in the trials

CS Appendix E reports subgroup analyses of patients who had received prior interferon-alfa, interferon-alfa as BAT, or ruxolitinib after crossover from receiving interferon-alfa as BAT. These subgroup, which are based on data pooled from RESPONSE and RESPONSE-2, have small sample sizes ranging from 13 to 30 patients and therefore their generalisability is uncertain.

3.2.8 Safety results

Adverse events in the RESPONSE and RESPONSE-2 trials are reported in CS sections B.2.10.1 and B.2.10.2 up to weeks 256 and 260 respectively and in Appendix F for earlier data cuts. Adverse events in the MAJIC-PV trial are reported up to 5 years in CS section B.2.11.3, CS Appendix Table 20 and the unpublished trial manuscript.[16]

NB as noted in section 3.2.4.4, in the CS adverse events are not reported consistently in the same format across the trials.

Most frequent adverse events

In RESPONSE and RESPONSE-2 the most frequent adverse events of any grade were seen in the BAT arms, especially for the disease symptom pruritus (BAT 32.6 and 31.9 per 100 patient years respectively; ruxolitinib 7.0 and 3.6 per 100 patient years respectively; crossover 6.1 and 3.4 per 100 patient years respectively). Thrombocytopaenia of any grade also had the highest rate in the BAT arms (BAT 16.3 and 15.0 per 100 patient years respectively; ruxolitinib 4.4 and 1.5 per 100 patient years respectively; crossover 1.2 and 1.5 per 100 patient years respectively). The most frequent adverse event of any grade that occurred more often in the ruxolitinib and crossover groups than in the BAT arms was anaemia (ruxolitinib 8.9 and 8.1 per 100 patient years respectively; crossover 8.8 and 9.2 per 100 patient years respectively).

In the MAJIC-PV trial, the most frequent adverse events were ███████ ██████████████████████████████████.[16] The CS highlights Grade 3 anaemia which occurred in █ of ruxolitinib patients compared to █ of BAT patients.

Most frequent serious adverse events

In the RESPONSE trial the most frequent serious adverse event was pneumonia, but with similar rates across the trial arms (1.2 to 1.8 per exposure adjusted 100 patient years). Several serious adverse events were recorded only in the ruxolitinib and crossover groups but not the BAT group, notably squamous cell carcinoma, basal cell carcinoma, rectal haemorrhage, and herpes zoster infection. Adverse events which were classified as serious adverse events are not reported in the CS for the RESPONSE-2 trial, nor in the trial manuscript for the MAJIC-PV trial.[16]

Infections

In the RESPONSE trial, the total rate of infections per 100 patient years was highest in the BAT arm (BAT 59.8; ruxolitinib 18.9; crossover 19.1). The total rate of infections is not reported for the RESPONSE-2 trial, although individual infections are reported. For both RESPONSE and RESPONSE-2, the herpes zoster infection appears to have only occurred in the ruxolitinib arms and the crossover groups, although the adverse events in CS Tables 15 and 17 are not reported consistently across the trials and infrequent infections might not have been captured due to the reporting thresholds used in the tables (RESPONSE: ≥0.5 per 100 patient years; RESPONSE-2: for ≥3% of patients in any arm).

In the MAJIC-PV trial, infections were more common in the ruxolitinib arm (█ Grade 3/4 events) compared to the BAT arm (█ Grade 3/4 events). The most common infections for ruxolitinib patients were ███████████████ and ██████████████. Herpes zoster infections at any Grade occurred in █ ruxolitinib patients compared to █ BAT patients. All infections are individually reported in Table S9B of the unpublished trial manuscript.[16]

Malignancies, including transformation to MF or AML

In the RESPONSE trial, second malignancies had a higher exposure-adjusted rate per 100 patient years in the ruxolitinib arm (7.0) and crossover group (4.5) than in the BAT arm (4.1); so too did rates of non-melanoma skin cancer: ruxolitinib arm (5.1), crossover group (2.7) and BAT arm (2.7). Exposure-adjusted rates per 100 patient years of transformation for both MF and AML were also higher in the ruxolitinib arm and crossover group although with slightly lower rates than reported for the malignancies.

In the RESPONSE-2 trial, second malignancies are reported in the CSR but not the CS: the week 260 CSR states that there may be some data overlap with this category.[32] Non-melanoma skin cancer had a slightly higher rate of occurrence in the ruxolitinib arm and crossover group than in the BAT arm, but rates of transformation to MF and AML were slightly higher in the BAT arm. No patients transformed to AML in the ruxolitinib arm or crossover group.

In the MAJIC-PV trial, certain malignancies were more common in the ruxolitinib arm compared to the BAT arm: ███████████████████ occurred in █ versus █ patients respectively, and ███████████ occurred in █ versus █ patients respectively. ███████████████████ was more common in the BAT arm: █ ruxolitinib patients compared to █ BAT patients. Further malignancies are fully reported in Table S9B of the unpublished manuscript.[16]

Thromboembolic events

In the RESPONSE and RESPONSE-2 trials, thromboembolic events had the highest rates (exposure-adjusted per 100 patient years) in the BAT arms (8.2 and 3.7 respectively), compared to the ruxolitinib arms (1.2 and 1.5 respectively) and crossover groups (2.7 and 2.9 respectively).

In the MAJIC-PV trial, Table S7 in the unpublished manuscript reports the number, toxicity, and CTCAE grade of minor and major thrombotic events but does not distinguish between the ruxolitinib and BAT arms.[16]

Deaths

One out of a total of six deaths in the RESPONSE trial was suspected to be related to the study drug (gastric adenocarcinoma) and none of the five deaths in the RESPONSE-2 trial were deemed to be related to the study drug. More deaths occurred in the MAJIC-PV population (███). The EAG speculate this may be due to the slightly older population and a greater proportion of patients (in the BAT arm) who had had a prior thromboembolic event (Appendix 9.2) indicating high-risk. However, only █ death in each treatment arm in the MAJIC-PV trial was considered related to the study drug and ██ of the deaths were infection-related.

EAG conclusions on safety results

Adverse events are difficult to compare across the trials due to inconsistent reporting formats. Safety results appear to be broadly consistent across the trials, the biggest

59

difference between them being the number of deaths occurring in the MAJIC-PV trial, reflecting high-risk population characteristics. The incidence rates of anaemia, specific infections including herpes zoster and non-melanoma skin cancers,were higher in the ruxolitinib arms and crossover groups. Overall rates of infections varied, being highest in the BAT arm of RESPONSE and the ruxolitinib arm of MAJIC-PV (not reported in the CS for RESPONSE-2). Overall no new safety signals were observed.

3.2.9 Pairwise meta-analysis of intervention studies

No pairwise meta-analysis was conducted because the three trials included by the company each included a different population subgroup (people with splenomegaly in the RESPONSE trial, those without palpable splenomegaly in RESPONSE-2, and a high-risk subgroup with or without splenomegaly in MAJIC-PV). The trials also differed in other characteristics including the presence and timing of crossovers and timing of outcome assessments. The EAG agree that a pairwise meta-analysis was not appropriate.

3.3 Critique of studies included in the indirect treatment comparison (ITC)

The company conducted an indirect treatment comparison which they refer to as a MAIC (matched adjusted indirect comparison). MAIC is a misnomer since the company had individual patient data (IPD) available from both cohorts being compared and used these in a propensity score matching analysis (MAIC, in contrast, is applicable when IPD are available for only one of the cohorts being compared[33]). In this report we refer to the indirect comparison as an ITC.

3.3.1 Rationale for the ITC

The RESPONSE and RESPONSE-2 trials experienced early crossover of patients from the BAT arm to the ruxolitinib arm, from week 32 in RESPONSE and from week 28 in RESPONSE-2. Estimates of the effect of ruxolitinib on overall survival would therefore be confounded by crossover. Adjustment for crossover was not feasible due to the low frequency of deaths (only two on-treatment events at week 256 in RESPONSE; CS section 2.7.1). An ITC was conducted to estimate the effect of ruxolitinib on overall survival without confounding, by comparing long-term survival in the randomised ruxolitinib trial arm of RESPONSE against that in an external BAT cohort, using propensity score matching to balance the characteristics of the ruxolitinib and BAT cohorts.

60

As discussed below, the ITC is based only on the RESPONSE trial (plus the matching external BAT cohort). The ITC therefore provides an estimate of the effect of ruxolitinib on overall survival specifically for the splenomegaly subgroup, but not for the no palpable splenomegaly subgroup. The company consider the ITC to be a "supportive analysis and presented for transparency and completeness" (clarification response A8). The ITC results for overall survival do not inform the company's economic analysis base case but do inform scenario analyses (section 5.2.2).

3.3.2 Identification, selection and feasibility assessment of studies for the ITC

The company did not include observational studies in their SLR (section 3.1 above), nor were other data sources for BAT considered (clarification response A8). An indirect comparison (referred to as a MAIC) containing a relevant PV registry (GEMFIN) is listed among the SLR results in CS Appendix D.1.3. The company acknowledge in their clarification response that a systematic search for other real-world registries was not performed, but they argue that a BAT cohort within the GEMFIN registry is likely to represent the most appropriate source of evidence at the time the analysis was conducted:

- The Spanish Registry of Polycythemia Vera set up in 2011 by GEMFIN (Grupo Español de Enfermedades Mieloproliferativas Filadelfia Negativas) referred to as the GEMFIN registry, is one of the largest registries of PV (N=██ as at October 2016) (clarification response A8).
- Results have been published for a subgroup of GEMFIN patients with PV treated with BAT who are resistant to or intolerant of hydroxycarbamide (N=184).[34]
- IPD from GEMFIN were available to the company (clarification response A8).

GEMFIN is a Spanish registry but both the EAG's clinical experts agreed that there is a general lack of robust long-term BAT data for PV patients who are resistant or intolerant to hydroxycarbamide and they were not aware of any registries or other cohorts that would be more relevant than GEMFIN.

An ITC using data from the week 208 analysis of the RESPONSE trial with a subgroup of patients from GEMFIN as the comparator cohort had previously been published as a conference poster by Alvarez-Larrán et al. 2018.[35] The CS provides an update of the ITC using week 256 data from the RESPONSE trial but the GEMFIN data from 2016 (median follow up 3 years) was not updated. The ITC is reported in CS section B.2.9, CS Appendix sections D.1.4. to D.1.8, and in a confidential company slideshow.[36] The ITC used 110

61

patients from RESPONSE and 184 resistant or intolerant to hydroxycarbamide patients from GEMFIN who had at least one follow-up visit.

A later study by Alvarez-Larran 2022 compared BAT (N=272) and ruxolitinib (N=105) cohorts from GEMFIN using an April 2021 data cut. For OS, they reported a hazard ratio of 0.8 (95% CI 0.4, 1.7) which did not reach statistical significance.

3.3.3   Clinical heterogeneity assessment

RESPONSE and GEMFIN were compared in terms of baseline characteristics (clarification response Table 2). Eight of 10 covariates were considered most likely to be prognostic or treatment effect modifiers by company experts (clarification response Table 1). The EAG's experts also considered resistance to hydroxycarbamide, inadequate HCT, and high WBC as prognostic but these were not reported in GEMFIN.

There are notable imbalances in terms of age (61 vs 69 years), cytopaenia at lowest hydroxycarbamide dose (15% versus 7%), male sex (60% versus 47%), time since diagnosis of PV (8.9 versus █ years), and diabetes (█% versus █%). JAK2 mutation status and leg ulcers also showed differences between studies (JAK2: 95% versus 89%; leg ulcers: █% versus █%). However, company experts did not rank either highly as a prognostic factor, and the EAG's experts concurred.

Hence, there are imbalances between RESPONSE and GEMFIN in terms of known prognostic factors between studies. Furthermore, other prognostic factors are not reported so differences between the cohorts are unknown.

3.3.4   Risk of bias assessment for studies included in the ITC

The company conducted a risk of bias assessment for the RESPONSE trial (CS Appendix D.3) but not for the GEMFIN cohort. We note that, in an ITC analysis, risks of bias can arise from within each included cohort (e.g. in selection of cases, management of patients, or assessment of outcomes) as well as from the matching method (e.g. inadequate control of confounding):

- In the RESPONSE trial the main risk of bias concern relevant to the ITC is that the trial was open-label, meaning that patient care in the ruxolitinib arm may have been influenced by investigators' knowledge of the treatment allocations (i.e. high risk of bias) (section 3.2.3).

- In the GEMFIN cohort, the retrospective ascertainment of cases could have led to selection bias (random selection from among the available cases could reduce this risk but would also reduce patient numbers)

- The propensity score matching analysis appears to have mitigated confounding to some extent but there is uncertainty as to whether residual confounding remains, due to the limited number of baseline characteristics that were included as covariates in the matching (section 3.4.1 below).

EAG comment on the studies included in the ITC

The EAG agree that GEMFIN is probably the best source of long-term BAT data available, although the availability of evidence has not been evaluated systematically. There are imbalances in prognostic factors between RESPONSE and GEMFIN and some prognostic factors were not reported.

3.4 Critique of the indirect treatment comparison

3.4.1 Data inputs to the ITC

Matching was conducted on OS only. Whilst the original propensity score matching used week 208 data for RESPONSE and week 728 data for GEMFIN [Alvarez-Larran et al. 2018[35] The analysis in the CS was updated using week 256 data from RESPONSE. The GEMFIN data for the matching were obtained in 2016 (median follow up 3 years). If a later cut of GEMFIN were used there would have been more patients and matching may have been more successful. However, as the data do not belong to the company, presumably this would not have been possible.

Evidence for selection of prognostic factors was based upon opinion of 2 clinicians, and those characteristics available and consistently reported in RESPONSE and GEMFIN. The top 8 prognostic factors were ranked by the experts (clarification response Table 1) but only 4 were included in the analysis. Experts were consulted for the Alvarez-Larran (2018) study [35] hence opinions are quite dated. Studies were matched on age, sex, history of thrombosis, and cytopaenia (CS Appendix D.1.6). Of the remaining 4 prognostic factors, uncontrolled myeloproliferation was excluded as there were no events in RESPONSE, duration of PV diagnosis was excluded as definitions differed by study. Diabetes was excluded as numbers were similar across studies and this factor was ranked low by experts (clarification response A11). No explanation is given for excluding failure to reduce massive splenomegaly but the variable is relatively balanced between studies (█% versus 1%).

No scenario analyses were conducted around variables selection as the company considered GEMFIN "insufficient to support further matching on lower ranked prognostic factors". The EAG disagree, as these rankings were based on the opinions of only two experts. We would have preferred the company to conduct scenario analyses to explore the broader effect of variable selection on ITC results. However, such analyses are unlikely to be feasible as the company do not own the GEMFIN database (clarification response A11[f]).

Whilst the population matching adjusted for some prognostic factors, others were excluded or not reported, and no scenario analyses around inclusion of prognostic factors were conducted.

The feasibility of combining RESPONSE and RESPONSE-2 in the matching exercise was assessed. However, the company say that results of this "exploratory analysis" could not be located, and the explanation provided as to how this analysis resulted in a "poor fit" is unclear (clarification response A9).

3.4.2 Statistical methods for the ITC

Propensity score matching is an appropriate methodology when the company have access to individual participant data (IPD) for both groups. The company matched RESPONSE with the GEMFIN registry.

Only patients from RESPONSE randomised to ruxolitinib were included in the analysis (patients who crossed from BAT to ruxolitinib were not included). Seven patients included in the original Alvarez-Larran et al. 2018 ITC[35] were excluded from the company submission due to a lack of follow up data subsequent to being identified as resistant or intolerant to hydroxycarbamide (clarification response A10). It is unclear why these patients would have been included in the Alvarez-Larran et al. 2018 analysis.[35]

Multivariate regression was conducted using nearest neighbour matching with prognostic factors as predictors and treatment as the dependent variable. Sample size was reduced from ▮ in GEMFIN and 110 in RESPONSE to ▮ post-matching. Studies were reasonably well-matched following matching (Table 12, document B), although there was a ▮% difference in males. Two sensitivity analyses were conducted: (i) using a wider nearest neighbour threshold, and (ii) using an optimal matching approach. Results were consistent with the base case.

3.4.3 Summary of the EAG's critique of the ITC

- The chosen propensity scoring methodology is appropriate for the ITC
- GEMFIN appears to be the best choice of dataset for the BAT cohort, although the available evidence has not been evaluated systematically
- The analysis uses a historical data cut of GEMFIN, but as the company do not have access to the dataset, cannot be updated
- Only a limited set of prognostic factors were included in the analysis and these were based on solicited responses from two experts back in 2016
- No scenario analyses around inclusion of variables in the analysis were conducted
- There may have been missing prognostic factors including those identified by EAG experts (e.g. resistance to hydroxycarbamide, inadequate HCT and high WBC counts)
- No scenario analyses were conducted including patients from RESPONSE-2 or MAJIC-PV
- The company list a number of uncertainties in the ITC results including whether GEMFIN was representative of a UK population (they concede low use of IFN-alfa) the generalisability of the GEMFIN population, shorter follow up for GEMFIN (3 years versus 5 years for RESPONSE), a failure to use RESPONSE-2 in the matching, and being unable to include many covariates in the matching (CS section B.2.9.2)
- A published comparison of patients from GEMFIN reported no statistically significant difference in OS between those who received ruxolitinib and BAT

In conclusion, based on the above, in our opinion the OS estimates from the company ITC are highly uncertain

3.5 Overall survival results from the ITC

The overall survival results are shown in Table 17. However, as noted above, we believe these are highly uncertain.

Table 17 Overall survival results from the indirect treatment comparison

Analysis	Number of patients		Number of events		HR (95% CI)[a]
	BAT	Ruxolitinib	BAT	Ruxolitinib	
Pre-matching[b]	■	■	■■■■■ ▎	■■ ■■■	■■ ■■■■■
Post-matching[b]	■	■	■■ ■■■	■■ ■■■	■■ ■■■■■
Post-matching[c]	■	■	■■ ■■■	■■ ■■■	■■ ■■■■■

3.6 Additional work on clinical effectiveness undertaken by the EAG

No additional analyses have been conducted by the EAG, as no statistical code nor input data for the ITC were provided to validate the results.

3.7 Conclusions on the clinical effectiveness evidence

The EAG have not identified any key issues in the clinical efficacy evidence that could be resolved by acquiring any additional data or by using alternative analysis approaches. Limitations of the existing data and reporting mean that the clinical efficacy outcomes are subject to uncertainty that would be difficult to resolve unless new evidence (and clearer reporting of studies) becomes available. The three RCTs are all at high risk of bias meaning that variance estimates such as 95% confidence intervals will underestimate the uncertainty present. HRQoL outcomes are particularly at risk of bias due to lack of clarity around missing data, subjectivity of the outcomes in relation to the open-label nature of the RCTs, and selective reporting. Inclusion of the MAJIC-PV trial to compensate for confounding after early crossover in the RESPONSE and RESPONSE-2 company trials is appropriate and has additional advantages, e.g. consisting of a wholly UK population, but is limited by superficial and ambiguous reporting of some aspects of the trial.

Residual uncertainty in the clinical efficacy evidence is summarised in Table 18 below. Although safety outcomes are difficult to compare across trials due to inconsistent reporting, adverse events were generally as expected and do not raise any new concerns.

Table 18 Residual clinical efficacy uncertainties identified by the EAG

#	Source of uncertainty	Effect on certainty of evidence	EAG comment/resolution
1	Radioactive phosphorus is included as a comparator in the NICE scope but excluded from the company's decision problem (section 2.3.2).	Trial BAT arm evidence may not be entirely representative of the NHS PV population receiving BAT who are R/I to HC.	The EAG's clinical experts confirmed that radioactive phosphorus is hardly ever used in clinical practice.
2	Lack of standardisation of definition of R/I to HC	The NHS PV population who are R/I to HC could be broader than in the trials and	The EAG's clinical experts confirmed baseline characteristics of the trials are generally reflective

66

	in clinical practice (section 2.3.1).	so influence the overall costs of introducing ruxolitinib.	of the NHS PV population and subgroups who are R/I to HC.
3	High risk of bias in all three RCTs (section 3.2.3 and Appendix 9.3) • Open label • Selective reporting • Handling of missing data	Uncertainty around the outcomes is not fully captured in the variance measures such as 95% confidence intervals, where reported.	Open label aspect was not justified, however cannot be changed retrospectively. Clarification could be sought on the randomisation process, selective reporting, and missing data around HRQoL specifically.
4	Lack of data from MAJIC-PV: trial publication is unpublished • selective reporting • IPD not available	There are ambiguities around some aspects of the MAJIC-PV trial, e.g. relating to crossovers, missing data and why EQ-5D was not reported.	MAJIC-PV was an investigator-led trial and IPD could not be made available to the company. Final publication of the draft trial manuscript might improve some aspects of clarity.
5	Non-RCT evidence was not systematically searched for: • the SLR was structured to only identify RCTs • Clarification response A8 confirms no systematic search was done to identify real-world studies for the ITC. • Provenance of a study used for additional scenario analyses is not reported.	Uncertainty whether the GEMFIN registry cohort (Alvarez-Larran et al. 2018)[35] used in the ITC is the most appropriate (externally valid) BAT cohort. Uncertainty whether the GEMFIN registry cohort (Alvarez-Larran et al. 2022)[37] used in additional scenario analyses is the most appropriate (externally valid) source of evidence.	The EAG's clinical experts were not aware of any other long term BAT cohorts that would be more relevant and considered the GEMFIN BAT cohort broadly generalisable to the UK. Secondly, the ITC is considered by the company as supportive and not critical evidence. The EAG did not identify a need for the ITC or observational study results to inform the economic model as the included RCTs are sufficiently representative.
6	ITC methods: The results of the ITC are highly uncertain due to: • Limited adjustment for imbalances in prognostic factors between the treatment groups. • High risk of bias in the existing RESPONSE study and in case selection from the GEMFIN registry. • Used an old data cut from the GEMFIN registry • Scenario analyses were not conducted	The overall survival estimates from the ITC are uncertain.	The EAG are not aware of any other data that would provide for a more robust analysis. Selection bias in the GEMFIN cohort was partly resolved by propensity score matching. An updated data cut from the GEMFIN registry was not available as the company do not have access to the dataset. The results inform overall survival estimates (and no further outcomes, except that the published study also analysed thrombosis) in scenario analyses only, not in the base case. Results from a recent comparison of BAT and ruxolitinib patients from

		GEMFIN did not find a statistically significant difference in overall survival.
around selection of variables or around including patients from RESPONSE-2 or MAJIC-PV		

BAT: best available therapy; HC: hydroxycarbamide; HRQoL: health-related quality of life; IPD: individual patient level data; ITC: indirect treatment comparison; PV: polycythaemia vera; RCTs: randomised controlled trials; R/I: resistant to or intolerant of; SLR: systematic literature review.

