CONFIDENTIAL UNTIL PUBLISHED

Evidence Review Group Report commissioned by the NIHR Systematic Reviews Programme on behalf of NICE

Avapritinib for treating unresectable or metastatic gastrointestinal stromal tumours

Produced by	**Southampton Health Technology Assessments Centre (SHTAC)**
Authors	Keith Cooper, Senior Research Fellow, Health Economics Neelam Kalita, Senior Research Fellow, Health Economics Inês Souto Ribeiro, Senior Research Assistant, Health Economics Lois Woods, Senior Research Assistant, Evidence Synthesis 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) Wessex Institute, Faculty of Medicine, University of Southampton Alpha House, Enterprise Road Southampton Science Park Southampton SO16 7NS www.southampton.ac.uk/shtac
Date completed	24th June 2020. Updated 13th July 2020 after factual accuracy check.

Source of Funding: This report was commissioned by the NIHR Systematic Reviews Programme as project number 131435.

Acknowledgements

We thank the following for providing clinical advice and comments on the draft report: Dr Peter Simmonds, Consultant Medical Oncologist, Southampton Oncology Centre, University Hospital Southampton NHS Foundation Trust; and Professor Ian Judson, Emeritus Professor of Cancer Pharmacology, The Institute of Cancer Research. We also thank Dr Jonathan Shepherd (SHTAC) for acting as an internal editor for the draft report.

Declared competing interests of the authors and expert clinical advisors

None from the authors. Dr Simmonds attended a local hospital event sponsored by Novartis (manufacturer of imatinib) at which the drug ribociclib was discussed. Professor Judson declared that he had provided informal opinion to the company on some aspects of their proposed submission to NICE at the British Sarcoma Group meeting, Glasgow, in February 2020. The opinion provided is stated in the CS on pages 96, 105, 111 and 114. Professor Judson confirmed that he was not formally instructed by the company, did not provide any written evidence to the company, did not enter into any contract with the company, and did not receive any reimbursement or other incentive from the company.

Rider on responsibility for report

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

This report should be referenced as follows:

Cooper K, Kalita N, Souto Ribeiro I, Woods L, Scott DA, Frampton G. Avapritinib for treating unresectable or metastatic gastrointestinal stromal tumours*:* A Single Technology Appraisal. Southampton Health Technology Assessments Centre, 2020.

Contributions of authors

Geoff Frampton critically appraised the clinical effectiveness systematic review, drafted the report, project managed the report and is the project guarantor. Keith Cooper critically appraised the health economic systematic review, critically appraised the economic

evaluation, and drafted the report; Neelam Kalita critically appraised the health economic systematic review, critically appraised the economic evaluation, and drafted the report; Inês Souto Ribeiro 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, conducted bibliographic searches, and drafted the report; David Scott critically appraised the clinical effectiveness systematic review and drafted the report.

Word count: 35,403

Commercial in confidence (CIC) information in blue
Academic in confidence (AIC) information in yellow

Table of Contents

LIST OF TABLES

LIST of FIGURES

LIST of APPENDICES

LIST OF ABBREVIATIONS

AE	Adverse event
AIC	Akaike information criterion
BIC	Bayesian information criterion
BNF	British National Formulary
BSC	Best supportive care
CBR	Clinical benefit rate
CHMP	Committee for Medicinal Products for Human Use
CNS	Central nervous system
CR	Complete response
CS	Company submission
CSR	Clinical study report
DoR	Duration of response
DSU	Decision Support Unit
ECG	Electrocardiogram
ECM	Established clinical management
ECOG	European Cooperative Oncology Group
EMA	European Medicines Agency
EORTC	European Organisation for Research and Treatment of Cancer
ERG	Evidence Review Group
EQ-5D	EuroQol 5-Dimension
FDA	Food and Drug Administration
GIST	Gastrointestinal stromal tumour(s)
HR	Hazard ratio
HRG	Healthcare Resource Group
HRQoL	Health-related quality of life
ICER	Incremental cost effectiveness ratio
IPW	Inverse probability weighting
ITC	Indirect treatment comparison
KM	Kaplan-Meier
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
ORR	Overall response rate
OS	Overall survival

PAS	Patient access scheme
PD	Progressive disease
PFS	Progression-free survival
PR	Partial response
PSSRU	Personal Social Services Research Unit
QALY	Quality adjusted life year
QD	Once daily (quaque die)
RCT	Randomised controlled trial
SAP	Statistical analysis plan
SOC	Standard of care
STA	Single Technology Appraisal
TKI	Tyrosine kinase inhibitor
ToT	Time on treatment
TSD	Technical Support Document
WHO	World Health Organisation

1 EXECUTIVE SUMMARY

The relevant population for this Technology Appraisal is patients who have unresectable or metastatic gastrointestinal stromal tumours (GIST) and who also have the *PDGFRA* D842V mutation. This is a very small subset of the overall GIST population, with approximately five incident cases expected in England and Wales each year. A key feature of the *PDGFRA* D842V mutation is that the tyrosine kinase inhibitors (TKIs) recommended by NICE for treating unresectable or metastatic GIST (first-line imatinib, second-line sunitinib or third-line regorafenib) are clinically ineffective in people who have this mutation.

The company submission compares the clinical effectiveness, safety, and cost-effectiveness of the intervention, avapritinib, against established clinical management (ECM), where ECM represents the use of TKIs and/or best supportive care (BSC). Given that the established TKI therapies are clinically ineffective in people with the *PDGFRA* D842V mutation, a majority of the patients in current clinical practice would be expected to receive BSC, although there may be some exceptions. Avapritinib is a new type of TKI inhibitor that inhibits *PDGFRA* D842V, thereby suppressing tumour cell proliferation, and is expected to be a first-line therapy if approved by NICE.

The company included three single-arm studies as sources of clinical effectiveness evidence for this Technology Appraisal. The Evidence Review Group (ERG) agree that these represent the best available evidence and that no relevant studies have been missed:

- NAVIGATOR: the pivotal company-sponsored prospective, single-arm, phase I/II study of avapritinib (N=56 *PDGFRA* D842V patients)
- BLU-285-1002: a company-sponsored retrospective chart review of ECM (N=19 *PDGFRA* D842V patients)
- An independent retrospective chart review study of ECM by Cassier et al (2012) which we refer to as the Cassier study (N=32 *PDGFRA* D842V patients)

In each study the relevant population of unresectable/metastatic GIST patients with the *PDGFRA* D842V mutation is a subset of a wider population of people with either unresectable/metastatic GIST (NAVIGATOR and Cassier studies) or locally advanced or unresectable/metastatic GIST (BLU-285-1002 study). NAVIGATOR is an ongoing study.