4 COST EFFECTIVENESS

4.1 EAG comments on the company's review of cost-effectiveness evidence

The company conducted a systematic search for literature on economic evaluations, health state utilities and UK resource use and costs for adults with PV (CS Appendix G). The search strategy was appropriate and reasonably up to date (last updated June 2022). The EAG do not have any concerns about the design or conduct of the reviews. We discuss results for the reviews of utilities and costs/resource use, respectively, in sections 4.2.7.1 and 4.2.8.1 below.

The review of economic evaluations identified five studies, including assessments of the cost-effectiveness of ruxolitinib compared with BAT in populations with PV resistant or intolerant to HC in Ireland (NCPE 2016), the United States (Hong et al. 2020) and Scotland (SMC 2019).[38-40] The SMC have also reported an assessment for ropeginterferon alfa-2b compared with ruxolitinib in a high-risk PV population (SMC 2022).[41] See CS Appendix G Tables 31, 33 and 34 for further details.

4.2 Summary and critique of the company's submitted economic evaluation

4.2.1 NICE reference case checklist

The company summarise key features of their economic evaluation in CS Table 19. The EAG assessment of the company's economic analysis against the NICE reference case checklist is shown in Table 19 below.[42] The company's analysis meets all reference case criteria, except for use of NICE's preferred measure of health-related quality of life, the EQ-5D. Instead, the company use a condition-specific preference-based measure developed for myelofibrosis, the MF-8D, for their base case analyses.[43] See section 4.2.7.2 below for discussion and EAG critique of this decision.

Table 19 NICE reference case checklist

Element of health technology assessment	Reference case	EAG agrees submission meets reference case
Defining the decision problem	The scope developed by NICE	Yes
Comparator(s)	As listed in the scope developed by NICE	Yes
Perspective on outcomes	All health effects, whether for patients or, when relevant, carers	Yes
Perspective on costs	NHS and personal social services (PSS)	Yes

Element of health technology assessment	Reference case	EAG agrees submission meets reference case
Type of economic evaluation	Cost–utility analysis with fully incremental analysis	Yes
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	Yes. Effectively lifetime (46 years from starting age at model entry)
Synthesis of evidence on health effects	Based on systematic review	Yes. Health effects from RESPONSE, RESPONSE-2 and MAJIC-PV trials. Scenario with OS HR from ITC. See 4.2.6 below for discussion.
Measuring and valuing health effects	Health effects should be expressed in QALYs. The EQ-5D is the preferred measure of health-related quality of life in adults.	No. Base case analysis uses MF-8D measure (EQ-5D in scenario). See section 4.2.7.2 below for discussion.
Source of data for measurement of health-related quality of life	Reported directly by patients and/or carers	Yes
Source of preference data for valuation of changes in health-related quality of life	Representative sample of the UK population	Yes. MF-8D and EQ-5D valuations from UK general population sample.
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit, except in specific circumstances	Yes. The criteria for use of QALY weighting for severity are not met, see Section 7 below.
Evidence on resource use and costs	Costs should relate to NHS and PSS resources and should be valued using the prices relevant to the NHS and PSS	Yes
Discounting	The same annual rate for both costs and health effects (currently 3.5%)	Yes
Source: assessment by EAG. Criteria from NICE health technology evaluations: the manual, January 2022		

70

4.2.2 Model structure

4.2.2.1 State-transition model for RESPONSE and RESPONSE-2 ('primary analysis')

For their primary analysis, the company use a cohort state-transition model (STM) for the licensed population subgroups with and without splenomegaly, based on populations in the RESPONSE and RESPONSE-2 trials respectively (see CS section B.3.2). The same model is used to calculate separate results for each subgroup (pooled results for the whole licensed population are not presented). The model is implemented in Microsoft Excel and employs a 28-day cycle length with a maximum 46-year time horizon, which is effectively lifetime given the age of the cohort at model entry. No half-cycle correction was applied due to the short cycle length. The model was developed with input from an advisory board comprising five UK-based haematologists with PV experience, as well as published literature.

Overview of the model structure

A schematic of the STM structure is provided in CS Figure 35. The model includes three main health states, defined by therapy phases as opposed to disease stages (an approach used in TA386 and TA756 for the treatment of myeloid fibrosis with ruxolitinib and fedratinib).[15][44] Patients enter the model in either the ruxolitinib state or the BAT state, depending on the treatment arm. Patients remain in the ruxolitinib state until discontinuation of ruxolitinib or death. After discontinuation of ruxolitinib, patients move into the BAT state. Patients in the BAT state remain there until death.

In the base case analysis, the BAT state is partitioned into three sub-states, which represent different stages of treatment: first BAT; second or subsequent BAT; and no treatment (discontinuation of all BAT). The company use this BAT partition to model progressive decline in health-related quality of life as patients move through the BAT regimens: utility declines between first, second/subsequent and no further treatment substates, see section 4.2.7.3 below. The BAT partition is implemented using a series of tunnel states, which capture time since initiation of BAT. A scenario analysis with no BAT partition is also presented.

Key complications associated with PV (thromboembolic events (TE), progression to MF, progression to AML and myelodysplastic syndrome (MDS), and haemorrhage) are modelled as events rather than as health states. Incidence rates for these complications and for therapeutic phlebotomy are lower in the ruxolitinib state than in the BAT state, but ruxolitinib is associated with a higher incidence of NMSC. One-off costs and QALY losses are applied for incident cases of TE, MF, AML/MDS, NMSC, haemorrhage and therapeutic phlebotomy.

71

The company argue that inclusion of these events as health states would be particularly challenging, as it would require many assumptions and data that are not available for the population (CS section B.3.2.2).

Approach to estimation of transition probabilities
The STM structure requires probability estimates for transitions between the ruxolitinib, BAT and death states. These probabilities are estimated from OS and time to treatment discontinuation (TTD) data from the trials. This is challenging for two reasons. Firstly, OS is immature in all three trials due to the relatively good prognosis for people with PV. Secondly, although five-year OS is available for the ruxolitinib arms in the RESPONSE and RESPONSE-2 trials, data for the BAT arms is confounded by cross-over (no patients remained on BAT after 80 weeks). Five-year comparative data are available from the MAJIC-PV trial, as this is unlikely to have been affected by cross-over (Harrison et al. 2022 supplementary Figure S5D).[16]

The company describe their approach to estimating time to treatment discontinuation and overall survival in CS sections B.3.1.2 to B.3.3.4 (note there is an error in the numbering of these sections in the CS). The estimation process is complex; an overview of the EAG's understanding of the process is as follows:

- TTD for reasons other than death is estimated for the ruxolitinib arm using competing-risk analyses of individual patient data from the RESPONSE and RESPONSE-2 trials. These analyses are conducted separately for the two trials and provide separate estimates of ruxolitinib TTD (with deaths censored) for the populations with and without splenomegaly.

- As the numbers of deaths observed in the trials were low, pre- and post-discontinuation survival for the ruxolitinib arm are estimated from pooled data from the RESPONSE and RESPONSE-2 trials.

- Parametric distributions are fitted to the ruxolitinib arm TTD, pre-discontinuation survival and post-discontinuation survival for each trial population. The model combines these extrapolations to estimate OS for the ruxolitinib arm.

- OS for the BAT arm is derived from the modelled OS for ruxolitinib adjusted downwards using a time-varying hazard ratio estimated from MAJIC-PV. The treatment effect is not estimated from the RESPONSE and RESPONSE-2 trials because of the problem with cross-over.

72

- The rates of discontinuation for the first BAT regimen and for all BAT regimen are estimated from MAJIC-PV data.

Further details and EAG critique of the company's approach to estimation of TTD and OS extrapolations are provided in sections 4.2.6.1 and 4.2.6.2 below.

Other model parameters

In addition to TTD and pre- and post-discontinuation survival, the model uses input parameters to estimate incidence rates for key events and adverse reactions, utilities and resource use/costs. The company present a summary of input parameters for the base case model in CS Table 37. They made some corrections to the parameter values reported in the CS in response to clarification questions and noted that the values in the model were correct. We discuss and critique the clinical effectiveness, utility and resource use/cost parameters in sections 4.2.6, 4.2.7 and 4.2.8 below.

4.2.2.2 Partitioned survival model for MAJIC-PV population ('subgroup analysis')

Individual patient data from the MAJIC-PV trial was not available to the company, as the trial is investigator-led. Consequently, the company employed a partitioned survival model (PSM) to estimate cost-effectiveness for the MAJIC-PV population. In this approach, the proportion of patients in each health state at each time point is estimated based on conventional survival outcomes (usually PFS and OS), and explicit modelling of transitions between the health states, which requires individual patient data, is not needed.[45][46]

As in the primary analysis, the model for the MAJIC-PV population has three health states, based on treatment: 'on ruxolitinib, 'on BAT' and death. Hence, the survival data required is TTD for ruxolitinib and OS. In this model, the BAT health state is not partitioned as with the primary analysis model. Conversely to the primary analysis, the OS for BAT is extrapolated directly from reconstructed KM data reported in the unpublished MAJIC-PV trial paper, with the OS for ruxolitinib estimated indirectly using a time-varying treatment effect.

4.2.2.3 EAG critique of model structure

EAG comments on the modelling approaches: STM vs. PSM (Key issue 4)

- In methodological terms, the state-transition approach has the advantage that the OS extrapolation is structurally related to ruxolitinib discontinuation, unlike the partitioned survival approach in which these outcomes are modelled independently.[45][46] In the current appraisal, the company report scenario analysis with their primary STM

73

model exploring uncertainty over the extrapolations of both pre-discontinuation survival and post-discontinuation survival.

- NICE DSU TSD19 notes that empirical comparisons have shown that the STM and PSM approaches can produce markedly different results, and that "it is not clear which approach is more reliable".[45] Consequently, TSD19 recommends that STMs should be presented alongside PSMs to verify the plausibility of the PSM extrapolations and to explore key uncertainties in the OS extrapolations.[45]

- A further uncertainty in the current appraisal is whether differences in results from the company's primary and subgroup models relate to the modelling technique (STM versus PSM), or to the different trial populations and contexts of treatment. Exploration of alternative modelling approaches might help to clarify this point. It is not currently possible for the company to conduct an STM analysis for the MAJIC-PV trial population, as they do not have access to individual patient date. However, it would be possible for the company to compare STM and PSM approaches for analysis of the RESPONSE and RESPONSE-2 trial populations.

EAG comments on model structure, states and events (Key issue 5)

- The company's decision to use therapy phases as states, rather than stages of disease, means that their model structure does not reflect the natural history of PV. Although discontinuation of ruxolitinib is likely to be related to long-term survival, other intermediate outcomes such as progression to more aggressive forms of cancer and major thromboembolic or haemorrhagic events are likely to be more strongly prognostic.

- The company cite TA386 and TA756 appraisals as precedent for the use of therapy-based health states for MF. However, a 'supportive care' state after discontinuation of treatment for MF was used in TA386 and TA756. We suggest that the supportive care state may be more directly related to decline in quality of life than the post-ruxolitinib BAT state for PV in the current appraisal.

- We understand that modelling multiple PV-related complications as states rather than as events would add complexity and require additional assumptions and parameter estimates and add uncertainty. However, we note that there are large uncertainties associated with the current model structure. In particular, we are concerned that extrapolation of all-cause mortality from the trials may not reflect the full impact of PV due to time lags between the onset of major complications and related mortality, and the increasing incidence of PV complications with age.

74

- A more conventional structure for the MAJIC-PV PSM would have been to use a measure of disease progression to define the health states, in addition to treatment discontinuation. For example, the MAJIC-PV manuscript reports KM curves and relative treatment effects for progression-free survival and event-free survival (see section 3.2.6.8 above). One of these intermediate survival outcomes could be used to define pre and post- progression/event health states in a standard three-state PSM structure. We suggest that the company consider an alternative model structure, incorporating an intermediate survival outcome.

EAG comments on partitioning of the BAT state

- Clinical advice to the EAG is that there is not a clear sequence of lines of BAT treatment and long-term cessation of all BAT is considered to be rare. In the absence of alternatives, patients with PV who are resistant or intolerant to hydroxycarbamide continue to switch between currently available medical treatments, with dose adjustments and interruptions to manage symptoms and risks, although this often results in suboptimal control. There is uncertainty over the long-term rate of discontinuation of all BAT therapies and over the assumptions about disutilities for the BAT substates (see sections 4.2.6.1 and 4.2.7.3 respectively). We therefore do not use the BAT partition in the EAG preferred analyses, but we include it in scenario analysis. This is not considered to be a key issue, as the impact on the cost-effectiveness results is modest.

4.2.3 Population

The decision problem population is adults with PV who are resistant or intolerant to hydroxycarbamide, in line with the marketing authorisation for ruxolitinib and the current decision problem (CS B.3.2.1).

The company report three sets of cost-effectiveness results for different subgroups of this population. The primary analysis uses data from the RESPONSE and RESPONSE-2 trials to model subgroups with splenomegaly and without splenomegaly respectively. In addition, the company report results for a 'high-risk subgroup', based on the population in the MAJIC-PV RCT. The company argue that all three trial populations are generalisable to England and Wales (CS Table 38). See section 3.2.2 above for discussion of baseline characteristics for patients in the three trials.

The company argue that, collectively, the trial populations with and without splenomegaly in RESPONSE and RESPONSE-2 represent the entire licensed population; with a split of approximately 20% with splenomegaly and 80% without (CS B.3.2.1). Estimates of the prevalence of splenomegaly in practice vary depending on the assessment method and it is difficult to compare estimates from the different trials. In the MAJIC-PV trial, ■ of the population had palpable splenomegaly at baseline (Appendix 9.2 below).

EAG comments on model population (Key issue 1)

- The baseline characteristics of patients in the three clinical trials on which the company's economic analyses based are broadly similar, with the exception of splenomegaly. The EAG clinical advisers agree that all three populations are generally reflective of NHS patients with PV who are resistant to or intolerant of hydroxycarbamide, but that the slightly older population in MAJIC-PV was closer to the patients who they see (section 3.2.2 above). However, we note that estimated survival in the MAJIC-PV population appears noticeably worse than in the RESPONSE and RESPONSE-2 trial populations.

- The NICE scope requests subgroup analysis for patients with and without splenomegaly, which is currently only available from the RESPONSE and RESPONSE-2 trial populations. Expert advice to the EAG is that splenomegaly status would be known at the time patients of consideration for ruxolitinib treatment as patients are assessed by palpation, so this subgroup is identifiable. The EAG experts suggested that people with splenomegaly are more likely to benefit from treatment with ruxolitinib than patients without splenomegaly, although evidence of a difference in treatment effect is lacking. Further analysis to compare cost-effectiveness results for people with and without splenomegaly should be conducted as and when subgroup analysis by baseline splenomegaly status becomes available for the MAJIC-PV trial.

4.2.4 Interventions and comparators

The economic model compares the incremental cost-effectiveness of ruxolitinib to best available therapy (BAT). The intervention and comparator are consistent with the NICE scope. See section 4.2.8.2 below for comments on the dosing assumptions and mix of current treatments in UK practice.

4.2.5 Perspective, time horizon and discounting

The company analyses take the perspective of the NHS and Personal Social Services (PSS) in England, which aligns with the NICE manual for health technology assessments.[42] Costs and outcomes (life years and QALYs) are discounted at 3.5%. The company uses a lifetime horizon to reflect the chronic nature of PV, where lifetime is assumed to be 46 years from the start of the model. Given that the starting age of the patient population in the model is approximately 60-66 years, the company's scenario analysis with a shorter time horizon of 30 years may be more appropriate. We include this scenario in EAG additional analysis (section 6.2.2 below).

4.2.6 Treatment effectiveness and extrapolation

The clinical parameters used in the model consist of time to treatment discontinuation (TTD), parameters required to estimate overall survival (OS) and incidence rates for key complications, therapeutic phlebotomy and adverse events. These parameters were estimated from RESPONSE and RESPONSE-2 trial data, and from the unpublished manuscript for the MAJIC-PV trial, as summarised in CS Table 21. We summarise the clinical parameters used in the company's primary and subgroup models in Table 20 and Table 21 respectively. Description and EAG critique of the company's approach to estimating these parameters is provided in the following sections of this report.

Table 20 Summary of clinical parameters in the primary model (RESPONSE and RESPONSE-2 trial populations)

Parameter	Base case analysis	Source
Time to treatment discontinuation (TTD)		
Ruxolitinib TTD (excluding death)	Odds spline with 1 knot for both subgroups, CS Figure 39	Competing-risk analyses of RESPONSE and RESPONSE-2 IPD for the two subgroups
BAT TTD 1st BAT	KM and Gompertz tail	Extrapolation of reconstructed KM data for discontinuation of first BAT regimen in MAJIC-PV
BAT TTD all BAT	BAT OS / HR (███)	HR approximated from numbers of deaths and discontinuations in the BAT arm of MAJIC-PV
Overall survival (OS)		
Ruxolitinib pre-discontinuation survival	Exponential for both subgroups (+ gen pop mortality constraint applied post- trial) CS Figure 41	Data from RESPONSE and RESPONSE-2 were pooled due to the small number of deaths observed within the trials (same extrapolations for both subgroups)
Ruxolitinib post-discontinuation survival	Exponential (+ gen pop mortality constraint over time horizon) CS Figure 44	
OS for ruxolitinib	Calculated indirectly by STM	-

77

OS for BAT	Ruxolitinib OS x time varying HR (█████████████████████, waning from year 5 to HR=1 at year 20) CS Figures 47 and 48	HR estimated from piecewise Cox proportional hazards analysis of reconstructed MAJIC-PV KM data
Event rates		
Key complications and phlebotomy (ruxolitinib)	Exposure-adjusted incidence rates while on ruxolitinib CS Table 24	Incidence rates estimated from relevant trial for population when available
Key complications and phlebotomy (BAT)	Incidence for ruxolitinib adjusted for BAT with IRR CS Tables 25	Incidence-rate ratios calculated from pooled RESPONSE, RESPONSE-2 and MAJIC-PV
Adverse events	Incidence rates CS Table 23	Exposure-adjusted incidence rates (any grade) pooled for RESPONSE and RESPONSE-2
Source: summary produced by EAG BAT best available treatment; CS company submission; CQ clarification question response; gen pop, general population; HR hazard ratio; IPD individual patient data; IRR incidence-rate ratios; KM Kaplan-Meier; OS overall survival; TTD time to treatment discontinuation		

Table 21 Summary of clinical parameters in the subgroup model (MAJIC-PV population)

Parameter	Base case analysis	Source
Time to treatment discontinuation (TTD)		
Ruxolitinib TTD	Ruxolitinib OS x HR for TTD vs. OS (███) See CS Figure 42 and CQ response B5	Ruxolitinib OS adjusted with HR for TTD vs. OS. HR estimated from reconstructed KM for ruxolitinib arm of MAJIC-PV
TTD all BAT	BAT OS / HR (███)	Estimated as above
Overall survival (OS)		
OS for BAT	Weibull extrapolation (+ gen pop mortality constraint over time horizon) CS Figure 46	Extrapolation fitted to MAJIC-PV reconstructed KM data for BAT arm
OS for ruxolitinib	BAT OS / time varying HR	BAT extrapolation adjusted by same HR as in primary analysis
Event rates		
Complications Phlebotomy Adverse events	Same as for primary analysis	
Source: summary produced by EAG BAT best available treatment; CS company submission; CQ clarification question; gen pop, general population; HR hazard ratio; KM Kaplan-Meier; OS overall survival; TTD time to treatment discontinuation		

4.2.6.1 Time to treatment discontinuation

4.2.6.1.1 Primary analysis (RESPONSE and RESPONSE-2 populations)

Ruxolitinib discontinuation

The TTD for ruxolitinib was modelled under a competing-risk framework, which is appropriate for the state-transition model. This allows the model to account for the increased likelihood of discontinuation due to death as patients age. The TTD for ruxolitinib due to reasons other than death and pre-discontinuation survival are initially modelled separately before being combined within the model 'trace' sheets.

The approach to fitting extrapolations for ruxolitinib discontinuation for reasons other than death is explained in CS section B.3.1.2. The analysis was conducted separately for people with and without splenomegaly, using individual patient data (with deaths censored) from the RESPONSE and RESPONSE-2 trials respectively (see CS Figure 38). The company followed recommended methods to fit and choose extrapolations in each population from NICE Decision Support Unit (TSD14).[47] See CS Appendix N.1 and N.2 for graphs and statistical measures of fit. For the base case, the company chose the odds spline model with one knot for both patients with and without splenomegaly (CS Figure 39). Other distributions were used in scenario analysis and the ICERs were moderately sensitive to the choice of distribution (CS Appendix P).
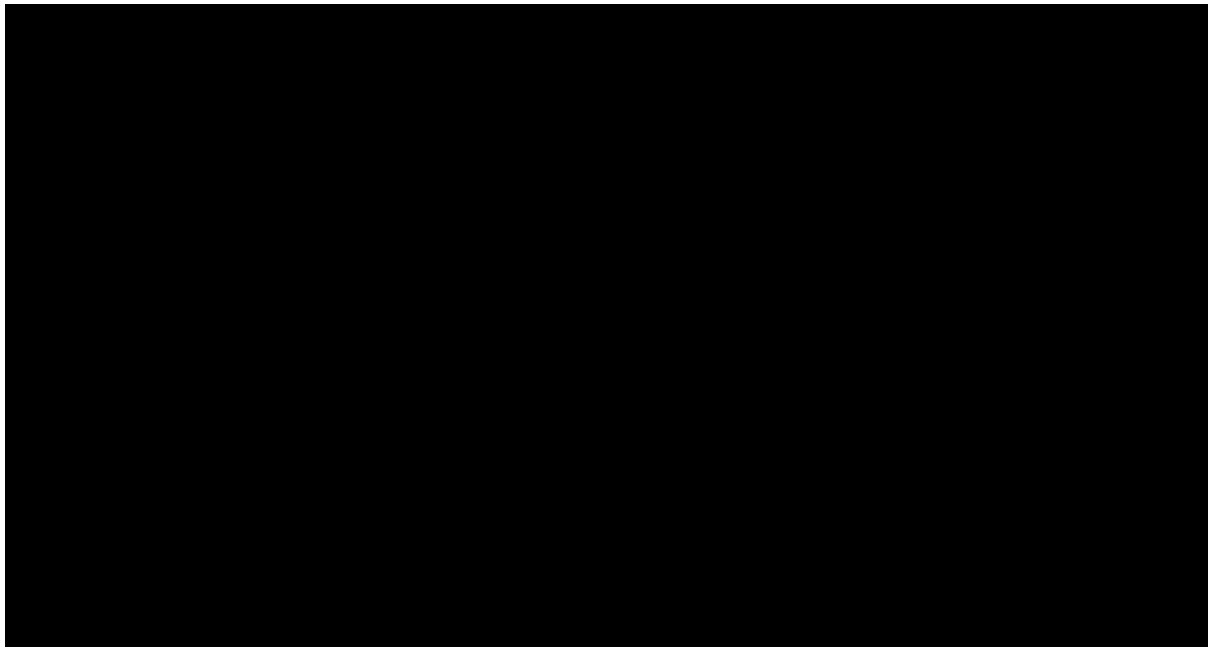
Figure 1 and Figure 2 below show the company's selected odds spline with one knot distribution and the EAG's preferred assumption of a Weibull distribution in comparison with KM data for TTD for ruxolitinib due to reasons other than death for the licensed population

with and without splenomegaly, respectively. We prefer the Weibull distribution, because it has a better statistical fit for the RESPONSE trial and similar fit for RESPONSE-2.