Given the lack of controlled trials, the company conducted indirect treatment comparisons (ITC) between avapritinib and ECM. An adjusted ITC was feasible for the comparison of NAVIGATOR against BLU-285-1002, i.e. adjusting for baseline imbalances in the population characteristics of the studies; but only an unadjusted (naïve) comparison was possible between NAVIGATOR and the Cassier study. The ERG agree that the overall approach to the data synthesis is appropriate.

Hazard ratios for ECM versus avapritinib from the adjusted ITC are [redacted] for overall survival (OS) and [redacted] for progression-free survival (PFS). As summarised below, and discussed in detail in this report, these results are subject to considerable uncertainty due to immaturity of the survival outcomes data (median OS was not reached), small sample sizes, inherent risks of bias, limitations in the company's studies, and limitations in the ITC methodology. However, these are currently the best data available for this technology appraisal.

A cohort partitioned survival model was developed by the company to assess the cost effectiveness of avapritinib compared to ECM. The model consists of five health states (i.e. first-line PFS, second-line PFS, third-line PFS, progressed disease, and death), and has monthly cycles and a lifetime horizon of 40 years. Patients transition to further lines of treatment according to their progression rate. Those in the avapritinib arm who progress are assumed to receive BSC and will not subsequently be treated with TKIs. Patients in the ECM arm are assumed to receive imatinib as first-line, sunitinib as second-line, and regorafenib as third-line therapy. After failing third-line therapy, patients are assumed to receive BSC (i.e. no further TKIs).

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

There are some minor differences between the NICE scope and company's decision problem in how ECM and BSC are described, but the ERG agree that the company's decision problem is appropriate. The key points to note are:

- Health-related quality of life (HRQoL) is specified as an outcome in the decision problem but was not assessed in the included clinical effectiveness studies (the company obtained HRQoL data for the economic model from the published literature).
- The NICE scope specifies that the company's economic analysis should include the costs of *PDGFRA* D842V mutation testing. However, the ERG believe that all

patients would be routinely tested for this mutation on diagnosis of GIST so there would be no mutation testing costs to include.

1.2 Summary of the key issues in the clinical effectiveness evidence

- There is uncertainty in the clinical treatment pathway, regarding the proportions of PDGFRA D842V patients who would receive imatinib, sunitinib, regorafenib and/or BSC. This differs between the company's clinical studies (the majority of patients received prior TKIs) and what would be expected UK clinical practice (most patients would receive BSC) (see sections 2.2.3.1 and 3.2.1.3 of this report). It is unclear whether some of this uncertainty could be resolved by wider clinical consultation or company clarification.
- Survival outcomes are immature which increases uncertainty (sections 3.2.4 and 3.2.5). This issue is not resolvable until the NAVIGATOR study is completed (or a more recent data cut provided).
- The clinical effectiveness evidence is based on small sample sizes which increases uncertainty (section 3.2.1). This issue is not resolvable unless additional data are collected – difficult due to the small number of people with the *PDGFRA* D842V mutation.
- The ECM comparators were retrospective and hence at risk of selection bias (risk of 'cherry-picking' existing data) (sections 3.2.2 and 3.4.5). This issue is not resolvable without conducting further, prospective, protocol-based, studies (or retrospective studies with random sampling and blinding) – difficult due to the small number of people with the *PDGFRA* D842V mutation.
- There is a lack of head-to-head comparative controlled studies of avapritinib versus ECM (sections 3.2.1 and 3.4.1). This issue is partly resolvable by conducting ITC analyses, albeit with uncertainties remaining due to inherent limitations in the studies and in the ITC methodology.
- Performance status score, tumour size and specific prior TKIs received could not be included as covariates in the analysis due to data limitations. It is unclear whether these would be influential as prognostic factors. This issue is not resolvable unless additional data are collected – difficult due to the small number of people with the *PDGFRA* D842V mutation.
- An adjusted ITC is not feasible for comparing the NAVIGATOR and Cassier studies due to limitations of reporting in the Cassier study; results of the alternative, unadjusted, ITC are highly uncertain (section 3.5.1). This issue might be resolvable if further data or clarification could be obtained from the Cassier study authors.

However, although the Cassier study is included in a scenario analysis (section 4.2.6), results of the unadjusted ITC do not inform the economic analysis.

- HRQoL data are lacking for people with the D842V mutation who receive avapritinib (section 3.2.5.7). This issue is partly resolvable by using HRQoL data from alternative sources (e.g. the published literature). Interim HRQoL data from a company-sponsored randomised controlled trial of avapritinib versus regorafenib (VOYAGER) are included in an ERG scenario analysis (section 6.2).

1.3 Summary of the key issues in the cost effectiveness evidence

- Whilst the model population is appropriate for the scope and the anticipated marketing authorisation, patients in the economic model are assumed to have no previous TKIs unlike those in the NAVIGATOR and BLU-285-1002 studies. Further, as noted above, the prior TKI use in these studies does not reflect the UK clinical practice. This means that there is uncertainty around the appropriateness of the modelled patient population (see sections **Error! Reference source not found.** and **Error! Reference source not found.**).
- The modelled outcomes provide a poor fit to observed OS Kaplan-Meier data for avapritinib (OS for avapritinib is overestimated). The model includes persistence of treatment benefits of avapritinib for five years with a gradual reduction of the treatment benefit over this time. Clinical experts advised the ERG that this was unlikely to be plausible and that patients who discontinue avapritinib would rapidly progress to a similar death rate as untreated patients (see section 4.2.6).
- The modelled outcomes do not provide a close fit to the observed Time on Treatment (ToT) Kaplan-Meier data for avapritinib. In addition, the ERG note that there are further inconsistencies in modelling ToT for the dose intensity of the comparator treatments. These issues produce a significant underestimate of the treatment cost for avapritinib (see section 4.2.6).
- Health utility values for first-line therapy for avapritinib and ECM appear to be implausible. The utility value used in the company's base case for patients with an initial age of [redacted] years is higher than the utility value of the general population in this age group. Clinical advice to the ERG suggests that these patients would have a lower or similar utility compared to that of the general population (see section 4.2.7).
- The survival models used, for OS and ToT, differ between treatment arms. To align with recommendations in NICE DSU TSD 14, we view it appropriate to use the same survival model for both treatment arms (see section 4.2.6).