Results with other selected distributions (lognormal, loglogistic, and the hazard spline with one knot) are shown in scenario analysis in Table 27.
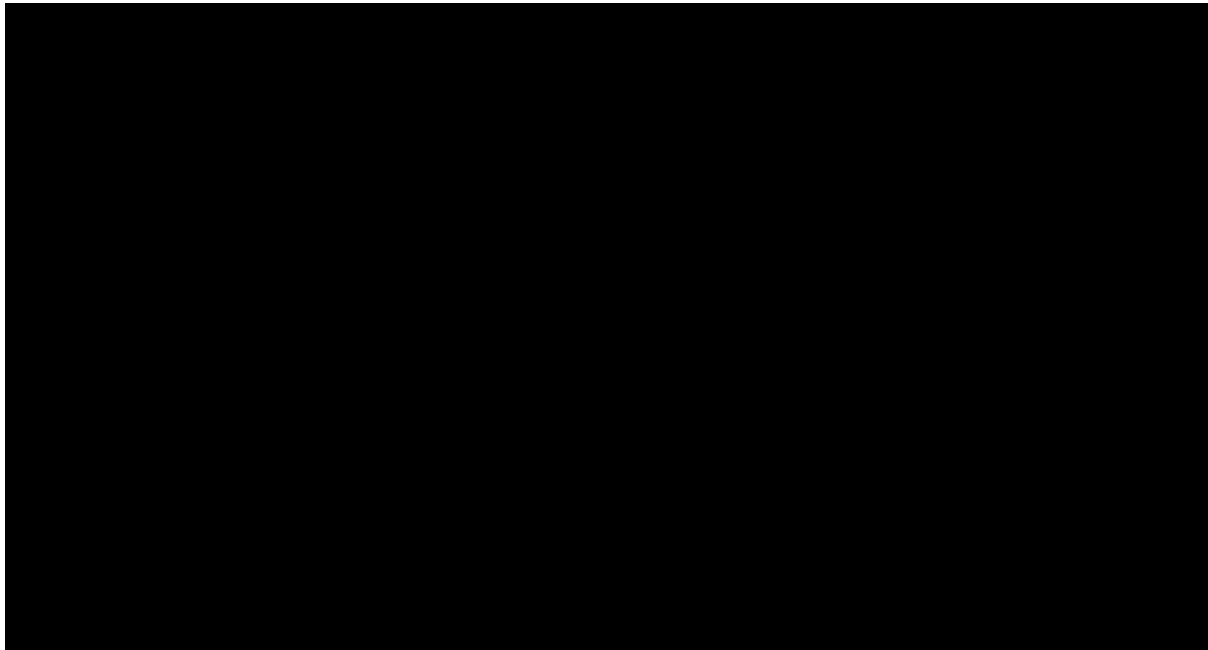
Figure 1 TTD for ruxolitinib for the licensed population with splenomegaly

Abbreviations: TTD: time to treatment discontinuation; KM: Kaplan-Meier; OS: overall survival.

Source: Reproduced from CS Appendix N Figure 18 using selected distributions.
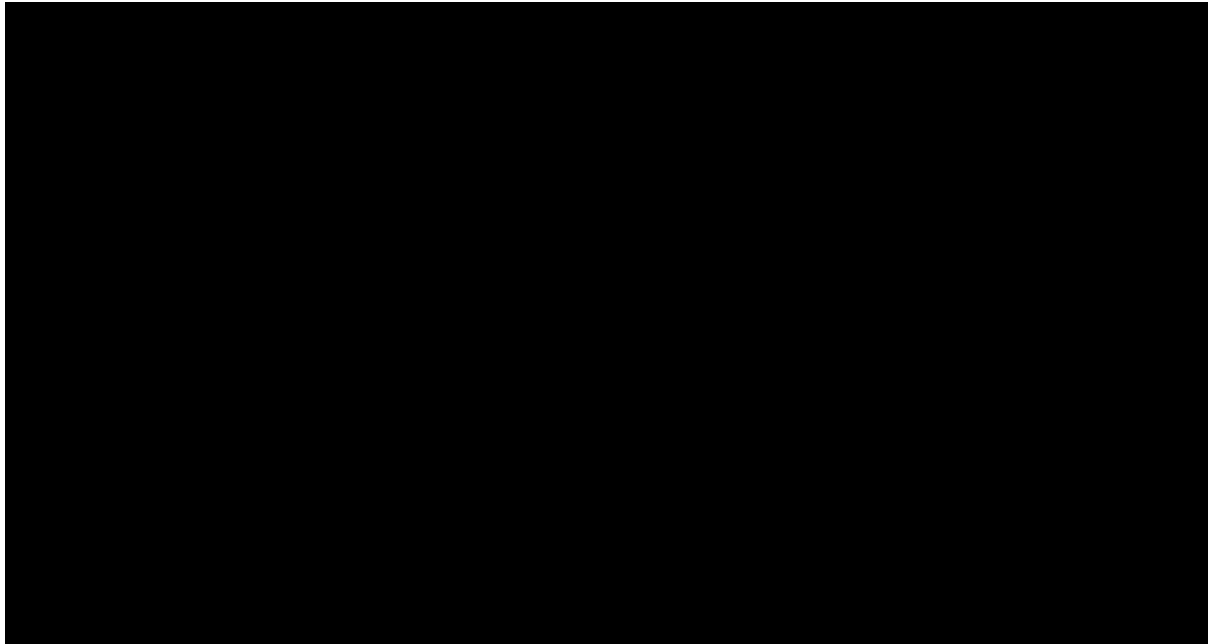
Figure 2 TTD for ruxolitinib for the licensed population without splenomegaly

Abbreviations: TTD: time to treatment discontinuation; KM: Kaplan-Meier; OS: overall survival.

Source: Reproduced from CS Appendix N Figure 19 using selected distributions.

BAT discontinuation

Time to discontinuation of the first BAT treatment was derived from reconstructed KM data from the MAJIC-PV trial (CS Figure 50). As with TTD for ruxolitinib, parametric extrapolations were fitted to the KM data and the fit and clinical plausibility assessed (see CS section B.3.3.5 and appendix N.6). As the data were mature, the company chose to use KM data directly for the 5-year follow up, with a Gompertz extrapolation for the remaining time horizon. The number of people remaining at risk in the KM at 4 and 5 years was █ and █ respectively.

The TTD for all BAT treatments is not reported in the MAJIC-PV manuscript; the company estimated a hazard ratio (HR) between OS and TTD of █ using the number of reported deaths and discontinuations in the BAT arm from the unpublished manuscript.

See CS Figure 49 for the resulting distribution between the three BAT substates in the company's base case model. The TTD for second and subsequent BAT is estimated as the difference between the TTD for first BAT and TTD for all BAT. The time in no treatment is taken as the difference between OS for BAT and the TTD for all BAT.

81

The company assume that after discontinuation of ruxolitinib, patients are distributed to the three BAT substates in the same proportions as patients who were initiated on BAT at the same model cycle.

4.2.6.1.2 Subgroup analysis (MAJIC-PV population)

Time to discontinuation for ruxolitinib in the partitioned-survival model for the MAJIC-PV population was not modelled under a competing-risk framework, as the company did not have access to individual patient data from the MAJIC-PV trial. Instead, a HR of ████████████████ was derived from reconstructed pseudo IPD for OS and TTD for ruxolitinib, which was then applied to the OS for ruxolitinib to obtain the TTD (note this HR was incorrectly reported in the CS, see correction in the company's response to clarification question B5). The company note that this approach follows clinical expert advice that TTD for ruxolitinib should be consistent with OS.

For discontinuation of BAT in the PSM, the same approach is used as described above for the primary STM model.

EAG comments on TTD extrapolations

- The company followed the recommended approach to fitting extrapolations for time to discontinuation of ruxolitinib and initial BAT treatment and provided clear reasons for their choice of distributions in the base case models.

- In the company's primary model, the distribution used for the extrapolation of TTD for ruxolitinib has a moderately large impact on the ICERs (CS Appendix P), because the STM structure means that TTD impacts on long-term survival as well as treatment-related utility and costs. (Note that this is not the case for the subgroup model (PSM) for the MAJIC-PV population, in which the TTD for ruxolitinib is linked via hazard ratio parameters to the OS extrapolation for BAT.)

- The company's choice of distribution (odds spline with one knot) for the extrapolation of ruxolitinib discontinuation in their primary is reasonable. We use a Weibull distribution in EAG preferred analysis, as this has a better statistical fit for the RESPONSE population. This results in a bigger difference in long-term continuation of ruxolitinib between the two subgroups, as shown in **Error! Reference source not found.** and Figure 2 above.

  Results with the Weibull and other selected distributions (lognormal, loglogistic, and the hazard spline with one knot) are shown in scenario analysis in Table 27.

- The model results are not sensitive to changes in the distributions used for extrapolation of time to discontinuation of BAT estimated from the MAJIC-PV trial (same distributions used in all three populations).

4.2.6.2 Overall survival

4.2.6.2.1 Treatment effect (OS HR for ruxolitinib versus BAT)

For the base case, the company used a time-varying HR estimated from reconstructed KM data from the MAJIC-PV trial. The company's clinical advisors noted that the KM curves appear to diverge after about ▮ years (see CS Figure 47), which was in line with the experts' expectations based on intermediate outcomes (CS B.3.3.4). The company fitted a piecewise Cox proportional hazards model to reconstructed MAJIC-PV KM data to estimate hazard ratios before and after this cut point. CS Appendix O shows log-log and Schoenfeld residuals plots based on reconstructed KM data, which the company used to assess the timing of the change in HR.

The company reported scenarios with different cut-points (▮▮▮ years) for their time-varying HR estimates. They also reported four other scenarios with fixed HR estimates applied throughout the time horizon: the HR from the unpublished report by the MAJIC-PV investigators; the estimate from the company's ITC analysis (see section 3.5 above); a propensity score adjusted incidence rate ratio (IRR) of death from a retrospective analysis of Spanish registry data (Alvarez-Larrán et al. 2022)[37]; and an HR estimated from pooled RESPONSE and RESPONSE-2 trial data, without adjustment for crossover. As might be expected, the ICERs were highly sensitive to these very different HR estimates (CS Appendix P and company response to clarification question B2).

Table 22 Treatment effect estimates used in company analysis

Analysis	HR for OS (ruxolitinib vs. BAT)	Source
MAJIC-PV time-varying HR (base case)	▮▮▮▮▮▮▮ ▮▮▮▮▮▮▮	CS section B.3.3.4
MAJIC-PV constant HR	▮▮▮▮▮▮	Harrison et al. 2022, Figure S5D [16]
Company ITC	▮▮▮▮▮	CS Table 13
Spanish registry data	0.8 (95% CI 0.4 to 1.5; p=0.4)	Alvarez-Larrán et al. 2022 [37]
Pooled RESPONSE and RESPONSE-2 data	▮▮▮▮▮	Company model
Source: EAG using data from company submission and model BAT best available treatment; CI confidence interval; HR hazard ratio; ITC indirect treatment comparison; OS overall survival		

83

Waning assumptions

In their base case, the company assume a gradual waning of the treatment effect after the trial period: with a linear increase in the HRs from the above estimates at year 5 to no effect (HR=1) at year 20 and beyond. This was based on clinical expert judgement that approximately twice the number of patients would be alive at 20 years with ruxolitinib compared with current treatment (see CS section B.3.3.4). The company tested various scenarios for the duration of the waning period, from 5 to 50 years. Results were sensitive to different waning assumptions.

EAG comments on the treatment effect for survival (Key issue 2)

- Evidence on the relative treatment effect on survival is highly uncertain. The confidence interval around the HR reported by the MAJIC-PV trial investigators is wide. The company's time-varying HR estimates are not unreasonable based on trends in the MAJIC-PV KM curves (CS Figure 47). The log-log and Schoenfeld residuals plots (CS Appendix Figures 26 and 27) provide support for the assumption of proportional hazards prior to ▮ years and increasing divergence after this timepoint. However, these estimates are also highly uncertain. For the EAG analysis, we prefer to use the constant HR estimate as reported by the MAJIC-PV trial investigators, but we report results with the company's time-varying HR estimates in scenario analysis.

- Other estimates of the treatment effect are used in the company's scenario analyses, including: estimates from pooled RESPONSE and RESPONSE-2 data, the ITC matched comparison with GEMFIN registry data, and the analysis of Spanish registry data (Alvarez-Larrán 2022)[37], see Table 22 above. We report EAG results with these scenarios for information but consider the MAJIC-PV trial to be the most robust source of evidence for relative treatment effects.

EAG comments on the waning of the OS treatment effect (Key issue 3)

- There is uncertainty over whether and how the treatment effect might change after the trial period. Given the uncertainties around the estimation of the treatment effect, we agree with the company's use of a waning assumption (linear increase in the HR from year 5 to HR=1 at year 20). We have not changed the waning period in EAG preferred analysis, but note a longer waning period, or the removal of waning, might be appropriate with the more conservative constant HR estimate that we use,

4.2.6.2.2 Ruxolitinib extrapolation for RESPONSE and REPONSE-2 populations

The OS for ruxolitinib was modelled indirectly using the extrapolations of TTD excluding death described above, and extrapolations of pooled data for pre-discontinuation survival and post-discontinuation survival(see CS sections B.3.1.2 and B3.3.2). Pooled data were used because of the small number of deaths observed in the trial, both pre- and post-ruxolitinib discontinuation.

The fitted extrapolations for pre-discontinuation survival are illustrated in CS Appendix N.3. The company choose an exponential distribution for their base case, which had the best statistical fit, with alternative distributions assessed in scenario analysis. They included a constraint to ensure that the hazard of death was no less than that for members of the general population of the same age and gender mix, but this was only applied after the trial period. In response to clarification question B4, the company added an option in the model to include the general population constraint throughout the time horizon (CQ response Figure 1).

Extrapolations for post-discontinuation survival are presented in CS B.3.3.2 and Appendix N.4. Again, the company chose an exponential distribution, which had the best fit to the trial data and was considered clinically plausible by the company's experts. The general population mortality constraint was applied throughout the time horizon. The resulting extrapolation is illustrated in CS Figure 44.

The STM model combines the extrapolations for time to ruxolitinib discontinuation, pre-discontinuation survival and post-discontinuation survival to estimate OS for ruxolitinib.

EAG comments on the ruxolitinib OS extrapolation (primary analysis):

- The use of a competing-risk framework to estimate TTD, and subsequently OS for ruxolitinib is appropriate for the STM structure of the company's primary analyses. We agree with the pooling of data from the RESPONSE and RESPONSE-2 trials for estimation of pre- and post-discontinuation survival extrapolations, given the small numbers of deaths observed. However, this means that the comparative results for the patients with and without splenomegaly may not fully reflect survival differences between these subgroups.
- The company's base case extrapolation for pre-discontinuation survival is not adjusted for general population mortality during the trial period. This results in a

lower mortality rate during the first five years of ruxolitinib treatment than for people of the same age and gender mix in the general population, which is not plausible. For the EAG preferred analysis we use the general population mortality constraint for pre-discontinuation survival throughout the time horizon. This results in mortality rates prior to discontinuation of ruxolitinib that are the same as for the general population, so he model is not sensitive to the distribution for extrapolation of pre-discontinuation survival.

- The model is somewhat sensitive to the distribution used for post-discontinuation survival. The company use an exponential extrapolation in their base case, which provides a reasonable fit to the trial data.

4.2.6.2.3 OS extrapolation for MAJIC-PV population

The OS for BAT was extrapolated directly from reconstructed OS KM data from the MAJIC-PV manuscript using a Weibull distribution (see CS B.3.3.3 and Appendix N.8).

The OS for BAT was derived by applying a relative treatment effect to the ruxolitinib OS extrapolation. In the base case analysis, the company used data from the MAJIC-PV trial, because comparative evidence from the RESPONSE and RESPONSE-2 trials was confounded by cross-over from the BAT arm to ruxolitinib.

The same estimates of the treatment effect were used in both STM and PSM models, and for all three trial populations (MAJIC-PV, RESPONSE and RESPONSE-2). See Table 22 below for the HR estimates used in the company's base case and scenario analyses.

The OS for ruxolitinib was derived from the OS for BAT by applying the time-varying treatment effect to the BAT OS, see discussion in section 4.2.6.2.1 above. Note that these HRs are the inverse of those used in the primary analysis, as the OS extrapolation for ruxolitinib in the MAJIC-PV population analysis was derived from the BAT OS extrapolation (in contrast with the primary analysis, where the OS extrapolation for BAT was estimated from the ruxolitinib OS extrapolation). The same gradual linear waning of the treatment effect from year 5 to year 20 employed in the primary analysis was also used in the MAJIC-PV population analysis.

4.2.6.3 Key complications (events)

The company incorporates five key complications as events in the economic model: TE, progression to AML or MDS, progression to MF, haemorrhage, and NMSC.

The incidence rates of key complications while on ruxolitinib were calculated based on the numbers of events reported in the RESPONSE, RESPONSE-2 and MAJIC-PV trials, adjusted by the duration of exposure to ruxolitinib or total follow-up time. CS Table 24 reports the exposure-adjusted incidence rates for patients on ruxolitinib for the three trials. Trial-specific data for the relevant population were used, where available.

The incidence of events whilst on BAT were estimated by applying a treatment effect in the form of incidence rate ratios (IRR) to the baseline incidence rate of events on ruxolitinib. To account for the small number of events and varying follow-up durations, the IRRs were estimated using the pooled number of events from the RESPONSE, RESPONSE-2 and MAJIC-PV trial. The IRRs used for each of the five events are reported in CS Table 25.

The company notes that none of the trials were powered to estimate the incidence of these key complications. They also note that assumptions were required for missing data, not reported for specific trials (see CS B.3.3.8).

EAG comment on estimated event rates for key complications (events)

- The incidence of key complications in the ruxolitinib arm was based on reported rates per patient year of exposure from the three trials. We note that these rates are fixed across the time horizon and are not adjusted for age.
- The incidence of the key events while patients were on BAT was estimated from relative rates (IRRs) from pooled trial data. This resulted in lower incidence of MF, TE and haemorrhage, and higher incidence of non-melanoma skin cancer while patients were on ruxolitinib than on BAT. There was very little difference between the treatments in estimated rates of conversion to AML/MDS.
- The company reported scenarios excluding the impact of the individual key events, and excluding all events in CS Appendix P. This showed limited impact on the ICERs.

4.2.6.4 Therapeutic phlebotomy

The rate of therapeutic phlebotomy for patients on ruxolitinib was derived from each of the three trials and applied to the respective analysis population: ███, ███, and ███ for RESPONSE, REPONSE-2, and MAJIC-PV, respectively (see CS section B.3.3.9). The unpublished MAJIC-PV manuscript did not report exposure time, and a total number of phlebotomy procedures was reported during the entire study period. as opposed to during

ruxolitinib treatment only. Therefore, total follow-up time estimated from the pseudo-IPD for OS was used. As with complications, the number of phlebotomy procedures across all trials and the exposure time for ruxolitinib and BAT were pooled to calculate a treatment effect IRR of ██, which was applied to the rates for ruxolitinib to acquire the rate of phlebotomy for patients on BAT.

4.2.6.5 Adverse events

The model included adverse events occurring at a rate of ≥5 per 100 patient-years of exposure and at a rate of ≥3 per 100 patient-years of exposure in either arm of the RESPONSE and REPONSE-2 trials, respectively. CS Table 23 reports the pooled exposure-adjusted rates of 67 AEs. All grades of AEs were included in the model, with Grades 1 and 2 having a lesser impact than Grades 3 and 4. In the primary analysis, the rates of AEs from the RESPONSE and RESPONSE-2 trials were pooled for both patients with and without splenomegaly. The unpublished MAJIC-PV manuscript only reports AE categories experienced by ≥10% of patients and does not have data regarding Grade 1 or 2 AEs nor on the duration of exposure; the analysis for this population therefore used the same incidence of AEs used for the primary analysis.

4.2.7 Health related quality of life

4.2.7.1 Systematic literature review for utilities

The company identified two studies that reported utility estimates for people with PV from their systematic review of literature on health-related quality of life (CS Appendix H). The study by Lelonek et al. (2018) reported EQ-5D-3L values with a UK tariff for 102 people with PV and the JAK2V617F mutation.[48] Mean (SD) utility scores were the same for people with and without aquagenic pruritus: 0.8 (0.1) (see CS Appendix Table 40).

The second study was the Scottish Medicines Consortium (SMC) review for ropeginterferon (2022).[41] This included EQ-5D-3L utility scores for 1,142 adults with PV from the PROUD-PV and CONTINUATION-PV studies. Mean (SD) utility scores were cited of 0.881 (0.152) for 892 people with JAK2<50 and 0.876 (0.148) for 250 people with JAK2≥50 (CS Appendix Table 41). The company state that these data were collected from an international study which did not include UK patients, and that the value set was not reported. It is therefore not clear that these estimates would meet NICE reference case requirements.

Neither study was specific to the population of interest in this current appraisal. So, as utility data was available from the RESPONSE and RESPONSE-2 trials, the company did not use the above estimates in the economic model. The EAG agree with this judgement.

See CS Table 28 for a summary of utility values used in the economic model.

4.2.7.2 Study-based health related quality of life

Treatment specific utility values were derived from individual patient data from the trials, using regression analysis, with treatment and baseline values as covariates (see CS B.3.4.3 and company response to clarification question B9). For their base case, the company use utility estimates for condition-specific preference based utility instrument (the MF-8D),[43] derived from EORTC QLQ-30 and MPN-SAF data from the RESPONSE trial.

The MF-8D was developed for use in myelofibrosis and uses three items from the EORTC QLQ-30 and five from the MF Symptom Assessment Form (MF-SAF). The MF-SAF is similar to the MPN-SAF, but with differences in the wording of two items used in the MF-8D. The company therefore had to make the following assumptions to use the MF-8D for the PV population in the RESPONSE trial:

- That "pain under ribs on the left side" in the MF-SAF is equivalent to "abdominal pain" in the MPN-SAF
- And that "bone or muscle pain" in the MF-SAF is equivalent to "bone pain" in the MPN-SAF.

The company justify their preference for the MF-8D on the basis that the EQ-5D is not appropriate for capturing the impact of PV on health-related quality of life (CS section B.3.4.1). Their argument is based on:

- Published psychometric analysis which indicates that the EQ-5D and EORTC QLQ-C30 instruments to do capture the key symptoms of myelofibrosis.[43 49]
- Precedent from two NICE MF appraisals (TA386 and TA756), in which the NICE committees accepted use of the condition-specific MF-8D.[15 44]
- The similar nature of symptoms for PV and MF, including fatigue, early satiety, abdominal discomfort, inactivity, concentration, night sweats, itching and bone pain. EAG clinical expert advisors agreed that the symptoms of MF and PV are generally similar in nature but vary in severity.

The company also report results from an 'exploratory' psychometric analysis of EQ-5D-5L (mapped to UK EQ-5D-3L utility values using the NICE recommended method)[42 50 51]) and MPN-SAF data from RESPONSE-2 (CS B.3.4.1). Further information about the exploratory psychometric analysis was provided in a PowerPoint report in response to clarification question B8. This analysis included a comparison of ceiling effects, item correlation and standardised measures of change from baseline for the EQ-5D and MPN-SAF TSS.

The company also report a scenario analysis based on EQ-5D-5L data from the RESPONSE-2 trial (CS section B.3.4.1). For this analysis, UK 3L utility values were derived using the algorithm developed by Hernández Alava et al. 2020, as currently recommended by NICE.[42 50 51]

Health state utilities are appropriately adjusted in the model for aging of the population, using UK general population utility data (Hernandez et al. 2022).[50]

4.2.7.3 Disutility for BAT substates

In the primary analysis, reductions in utility values and disutilities are assigned for the BAT sub-health states as follows:

- From baseline to 1st BAT sub-health state: ▮▮▮▮
- From baseline to 2nd+ BAT sub-health state: ▮▮▮▮
- No treatment sub-health state: -0.05.

The higher disutility for the no treatment sub-health state is in line with the greater decline in health for patients with high-risk PV who are not on treatment.

4.2.7.4 QALY loss associated with key events

The QALY loss for reduced utility associated with key complications were calculated based on estimates of disutility and life expectancy derived from the literature (CS Table 27). In response to clarification question B10, the company states that sources used to calculate these QALY losses were not derived by systematic review.

The EAG noted in clarification B12 that although the QALY losses associated with key events include utility lost during expected survival following an event, the QALY losses do not include QALY loss for shortened life expectancy due to an event. The company stated that extrapolation of overall survival beyond the observed trial period implicitly accounts for the increase in death caused by a key complication; incorporating years of life lost due to an

90

event could result in double counting. There is no possibility of determining the proportion of deaths due to a key event or due to other reasons, as overall survival is modelled directly for an average cohort and extrapolated over time, regardless of the cause of death.

For venesections, the company assume a QALY loss of -0.000103 per procedure, based on a decrement in utility of -0.037 procured from Matza et al. 2013 with the assumption that the decrement lasts for one day.[52] The company have confirmed in clarification response C3 the error in the company submission regarding the QALY loss associated with phlebotomy: the correct value of -0.000103 is implemented in the model.

4.2.7.5 QALY loss associated with adverse events

The impact of adverse events on HRQoL is not included in utility values but is captured in the model separately. The health disutility of an adverse event was based upon the health utility decrement and the duration of impact on quality of life of that particular adverse event. The company did not implement any health disutilities for Grade 1 or 2 adverse events, stating that this would simplify the model. CS Table 26 reports the disutilities and durations for the 36 categories of Grade 3 and 4 adverse events used in the model. Data for these adverse events were taken from values used in previous NICE appraisals and from the literature. For Grade 3 or 4 adverse events which no data could be sourced, the company assumed a disutility of -0.075 for a duration of seven days, based on results used in NICE TA772.[53]

EAG comments on health-related quality of life

- This provides some evidence in favour of the MF-8D, including greater responsiveness and lower susceptibility to ceiling effects.
- However, the MF-8D was not developed for use in PV, and the company had to make assumptions to substitute the PV symptom score for the myelofibrosis symptom score used in the MF-8D. There is also a lack of direct evidence validating the EQ-5D and MF-8D in a PV population.
- EAG clinical experts advised that the MPN-SAF TSS is mostly used in MF as that is the most symptomatic myeloproliferative disorder, but as there is extensive symptom overlap between MF, essential thrombocythemia and PV, they consider that the instrument would capture PV symptoms.
- We use EQ-5D utilities in the EAG preferred analysis. This follows the NICE preference for use of the EQ-5D when available from relevant clinical trials, as this provides consistency across NICE appraisals. There is some evidence in

91

favour of the MF-8D measure, but also uncertainty about its transferability from MF to PV.

- There is uncertainty regarding the accuracy of the QALY losses associated with key events, which do not consider the QALY loss associated with years of life lost. There is scope for further analyses regarding the QALY losses used, and whether more conservative QALY losses should be implemented to account for the lack of data regarding the potential decrease in life expectancy following a key event.

4.2.8 Resources and costs

4.2.8.1 Systematic literature review of costs and healthcare use

The company report the results of their review of cost and resource use data in CS Tables 45 and 46. They included three studies in their review, including the Scottish Medicines Consortium appraisal of ropeginterferon, but the company conclude that this data was not usable, because the population from which the data was sourced was not defined (SMC 2022).[38-40] The other two UK based studies were not used either, as one was considered too old and the other did not state the cost year.

4.2.8.2 Drug acquisition and administration

CS Table 29 reports the dosing schedule and costs of drugs used in the model. The 4-weekly treatment costs for ruxolitinib used in the model were ▮▮▮▮ and ▮▮▮▮ for the primary analysis with and without splenomegaly, respectively. As the unpublished MAJIC-PV trial manuscript does not contain data on dosage distribution, the company assumed a treatment cost for ruxolitinib in the MAJIC-PV population of ▮▮▮▮, the same as the RESPONSE-2 trial. These prices for ruxolitinib are using the current Patient Access Scheme (PAS) discount for myelofibrosis of ▮▮▮▮. ▮▮ Ruxolitinib is administered orally; there are no associated administration costs. The dosage for ruxolitinib in the model is based upon the RESPONSE and RESPONSE-2 trials for both primary and MAJIC-PV analyses.

For BAT, a 4-weekly treatment cost of £226.48 was used in the model for both primary and MAJIC-PV analyses. This was based on the distribution of treatments in the BAT arm of the MAJIC-PV trial (CS Table 20), but as pipobroman and radioactive phosphorus are no longer in use in England and Wales, they were excluded from the BAT composition in the model. Ruxolitinib, used in combination for a small number of patients in the MAJIC-PV BAT arm,

was also excluded. Unit costs for the included BAT medications are shown in CS Table 29. We note that the company use the cost for a pegylated derivative of interferon-alfa, as this is now routinely used in NHS practice.

All patients on interferon-alfa require training on how to self-inject the drug, which involves one or two visits with a nurse or GP. However, according to clinical experts, approximately 5%-10% of patients with PV using interferon-alfa require continuous help from a nurse to administer the injection; the remaining patients on interferon-alfa are able to self-inject once trained and do not incur administration costs. Therefore, the model implements a one-off cost of £24.71 for patients on BAT to include the cost of training and district nurse visits.

EAG comments on drug acquisition and administration
- Clinical experts advising the EAG have noted that the majority of patients would continue to be treated with interferon-alfa or hydroxycarbamide (despite being resistant or intolerant to the treatment). Anagrelide and busulfan are seldom prescribed. Approximately 10-15% of patients resistant or intolerant to hydroxycarbamide would have no other suitable alternative.
- We have also been advised that the majority of patients with PV on interferon-alfa would self-administer the drug, but between 2-10% would require on-going nurse help for injection.

4.2.8.3 Patient management and monitoring

There were no UK cost studies or NICE appraisals for PV identified in the company's economic SLR. Therefore, resource utilisation data was obtained from questionnaires completed by five UK clinical experts with experience in PV. The clinicians provide estimates for the management and monitoring of PV over three time intervals: 0-6 months, 7-12 months, and 13+ months of treatment. CS Table 32 provides the estimated resource use and unit cost per cycle for the different resource categories; the same resource use and costs were used for both primary and MAJIC-PV analyses.

The management and monitoring costs used in the model per cycle for patients on ruxolitinib were estimated to be ▇▇▇, ▇▇▇, and ▇▇▇ for 0-6 months, 7-12 months, and 13+ months of treatment, respectively. The corresponding costs used in the model for patients on BAT were ▇▇▇, ▇▇▇, and ▇▇▇. In the primary analysis where the BAT state is partitioned, patients in the "no treatment" sub-health state incurred an assumed cost of ▇▇▇ per cycle, twice the cost of patients on BAT, and was fixed across all time intervals. This sub-health

state was assumed to have a higher cost to represent the worsening of PV and a subsequent increase in management and monitoring when patients are no longer on treatment. The model also included a cost of £316 per therapeutic phlebotomy, and a cost of £6,774 for end of life care.

4.2.8.4 Adverse events and key events

The unit costs for Grade 3 and 4 adverse events are provided in CS Table 35, taken from the NHS reference costs 2020/21. Note that only 36 categories of the 67 adverse event categories were reported to have at least one Grade 3 or 4 event in either arm of the trials. The cost for the management of Grade 1 and 2 adverse events were assumed to be equivalent to the cost of two GP e-consultations at a total of £78.46.

CS Table 33 reports the management costs assumed for each of the five key events (TE, AML/MDS, MF, NMSC and haemorrhage). The company have noted in clarification response B13 and B14 the errors in costs in the table: the cost for the management of a TE event used in the model is £1,865, and the cost for a haemorrhage event is £2,023.

The cost for the management of a TE event, £1,865, was based upon the grade levels of events, unit costs, and the distribution of TE events in the ruxolitinib arms of RESPONSE and RESPONSE-2. CS Table 33 contains the unit costs from the NHS reference costs 2020/21 for Grade 3 and 4 TE events. The cost for an emergency department visit was assumed for the management of a Grade 1 or 2 TE event. The company have noted in clarification response B13 that the cost of an emergency department visit is stated incorrectly as £182 in the CS; the correct cost is £297.

The cost for the management of AML/MDS implemented in the model, £44,903, was also used in NICE TA386 and NICE TA756, and was taken from the results of a probabilistic decision model in AML by Wang et al. 2014.[15][44][54] The cost is the median value of the range of reported costs in Wang et al. 2014, who estimated 5-year medical costs for the management of AML in the UK.

The management cost for MF assumed in the model was £63,920. The costs for managing intermediate-2/high-risk MF which occurred in 57.3% of patients with MF was determined from TA386 using the total costs for ruxolitinib, £128,403, and BAT patients, £36,095.[15] The company were unable to find data on the management cost for the remaining 42.7% of patients with low/intermediate-1 MF, and so they assumed a cost of £72,190, double the

94

cost of intermediate-2/high-risk MF in BAT patients. The company note that patients with low/intermediate-1 MF generally have a more favourable prognosis than patients with intermediate-2/high-risk MF, and will consequently have an increased duration of treatment, leading to higher overall resource use.

The management costs for NMSC and bleeding/haemorrhaging events used in the model were £1,058 and £2,023, respectively. The cost for an NMSC event was based on results in Vallejo-Torres et al. 2013, whilst the cost for a major haemorrhaging event was based on Crathorne et al. 2018; the management cost for a minor bleed was assumed to be equivalent to the cost of one emergency department visit, £297.[55][56]

EAG comments on resources and costs

- Clinical advice to the EAG was that in addition to an emergency department visit, patients with a grade 1 or 2 thromboembolic event would need a D-Dimer test and an ultrasound doppler scan. We include an additional cost for these tests in the EAG preferred analysis.

- In addition, interim or long-term treatment with warfarin or an oral anticoagulant would be initiated for some patients after a grade 1 or 2 thromboembolic event. We therefore include the cost of a single dose of an anticoagulant, as stated in the NICE guideline NG158.[57] The effects of this cost change are discussed in section 5.3.3.2. We have not included the costs (or benefits) of thromboprophylaxis in our scenario analysis, as this would be difficult to estimate. However, we note that the impact of grade 1 or 2 thromboembolic events are likely to be underestimated in the model.

- Other estimated costs for adverse events were considered reasonable. It was noted that patients often consult with clinical nurse specialists for drug-related adverse effects, but the assumption of 1 or 2 GP online consultations was considered to be reasonable for the cost calculations. The company's use of a higher cost for low/intermediate-1 MF than for intermediate-2/high-risk MF was also considered reasonable due to the longer duration of treatment (median survival approximately 5-8 years and 1-3 years respectively).

5 COST EFFECTIVENESS RESULTS

5.1 Company's base case cost-effectiveness results for the primary analysis

The company report the deterministic base case results from their primary STM model in CS Table 39 for the licensed populations with and without splenomegaly (reproduced in Table 23 below). These and other results in this report use the current Patient Access Scheme (PAS) price for ruxolitinib (█████ price discount) agreed as part of the MF submission to NICE TA386,[15] with list prices used for all other drugs. Results with confidential discounts for comparator and concurrent medications are provided in a separate confidential addendum to this report.

Table 23 Company base case results: primary analysis

Treatment	Total			Incremental			ICER (£/QALY)
	Cost	LYG[a]	QALYs	Cost	LYG[a]	QALYs	
Licensed population with splenomegaly (RESPONSE trial population)							
BAT	£92,017	9.28	6.97	-	-	-	-
Ruxolitinib	███	███	███	███	2.17	███	███
Licensed population without splenomegaly (RESPONSE-2 trial population)							
BAT	£86,809	10.46	7.80	-	-	-	-
Ruxolitinib	███	███	███	███	1.79	███	███
Source: Reproduced from CS Table 39.							
BAT best available therapy; LYG life years gained; QALYs quality-adjusted life years; ICER incremental cost-effectiveness ratio.							
[a] Note: life years gained are not discounted.							

The base case results for the primary analysis show that for the licensed population without splenomegaly, ruxolitinib offers a mean QALY gain of ███ for an additional mean cost of ██████ compared with BAT, producing an ICER of █████ per QALY gained. For the licensed population with splenomegaly, ruxolitinib provides a QALY gain of ███ for an additional cost of ██████ against BAT, which results in an ICER of █████ per QALY gained.

5.1.1 Deterministic sensitivity analyses for the company's base case for the primary analysis

The company report deterministic sensitivity analysis results for the ten most influential parameters in CS Figure 53. The ranges of variation for the input parameters were based on 95% confidence intervals where available, or a range of +/- 20%. The company's results indicate that the assumptions regarding the treatment effect for OS are the main drivers of

the model results for the primary analysis, increasing the ICER to ▮▮▮▮ and ▮▮▮▮ per QALY for the licensed population with and without splenomegaly, respectively. The discount rates for both cost and benefits and assumptions regarding utility values also have a notable impact on the ICER for the primary analysis.

5.1.2 Scenario analyses for the company's base case for the primary analysis

The company consider almost 100 scenarios for the primary analysis (see CS Appendix P) and report the top 20 most impactful scenarios in CS Figure 54.

Licensed population with splenomegaly (RESPONSE population)

Changing the source of the treatment effect (HR OS) from the MAJIC-PV trial to the retrospective analysis of Spanish registry data (Alvarez-Larrán et al. 2022) had the largest impact on the ICER, increasing the ICER to ▮▮▮▮ per QALY, whilst limiting the treatment effect to 5 years has the second-largest effect, causing the ICER to rise to ▮▮▮▮ per QALY.[37] Of the 20 scenarios provided in the CS, the top seven scenarios that increase the ICER the most involve the source of the treatment effect, treatment effect waning, and the time horizon. Using the treatment effect from ITC comparison with GEMFIN results in the lowest ICER per QALY, at ▮▮▮▮. We note that the CS did not report results for the scenario with a constant HR OS from the MAJIC-PV trial, but this was provided in response to clarification question B2. This scenario increased the company's base case ICER for RESPONSE population to ▮▮▮▮ per QALY.

Licensed population without splenomegaly (RESPONSE-2 population)

Limiting the treatment effect to 5 years resulted in the highest impact on the ICER, which increases to ▮▮▮▮ per QALY; the second-largest effect arose by implementing Alvarez-Larrán et al. 2022 as the source of the treatment effect, giving an ICER of ▮▮▮▮ per QALY.[37] As with the licensed population with splenomegaly, the top seven scenarios causing the highest increase in ICERs involved the source of treatment effect, treatment effect waning, and the time horizon. Also in line with the licensed population with splenomegaly, applying the ITC treatment effect from the comparison with GEMFIN rather than MAJIC-PV gives the greatest reduction in the ICER at ▮▮▮▮ per QALY. The ICER for the scenario with the constant MAJIC-PV HR for the RESPONSE-2 population was ▮▮▮▮ per QALY (company response to clarification question B2).

97

5.1.3 Probabilistic sensitivity analysis for the company's base case for the primary analysis

The company conducted a probabilistic sensitivity analysis (PSA) with input parameter distributions as presented in CS Table 37. They used appropriate probability distributions for the different parameters. An arbitrary SE of 10% was assumed where the SE was not reported, namely for the QALY loss for key events, management costs and end of life cost.

The results from 2,000 iterations are reported in CS Table 41, and CS Figure 52 illustrates the extent of uncertainty around the results with cost-effectiveness scatterplots and cost-effectiveness acceptability curves (CEACs). The EAG confirm that the probabilistic results for the licensed population either with or without splenomegaly are similar to the deterministic results. The estimated probability that ruxolitinib meets a cost-effectiveness threshold of £30,000 per QALY gained at the current PAS price for ruxolitinib ▮▮ for both subgroups, with and without splenomegaly.

5.2 Company's base case cost-effectiveness results for the MAJIC-PV population

The company reports the results for the MAJIC-PV population in CS Table 42, reproduced in Table 24 below. This shows an estimated QALY gain of ▮▮ and additional cost of ▮▮▮ for ruxolitinib in comparison with current clinical management, resulting in an ICER of ▮▮ per QALY gained.

Table 24 Company base case results: MAJIC-PV population

Treatment	Total			Incremental			ICER (£/QALY)
	Cost	LYG*	QALYs	Cost	LYG*	QALYs	
BAT	£83,317	8.02	6.11	-	-	-	-
Ruxolitinib	▮▮	▮▮	▮▮	▮▮	1.63	▮▮	▮▮
Reproduced from CS Table 42. Best available therapy; LYG life years gained; QALYs quality-adjusted life years; ICER incremental cost-effectiveness ratio. *Note: life years gained are not discounted.							

5.2.1 Deterministic sensitivity analyses for the company's base case for the MAJIC-PV population

The company illustrate the results of the ten most influential parameters from their deterministic sensitivity analyses in CS Figure 56. As with the primary analysis, the company's results show that the model is most sensitive to the treatment effect for OS, with the ICER increasing to ▮▮▮ per QALY at the upper limit for the HR in the second time

98

period (year ▮). The discount rates for costs and benefits are also influential parameters for the MAJIC-PV population, as well as the hazard rate for the time to treatment discontinuation for ruxolitinib.

5.2.2 Scenario analysis for the company's base case for the MAJIC-PV population

The company report the results of the top 20 most impactful scenarios in CS Figure 56. Restricting the treatment effect to 5 years has the largest effect on the results, increasing the ICER to ▮▮▮▮ per QALY, and implementing the treatment effect for OS reported by Alvarez-Larrán et al. 2022 produced the next-highest ICER of ▮▮▮▮ per QALY.[37] Again, in line with the primary analysis, the most influential scenarios involve the treatment effect for OS and treatment effect waning, with the greatest reduction in the ICER obtained by from the ITC comparison with GEMFIN ▮▮▮▮ per QALY). The scenario with the constant HR estimated from the MAJIC-PV trial increase the ICER to ▮▮▮▮ per QALY (company response to clarification question B2).

5.2.3 Probabilistic sensitivity analysis for the company's base case for the MAJIC-PV population

Probabilistic results for the MAJIC-PV population are provided in CS Table 44 and CS Figure 55. The EAG confirm that the probabilistic results for the MAJIC-PV population are similar to the deterministic results. As with the base case results, the probability that the ICER is below £30,000 per QALY gained is ▮.

5.3 Model validation and face validity check

5.3.1 Company's model validation

The company state their approach to model validation in CS Section B.3.13. They report that two advisory board meetings were held with five clinical experts with experience in the management of patients with PV resistant or intolerant to hydroxycarbamide.[58][59]

The EAG note that the first advisory meeting, conducted on 24th June 2022, comprised only four clinical experts; however, the second cited advisory meeting took place over two dates (28th July 2022 and 8th August 2022) with five experts present.[58][59] Four of the five clinical experts who attended the advisory meetings are authors of the MAJIC-PV trial.

The model structure and appropriateness to the decision problem were discussed and validated with the clinical experts in these meetings, as well as the validity of model inputs such as costs and utilities. The company also report that a health economist, not involved in

99

the development of the model, reviewed the model for coding errors, inconsistencies, and plausibility of inputs, and also subjected the model to stress testing of extreme scenarios to detect modelling errors.

The company note the following points:

- Long term predictions could not be compared against external data as long term data for the patient population are not available.
- Predicted life years for the licensed population without splenomegaly was higher compared to the licensed pop with splenomegaly, despite using different model structures and inputs. This is in line with clinical expectations.
- Predicted life years for the MAJIC-PV population were lower compared to estimates from the primary analysis for the RESPONSE and RESPONSE-2 trial populations. This reflects the poorer prognosis of the MAJIC-PV population.
- Prediction for the MAJIC-PV population also aligns with that observed in Alvarez-Larrán et al. 2022.[37]

5.3.2 EAG model validation

5.3.2.1 EAG verification procedures

The EAG conducted a series of quality checks on the company model, assessing its transparency and validity. A range of tests were performed to verify model inputs, calculations, and outputs:

- Cross-checking all parameter inputs against values reported in the CS, model, and cite sources
- Checking all model outputs against results stated in the CS, including the base case, PSA, DSA, and company scenarios for both the primary and MAJIC-PV population analyses
- Checking the individual formulae within the model
- Manually running scenarios and verifying model outputs against results reported in the CS and appendices for the DSA and scenario analyses
- Applying a range of extreme value and logic tests to check the plausibility of changes in results when parameters are changed
- Checking Visual Basic (VBA) code for errors, and re-running the code to ensure expected outputs were produced.

The model is well implemented and no coding errors were identified, however the EAG considers the failure to apply a general population mortality constraint to pre-discontinuation mortality within the 5-year trial period to be an error (see section 5.3.3.1 below).

The EAG identified several discrepancies between parameter values cited in the CS and the values used in the model (clarification questions B5, B11 to B18 and C3). The company confirmed that in all cases these related to errors in the description of model inputs in the CS, and that the correct values had been used in the model. Note also that the company confirmed that the columns in the table of scenario analyses in CS Appendix P are incorrectly labelled (clarification question C4).