1.4 Summary of the ERG's preferred assumptions and resulting ICER

The ERG's preferred assumptions are shown below:

- Proportion of patients receiving TKIs in ECM assumed to be 20% imatinib, 10% sunitinib, 10% regorafenib.
- Dose intensity: Assumed the same for all TKIs.
- Duration of treatment waning: 1 month.
- Extrapolation of survival models for OS, PFS and ToT: Uses a Weibull distribution.
- Estimating ToT for avapritinib: Uses PFS as a proxy.
- All-cause mortality: Updated to ONS 2016-2018.
- Utility values for avapritinib / first-line TKI for ECM: use the general population norm.
- Resources for progressed disease: reduced resource use for patients with progressed disease (a third of patients would no longer have investigations).

The ICER using the ERG's preferred assumptions is shown in Table 1. The ICER for avapritinib versus ECM is [redacted] per QALY gained.

Table 1 ICER resulting from ERG's preferred assumptions

	Total Costs	Total QALYs	Change in costs	Change in QALYs	ICER £/QALY
Avapritinib	[redacted]	[redacted]	-	-	-
ECM	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

1.5 Summary of exploratory and sensitivity analyses undertaken by the ERG

The ERG conducted a range of scenario analyses using our preferred assumptions as outlined below:

- Varying patients' initial age;
- Using different model time horizons;
- Varying duration of treatment waning for avapritinib;
- Including drug costs of the additional TKIs in the BLU-285-1002 study, which are not currently approved for the treatment of GIST patients in England and Wales;
- Varying the percentage of incomplete loss of treatment benefit after discontinuation for the avapritinib arm;

- Varying the post-progression rate for the avapritinib arm;
- Using alternative sources to inform model parameters such as End of Life costs, resource use, and utilities;
- Using the Cassier study as a source for comparator clinical effectiveness; and
- Assigning different survival distributions to extrapolate OS and PFS.

Results and details of these analysis are provided in section **Error! Reference source not found.**.

Across all the scenarios, the ICERs for avapritinib versus ECM remain above £50,000 per QALY. The scenarios that significantly influence the cost-effectiveness results are: using a shorter time horizon, extrapolating the OS curves using the exponential distribution, varying the duration of treatment waning for avapritinib and using the Cassier study to inform ECM clinical effectiveness. The remaining scenarios also influence the cost effectiveness results, but to a lesser extent.

EVIDENCE REVIEW GROUP REPORT

2 INTRODUCTION AND BACKGROUND

2.1 Introduction

This report is a critique of the company's submission (CS) to NICE from Blueprint Medicines on the clinical effectiveness and cost effectiveness of avapritinib for treating gastrointestinal stromal tumours. It identifies the strengths and weaknesses of the CS. Clinical experts were consulted to advise the evidence review group (ERG) and to help inform this report.

Clarification on some aspects of the CS was requested from the company by the ERG via NICE on 12th May 2020. A response from the company via NICE was received by the ERG on 29th May 2020 and this can be seen in the NICE committee papers for this appraisal.

2.2 Background

Sections B.1.2 and B.1.3 of the company submission (CS) discuss the disease, gastrointestinal stromal tumours (GIST), the intervention (avapritinib) and its position in the treatment pathway. To support the evidence presented in the submission, the company carried out a survey of five clinicians, who are experts in the disease, to provide information on current clinical practice.[1] Of the two clinical experts advising the ERG, one agreed broadly with the opinions in the company's clinician survey whilst the other disagreed with some of the opinions, illustrating that there is uncertainty in clinical practice.

2.2.1 Background information on unresectable or metastatic gastrointestinal stromal tumours (GIST)

GIST is a type of gastrointestinal tumour that arises in the interstitial cells of the Cajal. It can occur anywhere along the gastrointestinal tract, but the most common site is the stomach.

It is possible for GISTs to be asymptomatic or silent, but where there are symptoms these may include abdominal pain, obstruction, palpable mass, upper or lower GI bleeding, anaemia, or dysphagia, and these may differ according to tumour site. The CS also lists non-specific systemic symptoms and discusses fatigue and fear in relation to the patient disease burden.

For patients presenting with localised disease surgery is expected as a cure, and only a small proportion of patients progress to or present with unresectable or metastatic disease. Patients with the *PDGFRA* D842V mutation generally have good prognosis and only around

5-6% progress to have unresectable or metastatic disease. However, when surgical resection fails, as this mutation is known to be resistant to current treatments, prognosis is the same as for any untreated patient with progressive disease.

GISTs are rare. They account for 0.1 to 3.0% of all gastrointestinal malignancies.[2] The most recent UK prevalence study estimates a prevalence of third-line treatment-eligible GIST of 1/100,000 and a prevalence count of 598.[3] This is similar to the European studies which estimate an incidence of 1 to 1.5/100,000 per year of GIST.[4,5] There are an estimated 650 new cases per year in the UK, 900 in total, and the median age at diagnosis is 60 to 65 years but the range is wide.[6]

The CS estimates that in England and Wales there are 30-40 patients with unresectable or metastatic GIST with the *PDGFRA* D842V mutation, with about 5 new cases per year.

2.2.2 Background information on avapritinib

- CS Table 2 presents information on avapritinib.

- Avapritinib is a Type 1 tyrosine kinase inhibitor (TKI) that has been shown in vitro to inhibit activity of several *PDGFRA* exon 18 mutants and several KIT exon 11, 11/17 and 17 mutants. The *PDGFRA* D842V mutation is the most common of the exon 18 mutations, and patients with this mutational status are the population of interest for this submission.