5.3.2.2 Comparison of company extrapolations with trial and cohort data

Figure 3 and Figure 4 below present the model predictions for overall survival and time to treatment discontinuation for ruxolitinib and BAT for the licensed population with and without splenomegaly, respectively.
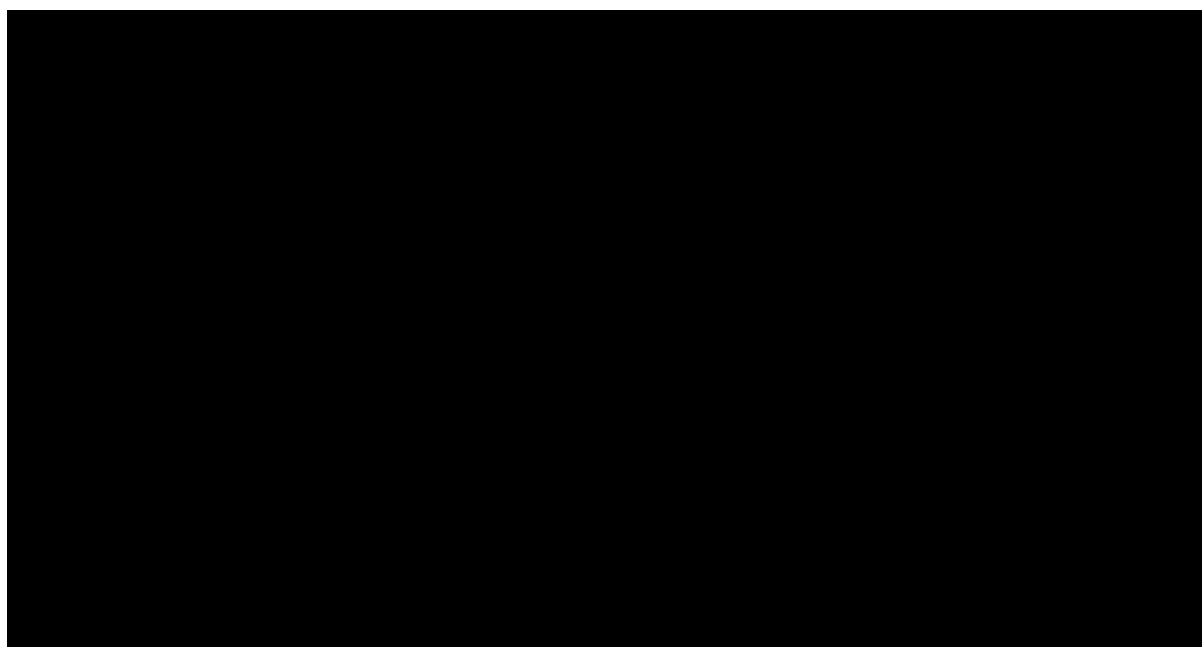
Figure 3 Predicted OS and TTD for ruxolitinib and BAT for the licensed population with splenomegaly

Abbreviations: OS: overall survival; TTD: time to treatment discontinuation; BAT: best available therapy; KM: Kaplan-Meier; Gen pop: general population mortality; ICER: incremental cost-effectiveness ratio; Rux: ruxolitinib. Source: Reproduced from CS Appendix J Figure 14.

101
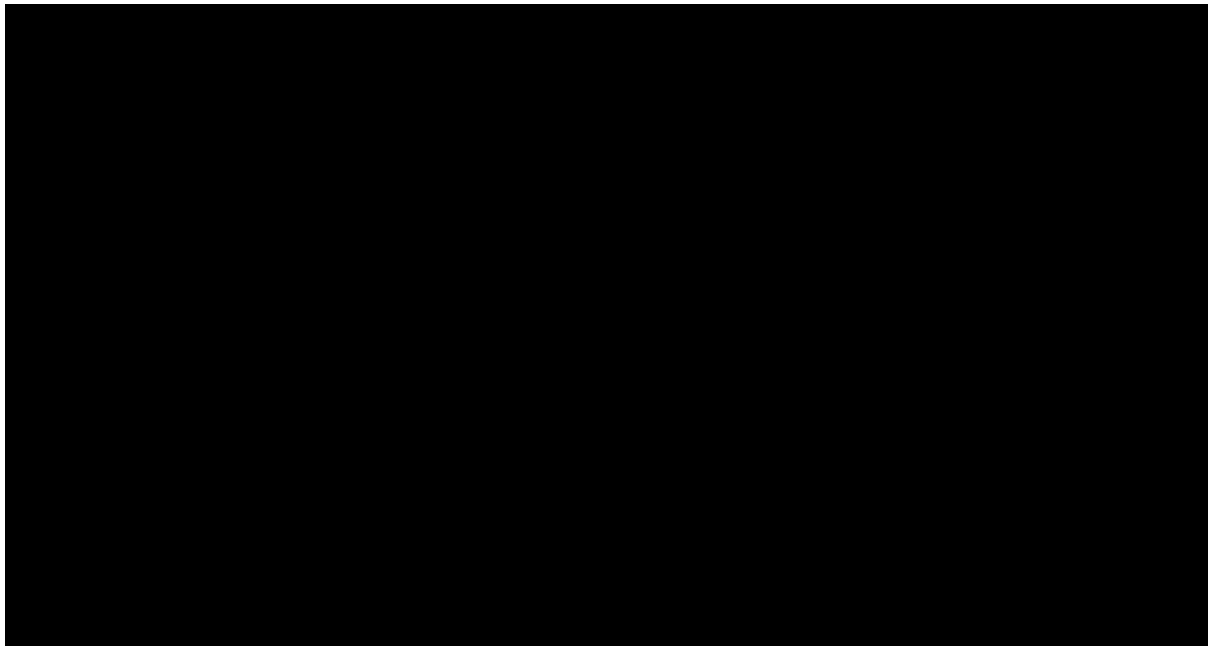
Figure 4 Predicted OS and TTD for ruxolitinib and BAT for the licensed population without splenomegaly

Abbreviations: OS: overall survival; TTD: time to treatment discontinuation; BAT: best available therapy; KM: Kaplan-Meier; Gen pop: general population mortality; ICER: incremental cost-effectiveness ratio; Rux: ruxolitinib. Source: Reproduced from CS Appendix J Figure 13.

EAG comments on extrapolation distributions

- The company selected an odds spline model with one knot for the extrapolation of TTD for ruxolitinib due to reasons other than death in the primary analysis. The company make note of the potential for spline models with more than one knot to overfit the data. The EAG opt for a standard parametric distribution, the Weibull distribution, in our preferred assumptions to remove the uncertainty around spline models and utilise a more conservative approach.

- The remaining distributions chosen by the company are deemed appropriate by the EAG. Scenario analyses showing outcomes of selected distributions for OS and TTD for the primary analysis and the MAJIC-PV population analysis are provided in section 6.1.

5.3.3 Corrections to the company model

5.3.3.1 General population mortality constraint for pre-discontinuation survival

In the company's analyses for the RESPONSE and RESPONSE-2 populations, pre-discontinuation survival for ruxolitinib is only adjusted for general population mortality after the 5-year period of trial observation, which results in better predicted survival while patients

102

remain on ruxolitinib than for people in the general population of the same age. The EAG raised this anomaly as a clarification question (B4), and the company provided an updated version of the model with an option to adjust pre-discontinuation survival for general population mortality over the entire time horizon. The ICERs for the RESPONSE and RESPONSE-2 populations with this adjustment were reported as a scenario analysis in Table 4 in the company's clarification response. We consider this a correction, as it is not plausible that people with PV would have better survival than the general population.

Full cost-effectiveness results for the company's primary base case analyses with the general population mortality correction applied are shown in Table 25 below. We use this correction in EAG additional in section 6.2. Note that as pre-discontinuation survival for ruxolitinib is only implemented in the primary analysis, the results for the MAJIC-PV population are unaffected.

Table 25 Company scenario analysis with the general population mortality constraint for pre-discontinuation survival: primary analysis

Treatment	Total			Incremental			ICER
	Cost	LYG[a]	QALYs	Cost	LYG[a]	QALYs	(£/QALY)
Licensed population with splenomegaly (RESPONSE trial population)							
BAT	£89,098	8.97	6.73	-	-	-	-
Ruxolitinib	■■■	■■■	■■■	■■■	2.20	■■■	■■■
Licensed population without splenomegaly (RESPONSE-2 trial population)							
BAT	£82,203	9.88	7.37	-	-	-	-
Ruxolitinib	■■■	■■■	■■■	■■■	1.87	■■■	■■■
Source: Company response to clarification question B4 and EAG analysis with company's model BAT best available therapy; LYG life years gained; QALYs quality-adjusted life years; ICER incremental cost-effectiveness ratio. [a] Note: life years gained are not discounted.							

5.3.3.2 EAG scenario analysis for the cost of a grade 1-2 thromboembolic event

The company assumed a cost for management of all Grade 1-2 thromboembolic events of £297, equivalent to the cost of one emergency department visit. However, EAG clinical expert advisers noted that a D-dimer test and a vascular ultrasound would also be required to investigate a suspected thromboembolic event, as well as a single low-dose of an anti-coagulant (as per the NICE guideline NG158).[57] For the EAG analysis, we include the cost of a laboratory D-dimer test at £6.79 (NG158),[57] a single dose of enoxaparin sodium at £8.79, (BNF 2022)[8] and a vascular ultrasound costing £96.99 (NHS Reference costs 2020/21)[60]. This results in a small reduction in the ICERs (see Table 27 below).

Table 26 EAG scenario analysis for cost of grade 1-2 thromboembolic event

Treatment	Total			Incremental			ICER
	Cost	**LYG[a]**	**QALYs**	**Cost**	**LYG[a]**	**QALYs**	**(£/QALY)**
Licensed population with splenomegaly (RESPONSE trial population)							
BAT	£92,035	9.28	6.97	-	-	-	-
Ruxolitinib	■	■	■	■	2.17	■	■
Licensed population without splenomegaly (RESPONSE-2 trial population)							
BAT	£86,849	10.46	7.80	-	-	-	-
Ruxolitinib	■	■	■	■	1.79	■	■
MAJIC-PV population							
BAT	£83,339	8.02	6.11	-	-	-	-
Ruxolitinib	■	■	■	■	1.63	■	■
Source: Company response to clarification question B4 and EAG analysis with company's model BAT best available therapy; LYG life years gained; QALYs quality-adjusted life years; ICER incremental cost-effectiveness ratio. [a] Note: life years gained are not discounted.							

5.3.4 EAG summary of key issues and additional analyses

The company summarise and justify assumptions in their primary and subgroup (MAJIC-PV population) economic analyses in CS Table 38. We highlight key areas of uncertainty and the rationale for additional EAG analyses in Appendix 9.5. Section 6.2 details the EAG's preferred assumptions and subsequent cost-effectiveness results. Additional scenario analyses are conducted on the EAG base case model in section 6.2.2.

6 EAG ADDITIONAL ANALYSES

6.1 Exploratory and sensitivity analyses undertaken by the EAG

Table 27 below shows cost-effectiveness results for selected company scenarios for the primary analysis for the licensed population with and without splenomegaly (RESPONSE and RESPONSE-2, respectively). As there are a large number of scenarios reported in CS Appendix P, we have selected scenarios that relate to key uncertainties and that have an impact on the ICERs.

Table 27 Selected scenarios applied to the company base case: primary analysis

Scenario	Treatment	RESPONSE			RESPONSE-2		
		Cost	**QALYs**	**ICER**	**Cost**	**QALYs**	**ICER**
Company base case	BAT	£92,017	6.97		£86,809	7.80	
	Ruxolitinib	■	■	■	■	■	■

Scenario	Treatment	Total costs	Total QALYs	ICER	Total costs	Total QALYs	ICER
HR OS: MAJIC-PV constant	BAT	£102,301	7.78		£94,479	8.52	
	Ruxolitinib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
HR OS: pooled RESPONSEtrials	BAT	£103,377	7.86		£95,125	8.58	
	Ruxolitinib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
HR OS: Alvarez-Larrán 2022	BAT	£105,234	8.01		£96,237	8.68	
	Ruxolitinib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
HR OS: matched GEMFIN (ITC)	BAT	£75,644	5.66		£77,734	6.95	
	Ruxolitinib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
No BAT partition	BAT	£94,485	7.04		£89,043	7.87	
	Ruxolitinib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
EQ-5D utilities	BAT	£92,017	6.47		£86,809	7.22	
	Ruxolitinib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Faster waning: 5 to 10 years	BAT	£98,816	7.50		£92,756	8.35	
	Ruxolitinib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Slower waning: 5 to 50 years	BAT	£86,097	6.50		£81,321	7.29	
	Ruxolitinib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Time horizon 30 years	BAT	£91,122	6.91		£86,368	7.77	
	Ruxolitinib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Ruxolitinib TTD lognormal	BAT	£94,803	7.18		£92,185	8.30	
	Ruxolitinib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Ruxolitinib TTD loglogistic	BAT	£93,096	7.05		£90,099	8.11	
	Ruxolitinib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Ruxolitinib TTD Weibull	BAT	£90,683	6.86		£88,983	8.00	
	Ruxolitinib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Ruxolitinib TTD hazard spline 1	BAT	£90,118	6.82		£85,402	7.67	
	Ruxolitinib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Ruxolitinib TTD Exponential	BAT	£85,860	6.48		£86,257	7.75	
	Ruxolitinib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Remove impact of key events	BAT	£56,318	7.03		£63,023	7.90	
	Ruxolitinib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Source: EAG analysis using company model and scenario analyses.

QALYs: quality-adjusted life years; ICER: incremental cost-effectiveness ratio; BAT: best available therapy; HR: hazard ratio; OS: overall survival; ITC: indirect treatment comparison; TTD: time to treatment discontinuation.

Figure 5 and Figure 6 below show the KM data with the company's choice of distribution for TTD for ruxolitinib due to reasons other than death in comparison with the selected scenario distributions from Table 27 above for the licensed population with and without splenomegaly.
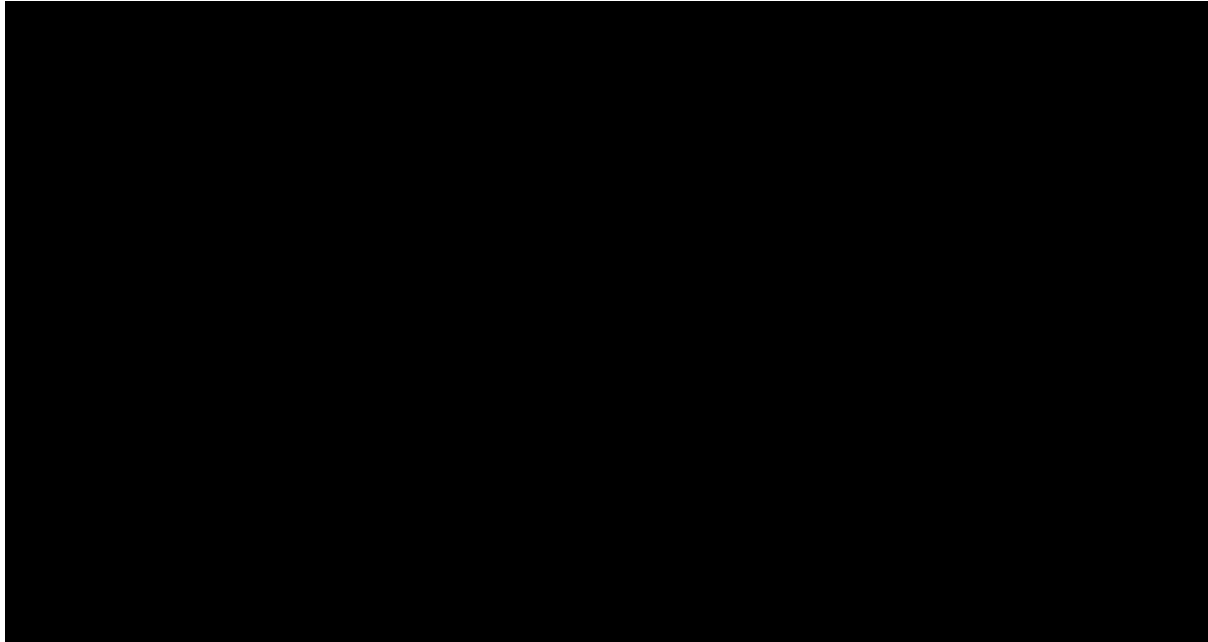
Figure 5 Comparison of selected scenario distributions for TTD for ruxolitinib for the licensed population with splenomegaly

Abbreviations: KM: Kaplan-Meier; TTD: time to treatment discontinuation; OS: overall survival.
Source: Reproduced from CS Appendix N Figure 18 using selected distributions.

Figure 6 Comparison of selected scenario distributions for TTD for ruxolitinib for the licensed population without splenomegaly

Abbreviations: KM: Kaplan-Meier; TTD: time to treatment discontinuation; OS: overall survival.

Source: Reproduced from CS Appendix N Figure 19 using selected distributions.

Table 28 below shows cost-effectiveness results for selected company scenarios for the MAJIC-PV population analysis. Again, from the many scenarios conducted by the company, we have selected scenarios that relate to key uncertainties and that have an impact on the ICERs.

Table 28 Selected scenarios applied to the company base case: MAJIC-PV population

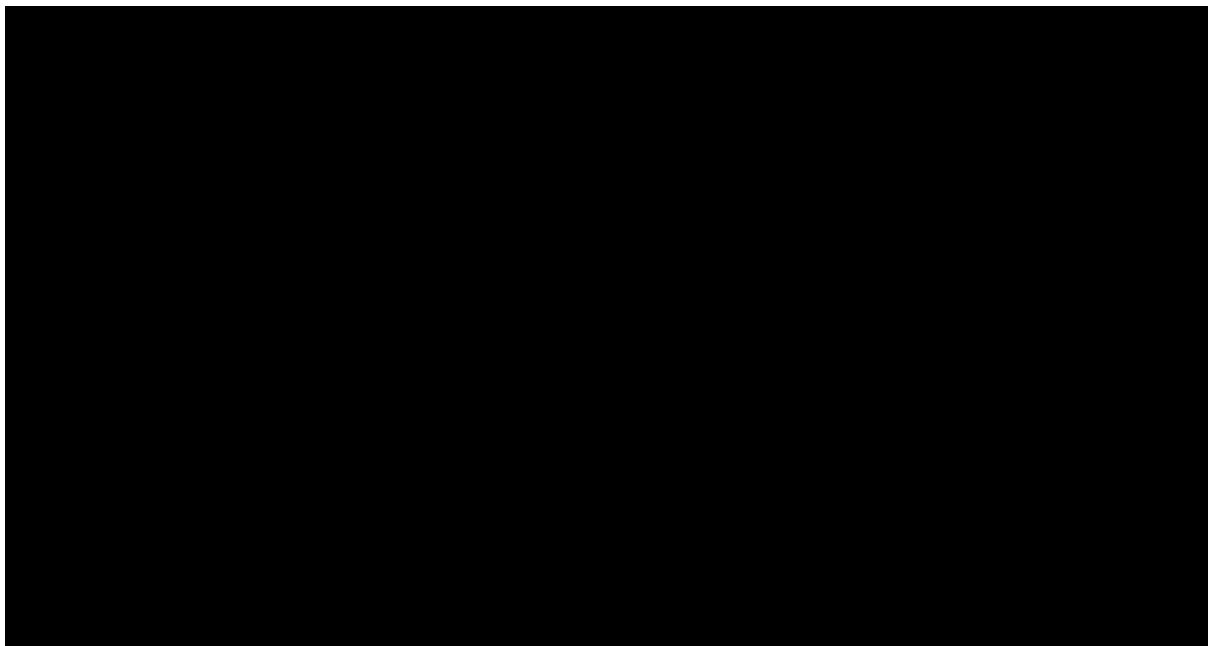| Scenario	Treatment	Cost	QALYs	ICER
Company base case	BAT	£83,317	6.11	
	Ruxolitinib	■■	■	■■
HR OS: MAJIC-PV constant	BAT	£83,317	6.11	
	Ruxolitinib	■■	■	■■
HR OS: Pooled RESPONSE-trials	BAT	£83,317	6.11	
	Ruxolitinib	■■	■	■■
HR OS: Alvarez-Larrán 2022	BAT	£83,317	6.11	
	Ruxolitinib	■■	■	■■
HR OS: matched GEMFIN (ITC)	BAT	£83,317	6.11	
	Ruxolitinib	■■	■	■■
EQ-5D utility values	BAT	£83,317	5.71	
	Ruxolitinib	■■	■	■■
Faster waning: 5 to 10 years	BAT	£83,317	6.11	
	Ruxolitinib	■■	■	■■
Slower waning: 5 to 50 years	BAT	£83,317	6.11	
	Ruxolitinib	■■	■	■■
BAT OS: lognormal	BAT	£101,095	7.43	
	Ruxolitinib	■■	■	■■
BAT OS: loglogistic	BAT	£94,943	6.97	
	Ruxolitinib	■■	■	■■
BAT OS: hazard spline 1	BAT	£98,348	7.23	
	Ruxolitinib	■■	■	■■
BAT OS: Gompertz	BAT	£70,476	5.13	
	Ruxolitinib	■■	■	■■
Time horizon: 30 years	BAT	£83,250	6.10	
	Ruxolitinib	■■	■	■■
Remove impact of key events	BAT	£57,187	6.18	
	Ruxolitinib	■■	■	■■
Source: EAG analysis using company model and scenario analyses. QALYs: quality-adjusted life years; ICER: incremental cost-effectiveness ratio; BAT: best available therapy; HR: hazard ratio; OS: overall survival; ITC: indirect treatment comparison.				

Figure 7 shows the KM data for overall survival for BAT in the MAJIC-PV population analysis in comparison with the company's chosen Weibull distribution and selected scenario distributions from Table 28 above.
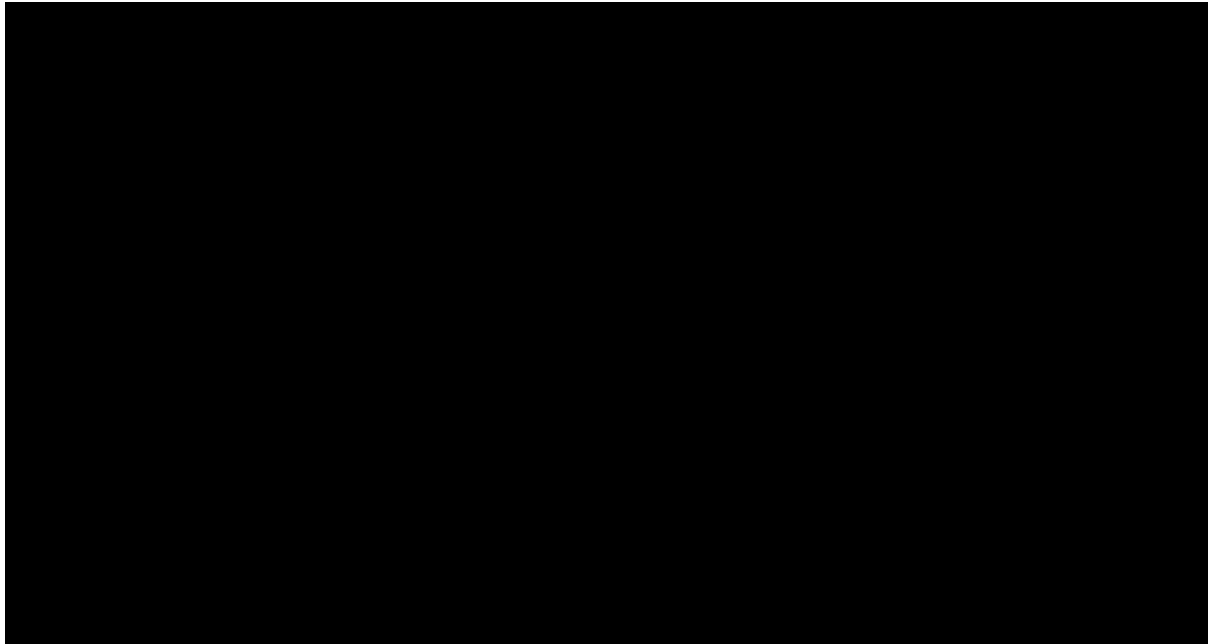
Figure 7 Comparison of KM with company base case distribution and selected scenario distributions for overall survival for BAT for the MAJIC-PV population analysis

Abbreviations: KM: Kaplan-Meier; BAT: best available therapy; OS: overall survival.
Source: Reproduced from CS Appendix N Figure 25 using selected distributions.

From the above tables, it is evident that the source of treatment effect for overall survival has a great impact on the ICER, with the exception of the hazard ratio derived from the ITC. As expected, reducing and increasing the treatment waning period also effects the ICER. Although the company implemented an extended time horizon of 46 years for patients starting in the model at age 66, a 30-year time horizon has minimal effect on the ICER.

6.2   EAG's preferred assumptions

Based on the critique of the company's model, the EAG have identified the following preferred model assumptions:

- Correction for general population mortality for pre-discontinuation survival in the primary analysis
- Weibull distribution for extrapolation of TTD for ruxolitinib due to reasons other than death in the primary analysis
- A constant hazard ratio derived from the MAJIC-PV trial for overall survival
- No partitioning of the BAT health state in the primary analysis
- EQ-5D utility values
- EAG estimated cost assumed for the management of Grade 1-2 thromboembolic events.

108

6.2.1 Results using the EAG preferred model assumptions

The results for this analysis for the three trial populations are shown in Table 29 below. We also report cumulative analyses for the three populations in Table 30, Table 31, and Table 32 below, showing the progression from the company's base case model to the EAG base case model by applying EAG preferred assumptions one at a time.

Table 29 EAG preferred analysis results

Treatment	Total			Incremental			ICER
	Cost	LYG[a]	QALYs	Cost	LYG[a]	QALYs	(£/QALY)
RESPONSE trial population (with splenomegaly)							
BAT	£100,281	9.90	7.02	-	-	-	-
Ruxolitinib	▮	▮	▮	▮	1.09	▮	▮
RESPONSE-2 trial population (without splenomegaly)							
BAT	£93,866	11.08	7.77	-	-	-	-
Ruxolitinib	▮	▮	▮	▮	0.91	▮	▮
MAJIC-PV trial population							
BAT	£83,339	8.02	5.71	-	-	-	-
Ruxolitinib	▮	▮	▮	▮	0.92	▮	▮
Source: EAG analysis using the company's model							
BAT best available therapy; LYG: life years gained; QALYs: quality-adjusted life years; ICER: incremental cost-effectiveness ratio.							
[a] Note: life years gained are not discounted.							

Table 30 Cumulative changes from the company base case model to the EAG preferred analysis: RESPONSE trial population (with splenomegaly)

Assumption	Treatment	RESPONSE		
		Cost	QALYs	ICER
Company base case	BAT	£92,017	6.97	
	Ruxolitinib	▮	▮	▮
+ General population mortality constraint	BAT	£89,098	6.73	
	Ruxolitinib	▮	▮	▮
+ Ruxolitinib TTD: Weibull	BAT	£87,837	6.64	
	Ruxolitinib	▮	▮	▮
+ HR OS: MAJIC-PV constant	BAT	£97,696	7.42	
	Ruxolitinib	▮	▮	▮
+ No BAT partition	BAT	£100,262	7.49	
	Ruxolitinib	▮	▮	▮
+ EQ-5D utilities	BAT	£100,262	7.02	
	Ruxolitinib	▮	▮	▮
+ Cost for Grade 1-2 TE events (EAG preferred analysis)	BAT	£100,281	7.02	
	Ruxolitinib	▮	▮	▮

Table 31 Cumulative changes from the company base case model to the EAG preferred analysis: RESPONSE-2 trial population (without splenomegaly)

Assumption	Treatment	RESPONSE-2		
		Cost	QALYs	ICER
Company base case	BAT	£86,809	7.80	
	Ruxolitinib	■	■	■
+ General population mortality constraint	BAT	£82,203	7.37	
	Ruxolitinib	■	■	■
+ Ruxolitinib TTD: Weibull	BAT	£84,052	7.54	
	Ruxolitinib	■	■	■
+ HR OS: MAJIC-PV constant	BAT	£91,411	8.23	
	Ruxolitinib	■	■	■
+ No BAT partition	BAT	£93,824	8.30	
	Ruxolitinib	■	■	■
+ EQ-5D utilities	BAT	£93,824	7.77	
	Ruxolitinib	■	■	■
+ Cost for Grade 1-2 TE events (EAG preferred analysis)	BAT	£93,866	7.77	
	Ruxolitinib	■	■	■

Table 32 Cumulative changes from the company base case model to the EAG preferred analysis: MAJIC-PV trial population

Assumption	Treatment	Cost	QALYs	ICER
Company base case	BAT	£83,317	6.11	
	Ruxolitinib	■	■	■
+ HR OS: MAJIC-PV constant	BAT	£83,317	6.11	
	Ruxolitinib	■	■	■
+ EQ-5D utilities	BAT	£83,317	5.71	
	Ruxolitinib	■	■	■
+ Cost for Grade 1-2 TE events (EAG preferred analysis)	BAT	£83,339	5.71	
	Ruxolitinib	■	■	■

6.2.2 Scenario analyses conducted on the EAG base case model

Table 33 and Table 34 below show selected scenario analyses applied to the EAG preferred analysis for the primary analysis (RESPONSE and RESPONSE-2 populations) and for the MAJIC-PV population analysis respectively. The scenarios included in these tables include company base case assumptions, as well as scenarios chosen to illustrate key uncertainties.

Table 33 Scenario analyses on the EAG base case model: primary analysis

Scenario	Treatment	RESPONSE			RESPONSE-2		
		Cost	QALYs	ICER	Cost	QALYs	ICER
EAG base case	BAT	£100,281	7.02		£93,866	7.77	
	Ruxolitinib	■	■	■	■	■	■
Ruxolitinib TTD odds spline 1	BAT	£101,830	7.13		£92,133	7.62	
	Ruxolitinib	■	■	■	■	■	■
HR OS: MAJIC-PV time varying	BAT	£90,278	6.28		£86,499	7.13	
	Ruxolitinib	■	■	■	■	■	■
BAT partition	BAT	£97,714	6.88		£91,454	7.61	
	Ruxolitinib	■	■	■	■	■	■
MF-8D utilities	BAT	£100,281	7.49		£93,866	8.30	
	Ruxolitinib	■	■	■	■	■	■
Company Grade 1-2 TE costs	BAT	£100,262	7.02		£93,824	7.77	
	Ruxolitinib	■	■	■	■	■	■
Waning from year 5 to 10	BAT	£103,118	7.22		£96,080	7.96	
	Ruxolitinib	■	■	■	■	■	■
Waning from year 5 to 30	BAT	£98,782	6.91		£92,542	7.66	
	Ruxolitinib	■	■	■	■	■	■
Waning from year 5 to 50	BAT	£97,525	6.82		£91,424	7.56	
	Ruxolitinib	■	■	■	■	■	■
Time horizon 30 years	BAT	£99,178	6.96		£93,194	7.73	
	Ruxolitinib	■	■	■	■	■	■
Remove impact of key events	BAT	£62,184	7.09		£68,639	7.87	
	Ruxolitinib	■	■	■	■	■	■

Source: EAG analysis using the company's model.

QALYs: quality-adjusted life years; ICER: incremental cost-effectiveness ratio; BAT: best available therapy; TTD: time to treatment discontinuation; HR: hazard ratio; OS: overall survival; TE: thromboembolic.

Table 34 Scenario analyses on the EAG base case model: MAJIC-PV population analysis

Scenario	Treatment	Cost	QALYs	ICER
EAG base case	BAT	£83,339	5.71	
	Ruxolitinib	■	■	■
HR OS: MAJIC-PV time-varying	BAT	£83,339	5.71	
	Ruxolitinib	■	■	■
MF-8D utilities	BAT	£83,339	6.11	
	Ruxolitinib	■	■	■

111

Company Grade 1-2 TE costs	BAT	£83,317	5.71	
	Ruxolitinib	■■■	■	■■
Waning from year 5 to 10	BAT	£83,339	5.71	
	Ruxolitinib	■■	■	■■
Waning from year 5 to 30	BAT	£83,339	5.71	
	Ruxolitinib	■■	■	■■
Waning from year 5 to 50	BAT	£83,339	5.71	
	Ruxolitinib	■■	■	■■
BAT OS: lognormal	BAT	£101,122	6.96	
	Ruxolitinib	■■	■	■■
BAT OS: loglogistic	BAT	£94,968	6.52	
	Ruxolitinib	■■	■	■■
BAT OS: hazard spline 1	BAT	£98,374	6.77	
	Ruxolitinib	■■	■	■■
BAT OS: Gompertz	BAT	£70,494	4.80	
	Ruxolitinib	■■	■	■■
Time horizon: 30 years	BAT	£83,271	5.71	
	Ruxolitinib	■■	■	■■
Remove impact of key events	BAT	£57,187	5.78	
	Ruxolitinib	■■	■	■■
Source: EAG analysis using the company's model. QALYs: quality-adjusted life years; ICER: incremental cost-effectiveness ratio; BAT: best available therapy; HR: hazard ratio; OS: overall survival; TE: thromboembolic.				

6.3 Conclusions on the cost effectiveness evidence

The company's model generated base case ICERs of ■■■, ■■■, and ■■■ per QALY for the licensed populations with and without splenomegaly and the MAJIC-PV population analysis, respectively. In response to clarification question B4, the company performed scenario analyses adjusting pre-discontinuation survival for general population mortality for the entire time horizon for the primary analysis. These scenarios produced ICERs of ■■■ and ■■■ for the licensed population with and without splenomegaly, respectively. The EAG considers this scenario as a correction (see section 5.3.3 above).

The EAG preferred model assumptions are the following:
- Correction to include the general population mortality constraint for pre-discontinuation survival throughout the time horizon (primary analysis)
- Weibull distribution for extrapolation of TTD for ruxolitinib due to reasons other than death, as we consider that this provides a better fit to the data than the odds spline model with one know that the company used (primary analysis)
- Treatment effect estimated using the constant HR estimate for OS, as reported by the MAJIC-PV trial investigators
- No partitioning of the BAT health state (primary analysis)

- EQ-5D utility values
- EAG estimated cost assumed for the management of Grade 1-2 thromboembolic events.

The EAG's correction and preferred assumptions increase the ICER to ▇▇▇ per QALY for the licensed population with splenomegaly, ▇▇▇ per QALY for the licensed population without splenomegaly, and ▇▇▇ per QALY for the MAJIC-PV population analysis. These estimates are most sensitive to the assumptions regarding the source of treatment effect for overall survival and the source of utility values.

Alternative assumptions about the waning of the treatment effect also affect the ICER, and we note that EAG clinical advisors have suggested that they do not have reason to expect that the effectiveness of ruxolitinib would wane over time.

We also report a scenario removing the QALY loss and costs for major complications of PV to illustrate the impact of the way in which this has been modelled, not because we believe that it might be appropriate to exclude these impacts.

7 SEVERITY MODIFIERS

The company state that the QALY shortfall criteria for severity weighting, as defined in the 2022 NICE health technology evaluations manual,[42] are not met (CS B.3.6 and Table 36). We show the absolute and proportional QALY shortfalls for the populations based on the company's base case analyses and EAG preferred assumptions in Table 35 below. The criteria for severity weighting are not met under the EAG's preferred assumptions.

Table 35 QALY shortfall analysis

Model (population)	Expected total QALYs [a]		QALY shortfall	
	General population [b]	Model	Absolute	Proportional
Company base case				
STM (RESPONSE population)	12.60	6.97	5.63	0.45
STM (RESPONSE-2 population)	11.13	7.80	3.32	0.30
PSM (MAJIC-PV population)	10.55	6.11	4.45	0.42
EAG preferred assumptions				
STM (RESPONSE population)	12.60	7.02	5.59	0.44
STM (RESPONSE-2 population)	11.13	7.77	3.36	0.30

PSM (MAJIC-PV population)	10.55	5.71	4.84	0.46

STM: state-transition model; PSM: partitioned survival model

ᵃ Discounted QALYs over the model time horizon (46 years from starting age)

ᵇ General population utilities by age and sex from Hernández Alava et al. 2022[51]

Source: Adapted from CS Table 36, with results for the EAG preferred analysis calculated from the company's model

8 REFERENCES

1. Stein BL, Moliterno AR, Tiu RV. Polycythemia vera disease burden: contributing factors, impact on quality of life, and emerging treatment options. *Annals of Hematology* 2014;93(12):1965-76. doi: 10.1007/s00277-014-2205-y [published Online First: 20141002]

2. MPN Voice. Polycythaemia vera 2020 [Available from: https://www.mpnvoice.org.uk/about-mpns/questions/polycythaemia-vera/ accessed 24/10/2022.

3. MPN Voice. About MPNs 2021 [Available from: https://www.mpnvoice.org.uk/about-mpns/ accessed 24/10/2022.

4. McMullin MF, Harrison CN, Ali S, et al. A guideline for the diagnosis and management of polycythaemia vera. A British Society for Haematology Guideline. *British Journal of Haematology* 2019;184(2):176-91.

5. The Haematological Malignancy Research Network (HMRN). Factsheets: Polycythaemia vera 2022 [Available from: www.hmrn.org/factsheets#polycythaemia_vera accessed 24/10/2022.

6. Harrison C, Sekhar M, Mitchell C, et al. Pan-London Haemato-Oncology Clinical Guidelines. Acute leukaemias and myeloid neoplasms. Part 4: Myeloproliferative neoplasms. London: RM Partners, South East London Cancer Alliance, North Central and East London Cancer Alliance, 2020.

7. Vannucchi AM, Barbui T, Cervantes F, et al. Philadelphia chromosome-negative chronic myeloproliferative neoplasms: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Annals of Oncology* 2015;26:v85-v99. doi: https://doi.org/10.1093/annonc/mdv203

8. BNF. British National Formulary: Joint Formulary Committee; 2022 [Available from: http://www.medicinescomplete.com accessed 05 August 2022.

9. Marchetti M, Vannucchi AM, Griesshammer M, et al. Appropriate management of polycythaemia vera with cytoreductive drug therapy: European LeukemiaNet 2021 recommendations. *The Lancet Haematology* 2022;9(4):e301-e11.

10. Barosi G, Birgegard G, Finazzi G, et al. A unified definition of clinical resistance and intolerance to hydroxycarbamide in polycythaemia vera and primary myelofibrosis: results of a European LeukemiaNet (ELN) consensus process. *British Journal of Haematology* 2010;148(6):961-63. doi: 10.1111/j.1365-2141.2009.08019.x [published Online First: 2009/11/26]

11. Medicines and Healthcare products Regulatory Agency (MHRA). Jakavi® 10 mg tablets: Summary of Product Characteristics. Available at: https://mhraproducts4853.blob.core.windows.net/docs/d61f7a73d3447fbfcf6d3936be d37a80946f480f [Last Accessed 08 September 2022] 2022 [

12. Medicines and Healthcare products Regulatory Agency (MHRA). Jakavi® 5 mg tablets: Summary of Product Characteristics. 2022 [Available from: https://mhraproducts4853.blob.core.windows.net/docs/9b08c7b88ef11cc23d00d10cd 3f73959b27ef099 accessed 31/10/2022.

13. Medicines and Healthcare products Regulatory Agency (MHRA). Jakavi® 15 mg tablets: Summary of Product Characteristics. 2022 [Available from: https://mhraproducts4853.blob.core.windows.net/docs/8fab59846945bb488753c25883e9e9131e288eac accessed 31/10/2022.

14. Medicines and Healthcare products Regulatory Agency (MHRA). Jakavi® 20 mg tablets: Summary of Product Characteristics. 2022 [Available from: https://mhraproducts4853.blob.core.windows.net/docs/6667a689dcc1eaf7cd7e8a5d11c7b1029da2c80a accessed 31/10/2022.

15. National Institute for Health and Care Excellence. TA386: Ruxolitinib for disease-related splenomegaly or symptoms in adults with myelofibrosis (review of TA289). 2016 [Available from: https://www.nice.org.uk/guidance/TA386 accessed 05 August 2022.

16. Harrison C, Nangalia J, Boucher R, et al. Ruxolitinib versus Best Available Therapy for PV intolerant or resistant to Hydroxycarbamide in a Randomized trial (Manuscript in Development). *Confidential unpublished manuscript* 2022

17. Pardanani A, Harrison C, Vannucchi A, et al. MPN-184: Safety and Efficacy of Fedratinib, an Oral, Selective JAK2 Inhibitor, in Patients with Polycythemia Vera (PV) Resistant or Intolerant to Hydroxyurea: Results from a Phase II Dose-Ranging and Dose-Expansion Study. *Clinical Lymphoma, Myeloma and Leukemia* 2020;20:S332-S33.

18. NCT00928707. Phase II Study of GIVINOSTAT (ITF2357) in Combination With Hydroxyurea in Polycythemia Vera (PV): ClinicalTrials.gov; [Available from: Https://clinicaltrials.gov/ct2/show/NCT00928707 accessed 08/09/2022.

19. NCT00726232. Study to Determine the Safety and Efficacy of INCB018424 in Patients With Polycythemia Vera or Essential Thrombocythemia.: ClinicalTrials.gov; [Available from: https://clinicaltrials.gov/ct2/show/NCT00726232 accessed 08/09/2022.

20. NCT01632904. Randomized Switch Study From Hydroxyurea to Ruxolitinib for RELIEF of Polycythemia Vera Symptoms: the Relief Study (NCT01632904). ClinicalTrials.gov; [Available from: https://clinicaltrials.gov/ct2/show/NCT01632904 accessed 09/09/2022.

21. Mesa R, Vannucchi AM, Yacoub A, et al. The efficacy and safety of continued hydroxycarbamide therapy versus switching to ruxolitinib in patients with polycythaemia vera: a randomized, double-blind, double-dummy, symptom study (RELIEF). *British Journal of Haematology* 2017;176(1):76-85.

22. Vannucchi AM, Kiladjian JJ, Griesshammer M, et al. Ruxolitinib versus standard therapy for the treatment of polycythemia vera. *New England Journal of Medicine* 2015;372(5):426-35.

23. Passamonti F, Griesshammer M, Palandri F, et al. Ruxolitinib for the treatment of inadequately controlled polycythaemia vera without splenomegaly (RESPONSE-2): a randomised, open-label, phase 3b study. *The Lancet Oncology* 2017;18(1):88-99.

24. Kiladjian J-J, Zachee P, Hino M, et al. Long-term efficacy and safety of ruxolitinib versus best available therapy in polycythaemia vera (RESPONSE): 5-year follow up of a phase 3 study. *The Lancet Haematology* 2020;7(3):e226-e37. doi: https://doi.org/10.1016/S2352-3026(19)30207-8

25. Barosi G, Birgegard G, Finazzi G, et al. Response criteria for essential thrombocythemia and polycythemia vera: result of a European LeukemiaNet consensus conference. *Blood* 2009;113(20):4829-33. doi: 10.1182/blood-2008-09-176818

26. Barosi G, Tefferi A, Besses C, et al. Clinical end points for drug treatment trials in BCR-ABL1-negative classic myeloproliferative neoplasms: consensus statements from European LeukemiaNET (ELN) and Internation Working Group-Myeloproliferative Neoplasms Research and Treatment (IWG-MRT). *Leukemia* 2015;29(1):20-26. doi: 10.1038/leu.2014.250

27. Scherber R, Dueck AC, Johansson P, et al. The Myeloproliferative Neoplasm Symptom Assessment Form (MPN-SAF): international prospective validation and reliability trial

in 402 patients. *Blood* 2011;118(2):401-08. doi: 10.1182/blood-2011-01-328955 [published Online First: 20110502]

28. Emanuel RM, Dueck AC, Geyer HL, et al. Myeloproliferative neoplasm (MPN) symptom assessment form total symptom score: prospective international assessment of an abbreviated symptom burden scoring system among patients with MPNs. *Journal of Clinical Oncology* 2012;30(33):4098.

29. Novartis. Data on File. Protocol CINC424B2301: Randomized, open label, multicenter phase III study of efficacy and safety in polycythemia vera subjects who are resistant to or intolerant of hydroxyurea: JAK inhibitor INC424 tablets versus best available care (The RESPONSE Trial).

30. Novartis. Data on File. Protocol CINC424B2401: Randomized, open label, multicenter phase IIIb study evaluating the efficacy and safety of ruxolitinib versus best available therapy in patients with Polycythemia Vera who are hydroxyurea resistant or intolerant (RESPONSE-2).

31. Harrison C. MAJIC Clinical Study Protocol / Chief Investigator: Claire Harrison. Version 9.0 ed: University of Birmingham, 2018.

32. Novartis. Data on File. Final Clinical Study Report (Week 260): Randomized, open-label, multi-center phase IIIb study evaluating the efficacy and safety of ruxolitinib versus best available therapy in patients with polycythemia vera who are hydroxyurea resistant or intolerant (RESPONSE-2).

33. Phillippo D, Ades A, Dias S, et al. NICE DSU Technical Support Document 18: Methods for population-adjusted indirect comparisons in submissions to NICE. Sheffield: Decision Support Unit, 2016.

34. Alvarez-Larrán A, Pereira A, Cervantes F, et al. Assessment and prognostic value of the European LeukemiaNet criteria for clinicohematologic response, resistance, and intolerance to hydroxyurea in polycythemia vera. *Blood* 2012;119(6):1363-9. doi: 10.1182/blood-2011-10-387787 [published Online First: 20111212]

35. Alvarez-Larran A, Verstovsek S, Perez-Encinas M, et al. PF628. Comparison of ruxolitinib and real-world best available therapy in terms of overall survival and thrombosis in patients with polycythemia vera who are resistant or intolerant to hydroxyurea. European Hematology Association Congress. Stockholm, Sweden, 2018.