- Avapritinib was granted an EMA orphan drug designation for the treatment of GIST in August 2017;[7] [redacted]. Avapritinib received an FDA fast track and orphan drug designation and was granted FDA approval for the treatment of adults with unresectable or metastatic GIST harbouring a *PDGFRA* exon 18 mutation (including D842V) in January 2020.[8]

- The intended licensed dosage is 300mg once daily, taken orally, until disease progression or unacceptable toxicity.

2.2.3 The position of avapritinib in the treatment pathway

2.2.3.1 Current treatment pathway

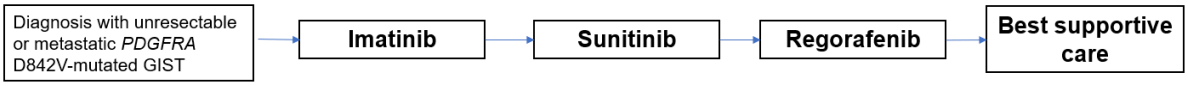
The CS outlines three current clinical guidelines for the treatment of GIST: The British Sarcoma Group for UK guidelines,[6] ESMO European guidelines,[9] and The National Comprehensive Cancer Network guidelines for the USA.[10] The ERG's clinical experts both

indicated that the guidelines most used in England are the UK (BSG) and European (ESMO) guidelines which are similar, with the BSG ones adapted to reflect UK drug availability.

On disease progression, NICE guidance approves sequential administration of the TKIs imatinib, sunitinib and regorafenib as first-, second- and third-line treatments respectively for unresectable or metastatic GIST.[11-13] Patients with the *PDGFRA* D842V mutation are known to be resistant to treatment with existing TKIs.[14,15] This is acknowledged in the guidelines; however, neither the NICE guidance, nor the clinical guidelines provide recommendations for treating unresectable or metastatic GIST in patients with the *PDGFRA* D842V mutation, as currently no known effective treatment is available. The clinical guidelines only say that patients failing on treatment can be considered for inclusion in clinical trials of new agents. Therefore, in UK clinical practice, it is not certain that patients with unresectable or metastatic GIST who have the *PDGFRA* D842V mutation would be treated with all three TKIs sequentially.

The company's view of the clinical pathway for patients with the *PDGFRA* D842V mutation, as used in the economic model, does not differ from the UK clinical pathway for the general unresectable or metastatic GIST population (CS section B.1.3.3.2). It is reproduced in Reproduced from CS Figure 1

Figure 1 below.



Reproduced from CS Figure 1

Figure 1 Company view of the current clinical pathway for patients with unresectable or metastatic GIST with the *PDGFRA* D842V mutation

The company refer to the use of TKIs (imatinib, sunitinib, regorafenib) and/or best supportive care (BSC) together as comprising "established clinical management" (ECM) in their decision problem (see section 2.3 below). The company do not explicitly define BSC. According to the ERG's clinical experts, BSC could include non-drug therapy such as surgery or ablation for specific lesions, with palliative intent.

According to the ERG's clinical experts, the company's clinical pathway for people who have unresectable or metastatic GIST and who have the *PDGFRA* D842V mutation is not reflective of UK clinical practice for the following reasons:

- Since imatinib, sunitinib and regorafenib lack clinical effectiveness among patients with the PFGFRA D842V mutation and carry a toxicity burden they would not usually be prescribed for this subgroup (in the company's clinician survey only two out of five clinicians responded that they would treat these patients with TKIs[1]).

- Patients with the *PDGFRA* D842V mutation might have received imatinib before their mutational diagnosis is known, for which they could wait up to three or four weeks. Most would discontinue imatinib once confirmed to have the *PDGFRA* D842V mutation.

- Among those patients who do receive imatinib, very few, if any, would subsequently receive sunitinib or regorafenib, due to lack of effectiveness and risk of toxicity.

In summary, whilst we agree that ECM (comprising TKIs and/or BSC) is an appropriate comparator, we disagree that the relative balance of TKIs and BSC in the company's clinical pathway reflects UK practice. Patients with the *PDGFRA* D842V mutation in UK clinical practice would predominantly receive BSC, with relatively few receiving imatinib and very few if any would go on to receive sunitinib or regorafenib. However, the clinical experts acknowledged that there is likely to be considerable variation in practice.

We note that whilst the whilst company's clinical pathway does not align with expected UK clinical practice, it does align with the company's studies, in which some patients with the *PDGFRA* D842V mutation received all three TKIs (imatinib, sunitinib, regorafenib). This is discussed further in section 3.2.1.3 below.

2.2.3.2 Treatment pathway with avapritinib

It is expected that, since the *PDGFRA* D842V mutation is resistant to other TKIs, avapritinib will be the first line of treatment after diagnosis with unresectable or metastatic disease and confirmation of mutational status.

There remains the possibility that a patient may have been receiving imatinib first line whilst waiting for the results of mutational diagnosis which can take up to three to four weeks from testing. In these cases, imatinib would be discontinued on confirmation of a D842V mutation and avapritinib would be given.

After failing to respond to avapritinib patients would receive BSC (CS section B.3.2.4).

ERG conclusion

- The ERG do not agree that the current clinical pathway, as represented in CS Figure 1, is representative of UK clinical practice for patients who have the *PDGFRA* D842V mutation.
- However, there is uncertainty around the use of TKIs for patients who have the *PDGFRA* mutation. Clinical experts consulted by both the company and the ERG had differing views around giving patients ineffective but toxic treatment (also at high monetary cost).

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

Error! Reference source not found. summarises the decision problem addressed by the company in the CS in relation to the final scope issued by NICE and the ERG's comments on this.

The company's decision problem is broadly consistent with the NICE scope, but the following points should be noted:

- The NICE scope and decision problem give different definitions of ECM (see Table 2). The NICE scope definition of ECM includes BSC but does not appear to include the TKIs (imatinib, sunitinib and regorafenib). However, the company's decision problem defines ECM as including the TKIs and BSC. We agree that the company's definitions of the comparators are appropriate and reflect how ECM is modelled in the economic analysis (see section 4.2.2). (NB: as discussed in section 2.2.3 above, whilst ECM is appropriate as an overall comparator, the relative balance of TKIs and BSC differs between the company's ECM pathway and that which would be expected in UK clinical practice.)
- The NICE scope specifies that the company should include the costs of *PDGFRA* D842V mutation testing in their economic analysis. However, mutational testing for *PDGFRA* D842V is done routinely on diagnosis of GIST, meaning that there are no additional mutation testing costs relevant to avapritinib that would need to be included.