36. Novartis. Data on File: Propensity Score Matching Analysis: Overall Survival at Week 256. Ruxolitinib (RESPONSE) versus BAT (GEMFIN).

37. Alvarez-Larrán A, Garrote M, Ferrer-Marín F, et al. Real-world analysis of main clinical outcomes in patients with polycythemia vera treated with ruxolitinib or best available therapy after developing resistance/intolerance to hydroxyurea. *Cancer* 2022;128(13):2441-48. doi: https://doi.org/10.1002/cncr.34195

38. Hong S, Veenstra DL. PCN153: Cost-Utility Analysis of Ruxolitinib versus Best Available Therapy for the Treatment of Hydroxyurea Resistant/Intolerant Polycythaemia Vera Without Splenomegaly in the United States. *Value in Health* 2020;23:S50. doi: 10.1016/j.jval.2020.04.1638

39. National Centre for P. Cost-effectiveness of ruxolitinib (Jakavi®) for polycythaemia vera. Available at: http://www.ncpe.ie/wp-content/uploads/2015/05/Web-summary-final.pdf. [Last Accessed 08 September 2022].

40. Scottish Medicines Consortium. Ruxolitinib (Jakavi) 2019 [Available from: https://www.scottishmedicines.org.uk/medicines-advice/ruxolitinib-jakavi-full-smc2213/ accessed 08/09/2022.

41. Scottish Medicines Consortium. Ropeginterferon alfa-2b 250 micrograms/0.5 mL solution for injection in pre-filled pen (Besremi®). 2022 [Available from: https://www.scottishmedicines.org.uk/medicines-advice/ropeginterferon-alfa-2b-besremi-full-smc2421/ accessed 08/09/2022.

42. National Institute for Health and Care Excellence. NICE health technology evaluations: the manual 2022 [Available from:

https://www.nice.org.uk/process/pmg36/chapter/introduction-to-health-technology-evaluation accessed 05/08/2022.

43. Mukuria C, Rowen D, Brazier JE, et al. Deriving a Preference-Based Measure for Myelofibrosis from the EORTC QLQ-C30 and the MF-SAF. *Value in Health* 2015;18(6):846-55. doi: https://doi.org/10.1016/j.jval.2015.07.004

44. National Institute for Health and Care Excellence. TA756: Fedratinib for treating disease-related splenomegaly or symptoms in myelofibrosis. Available at: https://www.nice.org.uk/guidance/ta756 [Last Accessed 05 August 2022] 2021 [Available from: https://www.nice.org.uk/guidance/ta756.

45. Woods B, Sideris E, Palmer S, et al. NICE DSU Technical Support Document 19: Partitioned Survival Analysis for Decision Modelling in. 2017

46. Woods BS, Sideris E, Palmer S, et al. Partitioned Survival and State Transition Models for Healthcare Decision Making in Oncology: Where Are We Now? *Value in Health* 2020;23(12):1613-21. doi: 10.1016/j.jval.2020.08.2094

47. Latimer N. NICE DSU Technical Support Document 14: Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data, 2011.

48. Lelonek E, Matusiak L, Wrobel T, et al. Burden of aquagenic pruritus in polycythaemia vera. *Acta dermato-venereologica* 2018;98(2):185-90. doi: 10.2340/00015555-2812 [published Online First: 2017/10/04]

49. Rafia R, Mukuria CW. PCN127 - Exploratory Psychometric Analysis of The Eq-5d In A Myelofibrosis Population. *Value in Health* 2015;18(3):A211. doi: https://doi.org/10.1016/j.jval.2015.03.1224

50. Hernández Alava M, Pudney S, Wailoo A. Estimating the relationship between EQ-5D-5L and EQ-5D-3L: results from an English Population Study: Health Economics and Decision Science, School of Health and Related Research, University of Sheffield, UK, 2020.

51. Wailoo MHA, Pudney S. Estimating EQ-5D by age and sex for the UK: The University of Sheffield, 2022.

52. Matza LS, Cong Z, Chung K, et al. Utilities associated with subcutaneous injections and intravenous infusions for treatment of patients with bone metastases. *Patient preference and adherence* 2013;7:855-65.

53. National Institute for Health and Care Excellence. TA722: Pemigatinib for treating relapsed or refractory advanced cholangiocarcinoma with FGFR2 fusion or rearrangement. Available at: https://www.nice.org.uk/guidance/ta722 [Last Accessed 05 August 2022] 2021 [Available from: https://www.nice.org.uk/guidance/ta722.

54. Wang H-I, Aas E, Howell D, et al. Long-Term Medical Costs and Life Expectancy of Acute Myeloid Leukemia: A Probabilistic Decision Model. *Value in Health* 2014;17(2):205-14. doi: https://doi.org/10.1016/j.jval.2013.12.007

55. Crathorne L, Reimer E, Harris S, et al. PSY79 - The Cost Of Major Bleeding Events Associated With Anticoagulation Treatment - An Estimation Of The Contribution Of The Reversal Agents To The Overall Cost Of Major Bleeding In A UK Setting. *Value in Health* 2018;21:S449. doi: 10.1016/j.jval.2018.09.2654

56. Vallejo-Torres L, Morris S, Kinge JM, et al. Measuring current and future cost of skin cancer in England. *Journal of Public Health (Oxford)* 2014;36(1):140-8. doi: 10.1093/pubmed/fdt032 [published Online First: 2013/04/05]

57. National Institute for Health and Care Excellence. Venous thromboembolic diseases: diagnosis, management and thrombophilia testing, NICE guideline NG158:

[A] Evidence reviews for D-dimer testing in the diagnosis of deep vein thrombosis and pulmonary embolism: National Institute for Health and Care Excellence, 2020.

58. Novartis. Data on File. Novartis UK Polycythemia Vera Advisory Board. 24th June 2022.

59. Novartis. Data on File. Novartis UK Polycythemia Vera Advisory Board. 08th August 2022.

60. Department of Health. National Schedule of Reference Costs 2020-2021. Available at: https://www.england.nhs.uk/costing-in-the-nhs/national-cost-collection/#ncc1819 [Last Accessed 05 August 2022] [

61. Passamonti F, Palandri F, Saydam G, et al. Ruxolitinib versus best available therapy in inadequately controlled polycythaemia vera without splenomegaly (RESPONSE-2): 5-year follow up of a randomised, phase 3b study. *The Lancet Haematology* 2022;9(7):E480-E92. doi: 10.1016/S2352-3026(22)00102-8

9 APPENDICES

9.1 EAG critique of the methods of review

Systematic review components and processes	EAG response (Yes, No, Unclear)	EAG comments
Was an appropriate review question clearly defined using the PICOD framework or an alternative?	Partly	The review question was clearly defined as identifying RCTs on the clinical efficacy and safety of any treatment in PV patients who are resistant or intolerant of HC (CS Appendix D.1), supported by a PICOS table for eligibility criteria (CS Appendix Table 8). However, limiting the study design to RCTs, and not searching for observational studies, meant the SLR could not identify relevant studies to support the ITC.
Were appropriate sources of literature searched?	Yes	The core bibliographic medical databases MEDLINE (including MEDLINE In-Process, etc.), Embase, and the Cochrane Library for CDSR and CENTRAL were searched. Several relevant haematology and oncology conferences, ClinicalTrials.gov, and the bibliographies of relevant systematic reviews and meta-analyses were searched (CS Appendix D.1.1).
Did the searches span an appropriate time period?	Yes	The original and update searches covered from database inception to 8 June 2022 (CS Appendix D.1.1).
Were appropriate search terms used and combined correctly?	Yes	Disease terms for PV were combined with RCT terms that were closely based on a published and validated search filter. Both subject headings and free text terms were used. All search strings were reported (CS Appendix D.1.1).
Were inclusion and exclusion criteria specified? If so, were these criteria appropriate and relevant to the decision problem?	Yes, except criteria for the intervention/ comparators were broader than the decision problem	The eligibility criteria for the SLR are defined in CS Appendix Table 8. They are appropriate and relevant but broader than the decision problem because they include any pharmacological intervention for the treatment of PV. This explains why 4 out of the 8 studies identified in the SLR were excluded (discussed above in section 3.2).
Were study selection criteria applied by two or more reviewers independently?	Yes	Two independent reviewers applied the study eligibility criteria. Consensus was achieved by comparison and discussion, and a third independent reviewer made a final decision if necessary (CS Appendix D.1.2).
Was data extraction performed by two or more reviewers independently?	No, but the process is adequate	A single individual extracted information with a second individual verifying and checking for missed data. A third individual arbitrated a

		final decision if necessary (CS Appendix D.1.2).
Was a risk of bias assessment or a quality assessment of the included studies undertaken? If so, which tool was used?	Yes, except for the GEMFIN registry cohort	All RCTs identified in the SLR were quality assessed using the CRD checklist (CS Appendix D.1.3 and D.3). However, the GEMFIN registry cohort used in the ITC was not assessed.
Was risk of bias assessment (or other study quality assessment) conducted by two or more reviewers independently?	No, but the process is adequate	A single individual assessed risk of bias and a second individual confirmed the conclusions. A third individual arbitrated a final decision if necessary (CS Appendix D.1.2).
Is sufficient detail on the individual studies presented?	Yes	Study details of all the included studies are tabulated in CS Appendix D.1.3. Some missing documents were provided in response to clarification questions A2 to A6. The CSR for RESPONSE week 32 was not provided.
If statistical evidence synthesis (e.g. pairwise meta-analysis, ITC, NMA) was undertaken, were appropriate methods used?	Yes	The company conducted an ITC (CS section B.2.9) using appropriate propensity score matching methods in order to estimate OS that was not confounded by crossover. Discussed in sections 3.3 to 3.5 of this report.
CDSR: Cochrane Database of Systematic Reviews; CENTRAL: Cochrane Central Register of Controlled Trials; CRD: Centre for Reviews and Dissemination; CSR: clinical study report; HC: hydroxycarbamide; OS: overall survival; PICOS: population, intervention, comparator, outcome, study design; PV: polycythaemia vera; RCTs: randomised controlled trials; SLR: systematic literature review.		

9.2 Baseline characteristics of the included studies

| Characteristic | RESPONSE | | RESPONSE-2 | | MAJIC-PV | | GEMFIN |
	Ruxolitinib (n=110)	BAT (n=112)	Ruxolitinib (n=74)	BAT (n=75)	Ruxolitinib (█)	BAT (█)	BAT (n=█)[e]
Age – years							
Median (range)	62.0 (34–90)	60.0 (33–84)	63 (NR)	67 (NR)	█	█	Mean ± SD █
IQR	-	-	54–71	61–74	-	-	-
>60 years – n (%)	-	-	46 (62)	57 (76)	-	-	-
Sex – n (%)							
Male	66 (60.0)	80 (71.4)	39 (53)	47 (63)	█	█	█
Female	44 (40.0)	32 (28.6)	35 (47)	28 (37)	█	█	-
Time since diagnosis – years							
Median (range)	8.2 (0.5–36)	9.3 (0.5–23)	6.5 (2.9–10.7)	6.7 (3.2–10.6)	-	-	-
Disease duration - months							
Median (range)	-	-	-	-	█	█	-
Previous lines of therapy							
Median (range)	-	-	-	-	█	█	-
Previous lines of antineoplastic therapy							
1	-	-	53 (72%)	52 (69%)	-	-	-
>1	-	-	21 (28%)	23 (31%)	-	-	-
Duration of prior HC/HU therapy – years							
Median (range)	3.1 (<0.1–20.9)	2.8 (<0.1–20.9)	2.83 (0.57–6.61) [a]	3.55 (0.57–7.03) [a]	-	-	
Resistance/intolerance (R/I) to hydroxycarbamide							
Both R/I – n (%)	-	-	-	-	█	█	-
Intolerant – n (%)	-	-	-	-	█	█	-
Resistant – n (%)	-	-	-	-	█	█	-

Characteristic	RESPONSE		RESPONSE-2		MAJIC-PV		GEMFIN
	Ruxolitinib (n=110)	BAT (n=112)	Ruxolitinib (n=74)	BAT (n=75)	Ruxolitinib (█)	BAT (█)	BAT (n=█)[e]
Previous HC/HU treatment status – n (%)							
Unacceptable side effects	59 (53.6)	61 (54.5)	44 (59)	45 (60)	-	-	-
Inadequate response	51 (46.4)	51 (45.5)	30 (41)	30 (40)	-	-	-
ECOG performance status – n (%) [b]							
0	76 (69.1)	77 (68.8)	-	-	█	█	-
1	31 (28.2)	34 (30.4)	-	-	█	█	-
2	3 (2.7)	1 (0.9)	-	-	█	█	-
Prior thromboembolic event							
n (%)	39 (35.5)	33 (29.5)	21 (28)	18 (24)	█	█	█ [f]
Presence of *JAK2* V617F mutation							
n (%)	104 (94.5)	107 (95.5)	72 (97) [c]	69 (92)	-	-	█
Allele burden – % ± SD	76.2 ± 17.8	75.0 ± 22.6	-	-	-	-	
JAK2 mutation status							
Wild type – n (%)	-	-	-	-	█	█	-
JAK2V617F – n (%)	-	-	-	-	█	█	-
JAK2 exon 12 – n (%)	-	-	-	-	█	█	-
Spleen length							
Below costal margin – cm							
Median (range)	7.0 (0–24.0)	7.0 (0–25.0)	-	-	-	-	-
<10 cm – n (%)	71 (64.5)	67 (59.8)	-	-	-	-	-
>20 cm – n (%)	2 (1.8)	4 (3.6)	-	-	-	-	-
Overall length by ultrasound – cm							
Median (range) [g]	-	-	-	-	█	█	-
Spleen volume – cm³							

122

Characteristic	RESPONSE		RESPONSE-2		MAJIC-PV		GEMFIN
	Ruxolitinib (n=110)	BAT (n=112)	Ruxolitinib (n=74)	BAT (n=75)	Ruxolitinib (▉)	BAT (▉)	BAT (n=▉)[e]
Median (range)	1195 (396–4631)	1322 (254–5147)	-	-	-	-	
Palpable splenomegaly							
n (%)	-	-	-	-	▉	▉	-
Percentage HCT level – % [d]							
Mean ± SD	43.6 ± 2.2	43.9 ± 2.2	42.8 ± 1.46	42.7 ± 1.44	-	-	-
Median (range or IQR)	43.3 (range: 39.2–50.5)	44.0 (range: 37.6–50.5)	43.0 (IQR: 41.7–44.0)	42.7 (IQR: 41.7–44.0)	▉	▉	-
HCT category – n (%)							
40–45%	79 (71.8)	83 (74.1)	-	-	-	-	-
>45%	28 (25.5)	25 (22.3)	-	-	-	-	-
WBC count × 10⁻⁹/L							
Mean ± SD	17.6 ± 9.6	19.0 ± 12.2	12.0 ± 8.19	13.0 ± 8.06	-	-	-
Median (range)	-	-	-	-	▉	▉	-
Platelet count × 10⁻⁹/L							
Mean ± SD	484.5 ± 323.3	499.4 ± 318.6	469.5 ± 295.96	471.5 ± 350.38	-	-	-
Median (range)	-	-	-	-	▉	▉	-
Haemoglobin g/L							
Median (range)	-	-	-	-	▉	▉	-
Phlebotomies within 24 weeks before screening							
≥2 – n (%)	-	-	58 (78)	57 (76)	-	-	-
Median (range)	2.0 (1–8)	2.0 (0–16)	-	-	-	-	
History of haemorrhage							
n (%)	-	-	-	-	▉	▉	-
Migraine or erythromelalgia							
n (%)	-	-	-	-	▉	▉	-
Diabetes							

123

Characteristic	RESPONSE		RESPONSE-2		MAJIC-PV		GEMFIN
	Ruxolitinib (n=110)	BAT (n=112)	Ruxolitinib (n=74)	BAT (n=75)	Ruxolitinib (■)	BAT (■)	BAT (n=■)[e]
n (%)	-	-	-	-	■	■	-
Hypertension							
n (%)	-	-	-	-	■	■	-
Cytopenia at lowest hydroxycarbamide dose							
n (%)	17 (15)	-	-	-	-	-	■

Sources: CS Table 7; CS Table 12; CS Appendix M.2.1; Clarification response A11 Table 2.

[a] Manually converted duration in months from the source to duration in years for consistency. [b] ECOG performance status ranges from 0 to 5, with 0 indicating no symptoms and higher numbers indicating increasing disability. [c] For five patients (ruxolitinib, n=2; BAT, n=3) the *JAK2* V617F mutation was not confirmed by central laboratory assessment. These patients were not included as *JAK2* V617F mutation positive. [d] Value at the end of the HCT control period before randomisation. Patients who had an HCT of 40–45% within 14 days before their day 1 visit could proceed to randomisation; however, the HCT at baseline may have been higher or lower. [e] Excludes 7 patients without follow-up beyond the date of being identified as resistant or intolerant to hydroxycarbamide (clarification response A10). [f] At time of resistance/intolerance. [g] from clarification response C1.

124

9.3 Company and EAG risk of bias assessments for the RCTs

Question	Assessor	Trial		
		RESPONSE	**RESPONSE-2**	**MAJIC-PV**
Was randomisation carried out appropriately?	Company	Unclear risk of bias, randomisation methods not reported	Low risk of bias, random assignment of participants (1:1), using an interactive voice and web response system.	Unclear risk of bias, randomisation methods were not reported
	EAG	**Probably low risk of bias** The trial protocol states that an IRT system will assign a randomization number to the participant to link them to a treatment arm. However, the trial publication[22] does not confirm that this process was followed in practice.	**Agree, low risk of bias** An interactive voice and web response system was used to assign randomisation numbers to participants to link each participant to a trial arm. [61]	**Agree, unclear risk of bias** The trial protocol states that "randomisation will be based on a minimisation algorithm prepared by the trial statistician", but not reported whether or how this was conducted.
Was the concealment of treatment allocation adequate?	Company	Unclear risk of bias, concealment of treatment was not reported	Low risk of bias, an interactive voice and web response system was contacted by the investigator	Unclear risk of bias, concealment of treatment was not reported
	EAG	**High risk of bias** Due to being an open-label trial (NB the full allocation process is not explained and the trial publication[22] does not confirm that the stated process was followed in practice).	**High risk of bias** Due to being an open-label trial (NB the full allocation process is not explained and the trial publication[61] does not confirm that the stated process was followed in practice).	**High risk of bias** Due to being an open-label trial. Some patients ▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓ (Figure S2 in the draft trial manuscript[16]).
Were the groups similar at the outset of the study in terms of prognostic factors, for example severity of disease?	Company	Low risk of bias, the authors of the primary publication reported that there were no significant differences between the two treatment groups with regard to baseline	Low risk of bias, baseline characteristics were generally similar between treatment groups. There were slight differences in median age and sex between the groups	Low risk of bias, authors reported that baseline characteristics at randomisation were balanced, however full patient characteristics were not reported

		characteristics and disease history		
	EAG	**Agree, low risk of bias** Baseline characteristics appear well balanced with minor exceptions (the ruxolitinib arm had 11% more females and 6% more people who had had a prior thromboembolic event than the BAT arm).	**Agree, low risk of bias** Baseline characteristics appear well balanced with minor exceptions (the ruxolitinib arm had 14% fewer people aged > 60 years and median age 4 years younger, 10% more females and a median 8.7 months less prior hydroxycarbamide therapy than the BAT arm).	**Unclear risk of bias** Most baseline characteristics appear balanced. However, ▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆ had prior thrombosis and the BAT arm also had a ▆▆▆▆▆ disease duration and number of previous lines of therapy; whilst ▆▆▆ patients in the ruxolitinib arm were both intolerant and resistant to hydroxycarbamide.[16]
Were the care providers, participants and outcome assessors blind to treatment allocation? If any of these people were not blind to treatment allocation, what might be the likely impact on the risk of bias (for each outcome)?	**Company**	High risk of bias, open-label study. There was a potential for bias, particularly in PROs. Bias for ruxolitinib versus hydroxycarbamide may be particularly relevant as patients were already known to be hydroxycarbamide -resistant/ intolerant	High risk of bias, open-label study. There was a potential for bias in outcomes, particularly PROs. Bias for ruxolitinib versus HC/HU may be particularly relevant as patients were already known to be hydroxycarbamide-resistant/ intolerant The assessors were unaware of the treatment group assignments until database lock	High risk of bias, open-label study. Potential for bias, particularly in symptom and QoL scores.
	EAG	**Agree, high risk of bias** Note that being open label the trial has high risks of bias relating to: (i) elective patient crossover, (ii) patient care, and (iii) recording of outcomes, (iv) analysis of outcomes.	**Agree, high risk of bias** Note that being open label the trial has high risks of bias relating to: (i) elective patient crossover, (ii) patient care, and (iii) recording of outcomes, (iv) analysis of outcomes.	**Agree, high risk of bias** Note that being open label the trial has high risks of bias relating to: (i) patient care, (ii) recording of outcomes and (iii) analysis of outcomes.

126

Were there any unexpected imbalances in drop-outs between groups? If so, were they explained or adjusted for?	Company	High risk of bias, patients were able to cross over from BAT treatment arm to ruxolitinib at Week 32; 96 patients crossed over at or after Week 32 – this would have been affected by the open label nature of the study	High risk of bias, patients were able to cross over from BAT treatment arm to ruxolitinib at Week 28; 51 patients crossed over at or after Week 28 – this would have been affected by the open label nature of the study.	Unclear risk of bias, drop-outs were not reported
	EAG	**≤ week 32: Unclear risk of bias** Unclear whether patients were informed that they could cross over at week 32 and if so whether this would have affected their outcomes prior to week 32. CONSORT chart (CS Appendix Figure 4) does not identify dropout numbers or reasons prior to week 32. **> week 32: Agree, high risk of bias** Reasons as stated by the company	**≤ week 28: Unclear risk of bias** Unclear whether patients were informed that they could cross over at week 32 and if so whether this would have affected their outcomes prior to week 32. CONSORT chart (CS Appendix Figure 5) does not identify dropout numbers or reasons prior to week 32. **> week 28: Agree, high risk of bias** Reasons as stated by the company	**Probably low risk of bias** Table S4 of the unpublished manuscript[16] suggests numbers and reasons for dropout were broadly similar between trial arms.
Is there any evidence to suggest that the authors measured more outcomes than they reported?	Company	Low risk of bias, the pre-defined outcome measures are all presented in the available records	High risk of bias, some outcomes measured are not reported, however analyses are promised in future publications but still not reported (e.g., changes ECOG status and spleen length)	High risk of bias, ISRCTN record lists outcome measures which are not reported in the available records
	EAG	**Efficacy outcomes: probably low risk of bias** Most of the pre-specified outcomes in the trial protocol	**Efficacy outcomes: Unclear risk of bias** The previously missing pre-specified outcomes (e.g. spleen	**Agree, high risk of bias** EQ-5D, MDASI and partial response rate are specified in the trial protocol, but results are not reported. Results

		have been reported, with some minor exceptions (MPN-PAF results not reported; overall clinico-haematologic response reported at 5 years but not at earlier timepoints). **HRQoL outcomes: high risk of bias** 32-week results for the MPN-SAF-TSS and PGIC are reported in the CS and publications only as % changes which have limited clinical interpretation, with no indication of the original scores, sample size or variance in scores. The week 32 CSR was not provided to the EAG.	length, ECOG performance status and WPAI score) are summarised in the week 260 CSR (NB individual patient ECOG PS scores are tabulated but not analysed). **HRQoL outcomes: high risk of bias** Changes in MPN-SAF TSS and PGIC are reported in the CS, publications and week 28 CSR only as % changes which have limited clinical interpretation, with no indication of the original scores, sample size or variance in scores. The week 28 CSR does report numbers achieving disease resolution, but only for a subgroup who had a baseline score of ≥20.	for the MPN-SAF are reported only as differences between arms, without the original scores for each arm.
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Company	Low risk of bias, ITT analysis was used, with data from all patients who underwent randomisation. Patients with missing assessments that prevented the evaluation of the primary and secondary endpoints were considered non-responders	Low risk of bias, ITT analysis was applied for the primary and key secondary endpoints, including data from all patients randomly assigned to treatment Patients with missing assessments that prevented the study of the primary and secondary endpoints endpoint were considered non-responders	Unclear risk of bias, an mITT analysis was used (those who commenced study treatment and had at least one response assessment) but details of how missing data were accounted for were not given.

128

	EAG	**Primary and key secondary outcomes: Low risk of bias** ITT analysis: Missing response data were considered non-responders and missing phlebotomy ineligibility data were considered phlebotomy eligible (number of missing observations not reported). **HRQoL outcomes: High risk of bias** Missing data excluded; number and reasons for missing data not reported. (sources: CS and trial protocol)	**Primary and key secondary outcomes: Low risk of bias** ITT analysis: Missing response data were considered non-responders and missing data for remission outcomes were considered to represent no remission. **HRQoL outcomes: High risk of bias** Missing data excluded; number and reasons for missing data not reported (sources: CS and trial protocol)	**Primary and secondary outcomes: unclear risk of bias** The trial protocol states that for secondary outcomes "the amount of missing data will be reported but not imputed". However the amount of missing data is not reported. **HRQoL outcomes: High risk of bias** Missing data probably excluded; number and reasons for missing data not reported. Sample size is unclear for MPN-SAF. **All outcomes: unclear risk of bias** Lack of clarity around crossovers from ruxolitinib to BAT and receipt of ruxolitinib on the BAT arm (see section 3.2.3 for discussion).
Also consider whether the authors of the study publication declared any conflicts of interest/study funding.	**Company**	Unclear risk of bias, sponsor (Incyte and Novartis) involvement in study design and data analysis not reported, Author affiliations were disclosed	Low risk of bias, study funding and author conflicts of interest declared. The study was sponsored and designed by Novartis. Data were analysed and interpreted by Novartis in collaboration with all the authors. Novartis was unaware of treatment group assignments until database lock	Unclear risk of bias, nothing declared. Funder: Leukaemia & Lymphoma Research (UK)
	EAG	Conflicts of interest is not an independent domain of bias. Any risks of bias arising through conflicts of interest would be reflected in the bias assessments already reported above. For example, Novartis' (lack		

		of) awareness of treatment assignment should already be captured under the allocation concealment and blinding questions which indicate a high risk of bias.

Source: CS Appendix Table 14 with EAG additions. BAT: best available therapy; IRT: interactive response technology; ITT: intention to treat; MDASI: M.D. Anderson Symptom Inventory; mITT: modified intention to treat; MPN-PAF: Myeloproliferative Neoplasm Pruritis Assessment Form

9.4   EAG summary of statistical methods in the RCTs

	RESPONSE	RESPONSE-2	MAJIC-PV
Analysis populations			
Summary	**Full analysis set: ITT analysis** (primary and two key secondary outcomes): all randomised patients included and analysed according to their hydroxycarbamide stratum and the treatment they were randomised to. **Safety set:** all randomised patients who received at least one dose of their allocated treatment, analysed according to the treatment they actually received. People randomized to the BAT arm who were intended to receive no therapy were included in the safety set. **Per protocol set:** A subset of the full analysis set patients who received at least one dose of study treatment and did not have a major protocol violation.	**Full analysis set: ITT analysis** (primary and key secondary outcome): all randomised patients included and analysed according to their hydroxycarbamide stratum and the treatment they were randomised to. **Safety set:** all randomised patients who received at least one dose of their allocated treatment, analysed according to the treatment they actually received. People randomized to the BAT arm who were intended to receive no therapy were included in the safety set. **Per protocol set:** A subset of the full analysis set patients who received at least one dose of study treatment and did not have a major protocol violation.	Modified ITT analysis: All patients who started treatment within one year of randomisation and had at least one response available. Safety population: Any patient starting treatment. The draft trial manuscript[16] states that ███ ████████████████████████ ████████████████, supporting analyses were performed █████████ ████████████████████████████ ████████████████████ ████████████████████. However, results of these analyses are not reported.
EAG comment	The analysis populations for the primary, two key secondary, and safety outcomes are appropriate. Analysis populations are not specified for the remaining secondary outcomes and HRQoL measures. The per protocol population is not referred to in the CS which is reasonable given that the full analysis set is more robust.	The analysis populations for the primary, key secondary, and safety outcomes are appropriate. Analysis populations are not specified for the remaining secondary outcomes and HRQoL measures. The per protocol population is not referred to in the CS which is reasonable given that the full analysis set is more robust.	Limited details of the analysis populations are reported; analysis populations for HRQoL outcomes are unclear. Potential for bias due to unaccounted for missing data (see Appendix 9.3).

Sample size and power calculations			
Summary	**Primary outcome:** Assuming an HCT control rate of 10% in the BAT arm and 30% in the ruxolitinib arm, a sample size of 200 patients was deemed to be required to detect a significant difference with a two-sided test (0.05 significance level and 94% power) (CS Table 9). **Key secondary outcomes:** **Durable primary response:** According to the trial protocol, assuming 24% and 8% primary outcome responders in the ruxolitinib and BAT arms respectively at week 48, a large sample normal approximation would give 87% statistical power. An observed response rate as low as 17.1% in the ruxolitinib arm would achieve statistical significance relative to an observed response rate of 8% in the BAT arm. **CHR at week 32:** According to the trial protocol, the power for complete haematological remission at 32 weeks would be approximately 99% using a large sample normal approximation, meaning that an observed response rate as low as 40% in the ruxolitinib arm would achieve statistical significance relative to an observed response rate for the BAT arm of 27%.	**Primary outcome:** Sample size was calculated based on the results for the HCT control portion of the compound primary outcome, assuming HCT control rates of 50% in the ruxolitinib group and 20% in the BAT group (corresponding to an OR of 4·0). A total of 116 patients were needed to detect a significant difference between treatment groups with two-sided t-test at alpha=0·05 and 90% power. Planned enrolment was 130 patients (65 in each group) to allow for an estimated 10% attrition rate (CS Table 9 and trial publication[23]). **Key secondary outcome** According to the trial protocol, a total of 116 patients (58 patients in each treatment arm) would provide 90% power to detect a 30% increase in the rate of CHR at Week 28, between a BAT arm rate of 20% and a ruxolitinib arm rate of 50% (corresponding to an OR of 4.0) at a 5% significance level.	The complete response rate for the control group was estimated to be ▮ and a clinically significant improvement would be ▮. Assuming complete response rates in the control and treatment group were ▮ and ▮ respectively, ▮ patients would be required in each arm to detect a clinically significant difference of ▮ with ▮ statistical power at a ▮ level of significance.[16] Apart from the primary outcome, additional hypotheses tests were unpowered, exploratory and not pre-specified [16]

EAG comment	The trial randomised 110 and 112 participants per arm so appears to be adequately powered for the primary outcome and probably also the two key secondary outcomes (the power calculation descriptions for the secondary outcomes do not specify the sample size). Adequacy of the sample size for detecting treatment effects in the remaining secondary outcomes is uncertain.	The trial randomised 74 and 75 participants per arm so appears to be adequately powered for the primary outcome and key secondary outcome. Adequacy of the sample size for detecting treatment effects in the remaining secondary outcomes is uncertain.	The complete response rates used for the power calculation in the ruxolitinib and BAT arms (▇ and ▇) ▇▇▇▇▇▇▇ the observed rates reported in the trial (▇ and ▇). The stated power calculation in the protocol uses a ▇ error rate ▇▇▇▇▇▇▇ to achieve ▇ power ▇▇▇▇▇▇▇. Nevertheless, a treatment effect on the primary outcome ▇▇▇▇▇▇▇). Adequacy of the sample size for detecting treatment effects in the remaining secondary outcomes is uncertain.
Methods to account for multiplicity			
Summary	A family wise α-level of 0.05 overall was applied for three pre-specified comparisons: the primary outcome and two key secondary outcomes. Conditional on significance of the primary outcome, treatment effects on the proportions of people achieving a CHR at week 32 and achieving a durable primary endpoint response at week 48 were tested at two-sided α = 0.05 for the two outcomes, controlling for multiplicity using the Hochberg procedure. [22] According to the trial protocol, no alpha adjustment was planned for the remaining secondary outcomes.	Not reported in the CS, week 28 CSR or study publication.[23] According to the trial protocol, the analysis of the key secondary outcome (proportion achieving CHR at week 28) was performed in a hierarchical manner (calculation method not specified). The key secondary outcome was tested at an α-level of 0.05 only if the primary outcome was significant at an α-level of 0.05. For all secondary efficacy outcomes, statistical tests were intended to be performed for descriptive purposes and not adjusted for multiple comparisons.	The CS, draft trial manuscript[16] and trial protocol do not mention whether any control for multiple outcome testing was applied.
EAG comment	The type I error control procedure is appropriate but only three outcomes are included. The likelihood of type I error in testing the remaining secondary outcomes is uncertain.	The type I error control procedure is appropriate but only two outcomes are included. The likelihood of type I error in testing the remaining secondary outcomes is uncertain.	No information available. The likelihood of nonsignificant treatment effects being declared significant is uncertain. Reliance on the statistical test results alone for inference is therefore inadvisable.

Analysis of outcomes			
Summary	**Primary outcome:** Responder rates were analysed using a Cochran-Mantel-Haenszel (CMH) test stratified by hydroxycarbamide tolerance status (resistant versus intolerant), 2-sided at the 5% significance level. The overall stratum-adjusted odds ratio was used as a measure of association between treatment and response. The adjusted proportion difference and its 95% CI were calculated using CMH weight and Wald-type CI or any other appropriate method (CS Table 9). The following is from the trial protocol (not reported in the CS): **Key secondary outcomes** (durable primary response and complete haematological response): Treatment groups were compared using a CMH test stratified on hydroxycarbamide tolerance as with the primary outcome. **All other secondary outcomes:** Are non-comparative in nature. These (except for durability of primary response and duration of primary response which can be evaluated in both treatment groups) will be evaluated only in the subjects originally randomized to ruxolitinib and will be summarised descriptively.	**Primary outcome:** A two-sided CMH test stratified by hydroxycarbamide tolerance status was conducted at the 5% level of significance. The odds ratio is presented with 95% Wald confidence limits (CS Table 9). The following is from the trial protocol (not reported in the CS or week 28 CSR): **Key secondary outcome** (complete hematological remission at week 28): Analysed using a two-sided stratified CMH test (stratification factors not reported in the CS, protocol or publications[23 61]). **Other secondary outcomes** (HCT control at weeks 52, and 80, complete hematological remission at weeks 52 and 80, and partial remission based on the ELN and IWG-MRT criteria at weeks 28, 52 and 80: A two-sided stratified CMH test at the 5% level of significance. **Other outcomes** (changes from baseline in HCT, summary of spleen length, number of phlebotomies from baseline to week 28, and HRQoL measures): Summarised with descriptive statistics.	**Primary outcome:** The trial protocol states that complete response was to be assessed using a normal test with continuity correction and unpooled variance and a ▇▇▇▇ considered statistically significant. Apart from the primary outcome, additional hypotheses tests were exploratory, unpowered, two-sided and considered ▇▇▇▇ statistically significant trial manuscript[16] and protocol). **HRQoL outcomes:** Changes from baseline and between-arm differences in change by timepoint were estimated using a linear mixed model which included covariates for categorical time point, treatment arm, and the interaction between time point and treatment arm. The difference between arms in proportion of patients with best post-baseline TSS response of 50% or greater was tested using a Chi-square test. [16] **Time-to-event outcomes:** Were predominantly analyzed using Kaplan-Meier methods, with differences in survival analyses determined using the Cox model, adjusting for the stratification factor (gender), and treatment (when not the primary variable of interest). [16]

EAG comment	The statistical methods are reported in different sources with varying levels of detail but appear generally appropriate. The CS does not state whether the analyses were checked or validated.	The statistical methods are reported in different sources with varying levels of detail but appear generally appropriate. The CS does not state whether the analyses were checked or validated.	The statistical methods appear generally appropriate. However, no justification is provided for using a ▮▮▮▮▮▮▮▮ p-threshold for determining statistical significance of the primary outcome (but not secondary outcomes) which gives a ▮▮▮▮ chance of nonsignificant effects being declared significant.
Handling of missing data			
Summary	**Primary and key secondary outcomes:** ITT analysis: Missing response data including patient withdrawals were considered non-responders and missing phlebotomy ineligibility data were considered phlebotomy eligible (number of missing observations not reported). **HRQoL outcomes:** Missing data excluded; number and reasons for missing data not reported. **Survival outcomes:** Censoring methods not reported (not specified in the CS, trial protocol or trial publication; the week 32 CSR was not provided to the EAG).	**Primary and key secondary outcomes:** ITT analysis: Missing response data including withdrawals were considered non-responders and missing data for remission outcomes were considered to represent no remission. **HRQoL outcomes:** Missing data excluded; number and reasons for missing data not reported **Survival outcomes:** (not stated in the CS; information from the trial protocol): For TFS, patients without an event by the analysis data cut-off were to be censored at the date of last adequate assessment. For OS, patients not known to have died before the data cut-off were to be censored at the date of the last assessment for patients who were on treatment or at the date of the last contact for patients in survival follow-up.	**Primary and secondary outcomes:** The trial protocol states that for secondary outcomes "the amount of missing data will be reported but not imputed". However the amount of missing data is not reported in the CS or trial draft manuscript. [16] **HRQoL outcomes:** Missing data probably excluded; number and reasons for missing data not reported. Sample size is unclear for MPN-SAF. **Survival outcomes:** Censoring methods not reported.
EAG comment	Methods for handling missing data were appropriate for primary and secondary outcomes. Missing data were not	Methods for handling missing data were appropriate for primary and secondary outcomes. Missing data were not	Overall missing data were not accounted for, and the amount of missing data and

	accounted for in analyses of HRQoL and other exploratory outcomes. Number and reasons for missing data not fully reported.	accounted for in analyses of HRQoL and other exploratory outcomes. Number and reasons for missing data not fully reported.	reasons for data being missing were not reported.
Subgroup analyses			
Summary	Pre-specified subgroup comparisons (trial protocol section 9.4.4) were: baseline palpable splenomegaly (<10cm versus ≥10cm below the costal margin), sex (male versus female), age group (≤60 years versus >60 years), hydroxycarbamide intolerance or resistance, region (US versus non-US), race (White or Caucasian versus other) and ethnicity (Hispanic or Latino versus other). The odds of achieving the primary composite response outcome at week 32 were compared across subgroups by calculating odds ratios and their confidence intervals using logistic regression and displaying these in a forest plot. Post-hoc subgroup comparisons (not specified in the trial protocol) are reported in CS Appendix E for patients who had received prior IFN-alfa, IFN-alfa as BAT, or ruxolitinib after crossover from receiving IFN as BAT. These subgroups pooled data from RESPONSE and RESPONSE-2.	Pre-specified subgroup comparisons (trial protocol section 10.4.4) were: hydroxycarbamide intolerance or resistance, sex (male versus female), age group (≤60 years versus >60 years), risk category (0 risk factors versus 1-2 risk factors including age >60 and/or previous thromboembolism). The odds of achieving HCT control at week 28 were compared across subgroups by calculating odds ratios and their confidence intervals using logistic regression and displaying these in a forest plot. Post-hoc subgroup comparisons (not specified in the trial protocol) are reported in CS Appendix E for patients who had received prior IFN-alfa, IFN-alfa as BAT, or ruxolitinib after crossover from receiving IFN as BAT. These subgroups pooled data from RESPONSE and RESPONSE-2.	Pre-specified subgroup comparisons (trial protocol section 13.3) were: hydroxycarbamide intolerance or resistance, blood count quartile at randomisation (3 classes), sex (male versus female), disease duration (5 classes), ruxolitinib starting dose (5mg or 10mg), number of prior treatments (4 classes), WBC count at trial entry (3 classes), haemoglobin at trial entry (4 classes), and splenomegaly (yes versus no). No analysis methods for subgroups were specified. The trial protocol states that due to the lack of statistical power for subgroup analyses, subgroup analysis results provided will be exploratory only. However, no subgroup analyses are reported in the CS or draft trial manuscript. [16]

136

EAG comment	The pre-specified subgroup analysis method is appropriate, but no justification is provided for the choice of subgroups analysed, which varied between the trials. The post-hoc IFN-alfa subgroups had small sample sizes ranging from 13 to 30 participants.	The pre-specified subgroup analysis method is appropriate, but no justification is provided for the choice of subgroups analysed, which varied between the trials. The post-hoc IFN-alfa subgroups had small sample sizes ranging from 13 to 30 participants.	No subgroup analysis method or results were reported.

BAT: best available therapy; CHR: complete haematological remission; CI: confidence interval; CMH test: Cochran-Mantel-Haenszel test; CSR: clinical study report; ELN: European LeukemiaNet; HRQoL: health-related quality of life; IFN: interferon; ITT: intention to treat; IWG-MRT: International Working Group - Myeloproliferative Neoplasms Research and Treatment; MPN-SAF: Myeloproliferative Neoplasm Symptom Assessment Form; OS: overall survival; TFS: transformation-free survival; US: United States; WBC: white blood cells.

9.5 EAG summary of key economic issues and additional analyses

Analysis	Company analysis	EAG comment	EAG additional analyses
Population and subgroups			
Primary analysis	Subgroup with splenomegaly (RESPONSE trial population)	All three trial populations represent subgroups of the population of interest	We report EAG analyses and scenarios for all three subgroups.
	Subgroup without splenomegaly (RESPONSE-2 trial population)	The EAG considers that the MAJIC-PV analysis is likely to be more relevant as the trial was wholly UK-based and it included the majority of the licensed population	
MAJIC-PV analysis	'High risk' subgroup (MAJIC-PV trial population)		
Model structure			
Primary analysis	STM with three health states (On ruxolitinib, On BAT, death) Key PV complications modelled as events with one-off costs and QALY losses Partition of the BAT state: BAT 1, BAT 2+ and no further treatment	In theory, the STM has the advantage of modelling dependency between discontinuation of ruxolitinib and OS beyond the trial period. Whereas in the PSM, OS and ruxolitinib discontinuation are extrapolated independently However, neither model structure reflects post-trial dependencies between the onset of major complications and survival The BAT partition is subject to uncertainty over long-term trends in cessation of all therapy and disutilities	We do not include partitioning of the BAT state in the EAG preferred analysis. The BAT partition is included in EAG scenario analysis We also note uncertainty over the OS extrapolations as mortality due to complications is not explicitly modelled.
MAJIC-PV analysis	PSM with the same health states as the primary analysis and key PV complications modelled as events No partition of the BAT state		
OS extrapolations			
Primary analysis Survival pre- and post-discontinuation of ruxolitinib (competing risk	Extrapolations fitted to pooled IPD from RESPONSE and RESPONSE-2 Exponential distribution used in base case for pre- and post-discontinuation survival extrapolations. Scenarios with	The competing risk approach is appropriate for the STM, as is the pooling of trial data, given the low numbers of observed events Methods used to fit the survival extrapolations are appropriate and the	We apply the general population mortality constraint throughout the time horizon (company response to CQ B5)

Analysis	Company analysis	EAG comment	EAG additional analyses
analysis)	other distributions are reported in CS Appendix P General population mortality constraint applied after the trial period for pre-discontinuation survival (but throughout the time horizon for post-discontinuation survival).	exponential is a reasonable choice for the base case It is not plausible that mortality rates should be lower in the first five years of ruxolitinib treatment than for people of the same age in the general population	
Treatment effect HR for OS (ruxolitinib vs. BAT)	HR estimated from piecewise Cox proportional hazards analysis of reconstructed MAJIC-PV KM data Scenarios: constant HR from MAJIC-PV trial report; indirect comparison with GEMFIN; Alvarez-Larrán analysis of Spanish data; and pooled HR from RESPONSE and RESPONSE-2 (not corrected for crossover) Waning assumption: linear decline from year 5 to HR=1 at year 20	MAJIC-PV is the best available source for estimation of the relative effect on survival The company's piecewise HR estimates have some face validity, but they are highly uncertain, with wide and overlapping confidence intervals. There is no clear rationale for the company's waning assumptions, but they do potentially mitigate against uncertainty.	We opt for the constant HR reported by the MAJIC-PV investigators, which is more appropriate from a statistical perspective. We also report scenarios with more conservative waning scenarios.
Treatment to treatment discontinuation			
Primary analysis TTD for ruxolitinib due to reasons other than death from competing risk analysis	Odds spline with 1 knot for RESPONSE and RESPONSE-2 (separate competing risk analyses)	There is the potential for overfitting data using an odds spline model, and a parametric model is preferred.	The EAG selects a Weibull distribution for the extrapolation of data for both RESPONSE and RESPONSE-2.
Utilities			
Health state utilities	MF-8D from RESPONSE trial for base case (EQ-5D from RESPONSE-2 for scenario).	Although the company comments on the use of the MF-8D for myelofibrosis in previous appraisals, the MF-8D was not designed for patients with	The EAG uses the EQ-5D utility values in our preferred analysis. This is in accordance with NICE preferred methods and allows for

139

Analysis	Company analysis	EAG comment	EAG additional analyses
		polycythaemia vera. Assumptions were made in order to obtain PV symptom scores in place of myelofibrosis symptoms scores. There is a lack of direct evidence validating the EQ-5D and MF-8D in patients with PV.	consistency across NICE appraisals.
Resource use and costs			
Thromboembolic events	The company assume a cost equivalent with one emergency department visit, £297, for the management of all Grade 1 and 2 thromboembolic events.	EAG clinical experts suggested a higher cost associated with the management of Grade 1 and 2 thromboembolic events, taking into account the processes required to confirm and treat such an event.	The EAG applies additional costs in the base case for a D-dimer test, vascular ultrasound, and a single dose of an anticoagulant.

140