Table 2 Summary of the decision problem

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	ERG comments
Population	Adults with unresectable or metastatic GIST and the platelet-derived growth factor receptor alpha (*PDGFRA*)	[redacted].	This is the population for which avapritinib is anticipated to receive its marketing authorisation	The decision problem population matches the NICE scope and the intended licensed

	D842V mutation regardless of prior therapy.		from the EMA and is in line with the evidence presented in the pivotal NAVIGATOR study.	indication and is consistent with the *PDGFRA* D842V mutation subgroup in the pivotal NAVIGATOR study.
Intervention	Avapritinib	Avapritinib	Not applicable	Not applicable
Comparators	• Imatinib (for adults who have KIT [CD117]-positive tumours) • Sunitinib (for adults whose treatment with imatinib has failed due to resistance or intolerance) • Regorafenib (for adults whose disease has progressed on, or who are intolerant to, prior treatment with imatinib and sunitinib) • Established clinical management without avapritinib including best supportive care	Established clinical management without avapritinib including: • Imatinib • Sunitinib (for adults whose treatment with imatinib has failed due to resistance or intolerance) • Regorafenib (for adults whose disease has progressed on, or who are intolerant to, prior treatment with imatinib and sunitinib) • Best supportive care	The appropriate comparators have been selected for the anticipated licensed population for avapritinib in line with clinical opinion.	The comparators are worded differently in the NICE scope and decision problem. However, we agree with the company's definition of the comparators which aligns with how ECM is modelled in their economic analysis (an ECM comparator arm includes imatinib, sunitinib, regorafenib and BSC).

Outcomes	The outcome measures to be considered include: • Overall survival • Response rate (including partial response rate and duration of response) • Progression-free survival • Adverse effects of treatment • Health-related quality of life	The outcome measures to be considered include: • Overall survival • Response rate (including partial response rate and duration of response) • Progression-free survival • Adverse effects of treatment • Health-related quality of life • Time on treatment	Time on treatment is an important outcome of interest for use in the economic model, as tracking patient outcomes via line of therapy avoids the issue of noncomparability of progression across treatments	All outcomes in the NICE scope are included in the decision problem. We note that health-related quality of life (HRQoL) was not assessed in the pivotal avapritinib study and the company have sourced HRQoL data in their economic analysis from other sources (section 4.2.7). The additional inclusion of time on treatment in the decision problem is appropriate, as this outcome informs the economic model.
Economic analysis	• The cost effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year	• The cost effectiveness of treatments is expressed in terms of incremental cost per quality-adjusted life year	According to clinical experts [company's expert opinion survey], nearly all patients will have their mutational status known before or within three	The company's assumption that all GIST patients would be routinely tested for the *PDGFRA* D842V mutation in clinical practice is appropriate.

	• The time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared • Costs will be considered from an NHS and Personal Social Services perspective • The availability of any commercial arrangements for the intervention, comparator and subsequent treatment technologies will be taken into account • The use of avapritinib is conditional on the presence of the *PDGFRA* D842V mutation. The economic modelling should include	• The time horizon runs until over 99% of patients have died in both treatment arms • Costs are considered from an NHS and Personal Social Services perspective • Where known, commercial arrangements for the intervention, comparator and subsequent treatment technologies are taken into account • The clinical evidence is based only on eligible (i.e. metastatic or unresectable) patients with the *PDGFRA* D842V mutation	weeks of diagnosis with unresectable or metastatic GIST.	Routine *PDGFRA* D842V mutation testing is recommended by the relevant UK guidelines (the British Sarcoma Group guidelines say that "mutational testing is obligatory" in GIST [6]) and the ERG's clinical advisors agreed that all GIST patients would be routinely tested for this mutation on diagnosis. There are therefore no additional mutation testing costs relevant to avapritinib that would need to be included.

	the costs associated with diagnostic testing for the *PDGFRA* D842V mutation in people with unresectable or metastatic GIST who would not otherwise have been tested. A sensitivity analysis should be provided without the cost of the diagnostic test. See section 5.9 of the Guide to the Methods of Technology Appraisals			
Subgroups	Not applicable	Not applicable	Not applicable	Not applicable
Special considerations including issues related to equity or equality	Not applicable	Not applicable	Not applicable	Not applicable
Source: CS Table 1				

3 CLINICAL EFFECTIVENESS

3.1 Critique of the methods of the company's systematic literature review

The systematic literature review performed by the company is reported in CS Appendix D, and the ERG's assessment of the review is summarised in Table 3 below. Overall the company's review is fit for purpose and we believe all relevant studies have been identified. However, we disagree with the company's risk of bias assessment approach, as explained in section 3.2.2 and Appendix 2 in this report.

Table 3 ERG appraisal of systematic review methods

Systematic review components and processes	ERG response	ERG comments
Was the review question clearly defined using the PICOD framework or an alternative?	Yes	The PICOS (S=study type) is defined in Appendix Table 5 for the eligibility criteria. It matches the decision problem.
Searches: was the literature review carried out appropriately (sources, date range, in line with PICOD, correct search terms/syntax, etc.)?	Yes	Reported in CS Appendix D.1 See Appendix 3 for detailed ERG comments.
Searches: were any relevant studies missed?	No	The identified studies are listed in CS Appendix Tables 13 and 14. The ERG and our clinical experts are not aware of any missing studies.
Were inclusion and exclusion criteria specified? If so, were these criteria appropriate and relevant to the decision problem?	Yes Yes	CS Appendix Table 5.
Were study selection criteria applied by two or more reviewers independently?	Yes	CS Appendix D.1 page 19 Two independent reviewers for both level 1 and 2 screening, with disagreements checked by a third reviewer.

Was data extraction performed to a reasonable standard (e.g. use of two reviewers)?	Yes	CS Appendix D.1 page 19 Data extraction was performed by one researcher and verified against the original source by a second researcher.
Was a risk of bias assessment or a quality assessment of the included studies undertaken? If so, which tool was used?	Yes	Downs and Black checklist.[16] CS Appendix D.3 and CS Appendix Table 19. Discussed in section 3.2.2 and Appendix 2 in this report.
Was risk of bias assessment (or other study assessment) conducted by two or more reviewers independently?	Yes	Not reported but clarified by company at factual error check stage (NB ERG disagree with company approach to risk of bias assessment – see section 3.2.2)
Is sufficient detail on the individual studies presented?	Partly	CS Tables 14 and 15 and CS Appendix Tables 14 and 15 (baseline characteristics). Limited data for BLU-285-1002 are given in the CS, so ERG have sourced these from the CSR.
If statistical evidence synthesis (e.g. pairwise meta-analysis, ITC, NMA) was undertaken, were appropriate methods used?	Yes	See section 3.4 of this report

3.1.1 ERG summary of the company's literature searches

The company performed a sensitive search of the literature including all relevant and recommended sources. The search included terms for all approved or investigational pharmacological interventions used to treat GIST except that there were no terms used to express BSC and, therefore, the search may have missed any BSC-only studies. By the time of receipt of the CS the searches were over five months out of date. We therefore ran updated searches in Medline, Embase and the Cochrane Library, and checked Medline for any BSC-only studies with no date limit. We found no further relevant studies. The ERG is satisfied that the review was carried out to a good standard, albeit with some lack of clarity of reporting, and that it was appropriate to this appraisal. For reference, detailed ERG comments on the searches are given in Appendix 3.

3.2 ERG critique of the included clinical effectiveness studies

3.2.1 Included studies

The company's systematic literature search and study selection process identified the following seven studies relevant to the decision problem. All of these studies except BLU-285-1002 included a mix of GIST patients with and without the *PDGFRA* D842V mutation, meaning that the *PDGFRA* D428V population relevant to the current appraisal is a subgroup from each study (except BLU-285-1002). As a consequence, sample sizes (N) are small (only 3 to 12 patients in four studies), with the largest *PDGFRA* D842V subgroup sizes being in the NAVIGATOR, BLU-285-1002 and Cassier studies (22 to 56 patients):

- NAVIGATOR; a company-sponsored multinational single arm prospective study on avapritinib (N=56)[17-21]
- BLU-285-1002: a company-sponsored retrospective chart review of patients at three centres in the USA who had received imatinib, sunitinib and regorafenib (N=22)[22,23]
- Cassier et al. 2012: an international survey of GIST referral centres on patients who had received imatinib first-line and sunitinib (N=32)[24,25]
- Rutkowski et al. 2012: a retrospective single-centre registry of Polish GIST patients who had received sunitinib (N=12)[26]
- Yoo et al. 2016: a retrospective single-centre registry of Korean GIST patients who had received imatinib and sunitinib (N=9)[27]
- Osuch et al. 2014: a retrospective multi-centre registry of Polish GIST patients who received imatinib (N=8)[28]
- B222: a multi-centre randomised controlled trial (RCT) of GIST patients who received two doses of imatinib (N=3)[29]

The NAVIGATOR, BLU-285-1002 and Cassier studies contribute to the company's economic analysis as follows:

- NAVIGATOR (avapritinib) and BLU-285-1002 (ECM) were compared in an adjusted indirect treatment comparison (ITC) which informs the company's economic model base case.
- NAVIGATOR (avapritinib) and the Cassier study (ECM) were compared in an unadjusted ITC (CS Appendix Table 15 and Appendix P), with the survival outcomes informing a scenario analysis (CS section B.3.8.3).

We have therefore focused the current report on the characteristics and results of the NAVIGATOR, BLU-285-1002 and Cassier studies. The remaining four studies (B222, Rutkowski, Osuch, Yoo) are discussed narratively by the company (CS Appendix Tables 13, 14, 16, 17 and accompanying text) and the B222 trial was considered by the company as a potential source of health utility data (as noted in section 4.2.7 below). These four studies do not inform the economic model and are not discussed further in this report because they are limited by their very small sample sizes (and none included any UK patients), among other limitations which are summarised in CS Table 14.

Ongoing studies

The NAVIGATOR study is currently ongoing, with incomplete follow-up of survival outcomes. Study outcomes are reported for two interim data cuts (see Table 4 below). The company advised in clarification response A1 that the final CSR for NAVIGATOR will not be available until [redacted].

One other relevant ongoing study, VOYAGER, was identified by the company (CS section B.2.11). This is an open-label company-sponsored RCT comparing avapritinib against regorafenib in patients with locally advanced, unresectable or metastatic GIST previously treated with imatinib and one or two other TKIs. The numbers enrolled are not clearly reported. VOYAGER includes a subgroup of patients with the *PDGFRA* D842V mutation, but only 12 of these patients have been recruited (six in each treatment group). The company confirmed in clarification response A1 that a CSR for VOYAGER is not currently available ([redacted]). However, on request from the ERG (clarification question B6) the company provided HRQoL data from VOYAGER for inclusion in an ERG scenario analysis (see section 6.2). These HRQoL data were the only VOYAGER outcomes available for the ERG to consider at the time of preparation of this report.

The ERG searches did not identify any other ongoing studies of avapritinib or the comparators in the decision problem that would be completed within the timeframe of the current appraisal. Ongoing studies that we are aware of are:

- INVICTUS (RCT: ripretinib versus placebo) is expected to complete in December 2020. This trial only has 3 *PGDFRA* D842V patients.
- INTRIGUE (RCT: ripretinib versus sunitinib) is not due to complete until March 2022. No data have been published yet. It is unclear how many *PDGFRA* D842V patients have been enrolled so far.

3.2.1.1 Study characteristics

Characteristics of the three key studies are summarised in Table 4. Few details of BLU-285-1002 are reported in the CS and so we have sourced these from the CSR. The NAVIGATOR study was conducted prospectively whilst both the comparator studies retrospectively collated patient data from clinical records.

The NAVIGATOR study included patients who received a range of daily doses of avapritinib (30mg, 60mg, 90mg, 135mg, 200mg, 300mg, 400mg, 600mg) and the company's analyses of clinical effectiveness outcomes are based on the "all doses" pooled group to maximise the available sample size (N=56). The analysis population therefore included [redacted] patients ([redacted]%) who had received the intended licensed indication dose of 300mg, [redacted] patients ([redacted]%) who had received lower doses, and [redacted] patients ([redacted]%) who had received higher doses. Clinical experts advising the ERG consider that the company's dose pooling approach is appropriate. This is based on experience with other TKIs that suggests clinical effectiveness outcomes would be unlikely to differ markedly across the included doses, with one expert commenting that dose pooling to increase the sample size is a standard practice in phase I/II studies. The company have provided data separately for the 300mg, 400mg and all-doses groups (but not the lower-dose groups) for baseline characteristics and effectiveness outcomes (CS Appendix L) and for safety outcomes (CS Appendix F). We consider the homogeneity of these dose groups in relation to patients' baseline characteristics (see below); clinical effectiveness outcomes (see section 3.2.5) and safety outcomes (see section 3.3).

The BLU-285-1002 study was designed to serve as a "historical control for efficacy studies of avapritinib" (CS section B.2.9). However, in BLU-285-1002 most patients had initially received adjuvant TKI therapy for locally advanced GIST, which has a different prognosis to the decision problem population, i.e. people with unresectable or metastatic disease. To enable a comparison of BLU-285-1002 against NAVIGATOR, the company reviewed the records of patients in BLU-285-1002 to separate the TKI use that had been received in the adjuvant setting from the TKI use that had been received for unresectable or metastatic disease. The company did this by identifying the first TKI that each patient had received for unresectable or metastatic disease and then including only the patient's data from that point onwards in analyses. This approach enabled 19 of the 22 patients in BLU-285-1002 to be included in comparisons against the NAVIGATOR study.

The Cassier study included 32 patients with unresectable or metastatic GIST who had the *PDGFRA* D842V mutation and can be compared with the NAVIGATOR study population. As discussed further below, a limitation of the Cassier study is that patients' baseline characteristics are reported for the whole study group, not specifically for those with the *PDGFRA* D842V mutation.

Table 4 Overview of the intervention and comparator studies

	Study		
Study feature	**NAVIGATOR**	**BLU-285-1002**	**Cassier et al 2012**
Study design	Prospective, single arm phase I/II study	Retrospective chart review	Retrospective chart review
Status	Ongoing, unpublished	Complete, unpublished	Complete, published
Study population	[redacted]	[redacted]	Adults who had *PDGFRA* mutant advanced or metastatic GIST and had been treated with imatinib in a non-adjuvant setting
Number and location of centres/ data sources	[redacted]	[redacted]	12 European centres plus 2 EORTC clinical trials
Total study population	[redacted]	[redacted]	N=58
Number with the *PDGFRA* D842V mutation	N=56 [redacted]	[redacted]	N=32
Number of UK patients	[redacted]	[redacted]	Not reported whether any UK centres were included
TKI dosing regimens included	[redacted]	[redacted]	Imatinib 400mg QD (n=44) 800mg QD (n=14)
Primary analysis group used for the current appraisal	[redacted] (N=56)	[redacted] (N=19)	*PDGFRA* D842V mutation subgroup (N=32)
Outcomes	[redacted]	[redacted]	Response (CR, PR, SD, PD); OS; PFS
Latest available data	[redacted]	[redacted]	Date of starting imatinib ranged from January 2001 to November 2010
Median duration of follow-up	[redacted]	[redacted]	45.3 months

Source: CS for NAVIGATOR data; CSR for BLU-285-1002 data; Cassier study data from study publication.

CBR: clinical benefit rate; CR: complete response; CSR: clinical study report; DoR: duration of response; EORTC: European Organisation for Research and Treatment of Cancer; KM: Kaplan-Meier; ORR: overall response rate; OS: overall survival; PD: progressive disease; PFS: progression-free survival; PR: partial response; QD: once daily; SD: stable disease; TKI: tyrosine kinase inhibitor; ToT: time on treatment

3.2.1.2 Patients' baseline characteristics

The baseline characteristics of the three key studies are summarised in Table 5. The population characteristics for the BLU-285-1002 and Cassier studies are incompletely reported in the CS and we have therefore sourced these from the CSR[22] and study publication[25] respectively. Some of the characteristics shown in Table 5 were considered by the company to be potential prognostic factors and were adjusted for in the company's indirect comparison between the NAVIGATOR and BLU-285-1002 studies (for discussion of the ITC see sections 3.4 and 3.5).

When comparing the baseline characteristics between the studies it should be borne in mind that the characteristics reported for NAVIGATOR and BLU-285-1002 are for people with the *PDGFRA* D842V mutation whereas the characteristics in the Cassier study are reported for the whole study population, i.e. people with and without the mutation. In the Cassier study just over half of the patients (32/58; 55%) had the *PDGFRA* D842V mutation. Clinical experts advising the ERG suggested that baseline characteristics would be unlikely to differ between people with unresectable or metastatic GIST with or without the PGDFRA D842V mutation. However, the ERG are concerned that differences in TKI use might be expected but this is unclear in the Cassier study due to lack of reporting of patients' baseline characteristics for the mutation subgroup.

Clinical experts advising the ERG agreed that the characteristics of participants in the comparator studies would be similar to those expected in UK clinical practice, with the following exceptions:

- A key feature of participants in NAVIGATOR and BLU-285-1002 is that they received a higher frequency of prior TKIs than would be expected in UK clinical practice (discussed further below – see section 3.2.1.3).
- Participants in the NAVIGATOR and Cassier studies had relatively low ECOG/WHO performance status scores (almost all PS=0 or PS=1) so would have had better performance status than would be expected in clinical practice. Performance status is uncertain for the BLU-285-1002 study due to the majority of data being missing.

- One of the clinical experts advising the ERG commented that the sex distribution of patients, with more than half (range 59% to 70%) being male in each study, is atypical of clinical practice, where a more balanced sex ratio would be expected. The explanation for this difference is unclear but sex is not known to be a prognostic factor and so we believe that the imbalance would not influence effectiveness or safety results. In an analysis of the NAVIGATOR data by the US FDA,[30] sex did not appear to influence treatment outcomes, although firm conclusions are hindered by the small sample size.

Table 5 Patient baseline characteristics in the intervention and comparator studies

Study		**NAVIGATOR**	**BLU-285-1002**	**Cassier et al 2012**
Intervention Population group **Baseline characteristics**		**Avapritinib** All doses group [a] (N = 56)	**Imatinib** Participants receiving their first TKI for unresectable or metastatic disease (N=19)	**Imatinib** Full study population (not limited to *PDGFRA* D842V mutation group) (N=58)
Sex, n (%)	Male	[redacted]	[redacted]	34 (59)
	Female	[redacted]	[redacted]	24 (41)
Age, years, n (%)	< 60	[redacted]	[redacted]	Not reported
	≥ 60	[redacted]	[redacted]	Not reported
Age, years, median (range)		[redacted]	[redacted]	61 (19–83)
Race, n (%)	White	[redacted]	[redacted]	Not reported
	Non-white	[redacted]	[redacted]	Not reported
	Missing	[redacted]	[redacted]	Not reported
Region	US	[redacted]	[redacted]	0
	Europe	[redacted]	[redacted]	58 (100)
	Asia	[redacted]	[redacted]	0
Anatomical site, n (%)	Gastric (stomach)	[redacted]	[redacted]	40 (69)
	Other	[redacted]	[redacted]	18 (31)
Metastatic disease, n (%)		[redacted]	[redacted]	56 (97)
ECOG/WHO performance status, n (%)	0	[redacted]	[redacted]	28 (48)
	1	[redacted]	[redacted]	19 (33)
	2	Not reported	Not reported	2 (3)

	2+	[redacted]	[redacted]	Not reported
	Missing	[redacted]	[redacted]	9 (16)
Duration of disease, n (%)	< 3 years	[redacted]	[redacted]	Not reported
	≥ 3 years	[redacted]	[redacted]	Not reported
Total number of TKI, n (%)	1	[redacted]	[redacted]	29 (50) [e]
	2	[redacted]	[redacted]	21 (36) [e]
	3	[redacted]	[redacted]	5 (9) [e]
	4+	[redacted]	[redacted]	0 [e]
	Unclear	[redacted]	[redacted]	3 (5) [e]
Prior imatinib, n (%)		[redacted]	See footnote [f]	15 (26) [e]
Prior sunitinib, n (%)		[redacted]	See footnote [f]	18 (31) [e]
Prior regorafenib, n (%)		[redacted]	See footnote [f]	0 [e]
Largest target lesion/ primary tumour size by n (%)	≤ 5 cm	[redacted]	See footnote [h]	Not reported
	> 5 to ≤ 10 cm	[redacted]	See footnote [h]	Not reported
	> 10 cm	[redacted]	See footnote [h]	Not reported
	missing	[redacted]	See footnote [h]	Not reported
Other baseline characteristics which are reported in the CS and/or study publication but are not extracted here		[redacted] (CS Table 7 and CSR)	[redacted]	NIH risk group; Miettinen risk group; site of metastatic disease; median tumour size; mitotic rate (Cassier study paper)

Source: CS Tables 7 and 15; CS Appendix Table 14; BLU-285-1002 CSR; [22] Cassier study publication[25]
[a] includes patients with < 300 mg QD and 600 mg QD starting dose.
[b] error in denominator for % in CS Table 15 and ITC report;[31] corrected by ERG
[c] incorrectly reported in CS Appendix Table 14 as 68%
[d] Note these ECOG performance status data are unreliable due to 18/19 (95%) missing
[e] calculated by ERG based on data reported in the text of the study publication (see Appendix 1)
[f] reported in CSR Table 8, but only for the full study population (N=22), therefore would include adjuvant therapy for locally advanced disease so not relevant to the current appraisal. Also reported in CSR Tables 14.1.1.7B and 14.1.1.7C for the unresectable/metastatic GIST group (N=19) but only as partial data that are not comparable with those from NAVIGATOR.
[g] refers to size of the "target lesion" by central radiological assessment
[h] data on the size of the "primary tumour" are reported in the CSR but are for the full study population (N=22) and therefore may not be reflective of tumour size specifically in unresectable or metastatic GIST patients

The company present baseline characteristics for the NAVIGATOR population separately for the 300mg and 400mg dose groups (but not the lower-dose groups) in CS Appendix L (not reproduced here). Due to the relatively small sample sizes it is difficult to draw any firm conclusions about whether the baseline characteristics differ systematically between the

300mg (N=[redacted]), 400mg (N=[redacted]) and all-doses (N=56) groups, but no substantive differences are evident (CS Appendix Table 50).

3.2.1.3 Prior TKI use in the included studies

The frequency of prior imatinib, sunitinib and regorafenib use (Table 5) was higher in the NAVIGATOR study than would be expected in UK clinical practice (see section 2.2.3.1 above). The CS does not explain why patients with the PDGFRA D842V mutation in the NAVIGATOR study received prior TKIs.

Prior TKI use in the BLU-285-1002 study is only reported in the CSR and was also higher than would be expected in UK clinical practice; however, the data are for the full study population (N=22) so presumably include TKI use in the locally advanced GIST setting which would not be relevant to the current appraisal (Table 5).

Baseline characteristics in the Cassier study, including prior TKI use, are reported only for the overall unresectable/metastatic GIST population, not specifically the *PDGFRA* D842V subgroup. Although clinical experts advising the ERG suggested that these two populations would not be expected to differ on baseline characteristics, we are uncertain whether that would apply to prior TKI use which, theoretically, should be different between these populations given that the TKIs are not clinically effective in the *PDGFRA* D842V subgroup.

Clinical experts advising the ERG agreed that the following explanations for the difference in prior TKI use between the avapritinib and ECM studies and UK clinical practice would be plausible:

- Both company studies included US centres (8/19 in NAVIGATOR, 3/3 in BLU-285-1002) which may not be reflective of UK TKI use due to major differences in the US and UK healthcare systems (e.g. US oncologists could receive financial benefits for prescribing ineffective and expensive treatments).
- Patients may have commenced imatinib per standard therapy for unresectable metastatic GIST while awaiting their *PDGFRA* D842V mutation test result. Although patients would usually discontinue imatinib when the *PDGFRA* D842V mutation is confirmed, some might be continued on imatinib for symptom (e.g. pain) control. The CS does not report the time to mutation test results nor the proportion of patients who received late test results in the company studies. Clinical experts advising the ERG noted that mutational status may not be standardised across countries.

- Due to the small size of the *PDGFRA* D842V subgroup and relative lack of clinical experience in treating these patients, clinicians may be heterogeneous in the therapy they provide, perhaps prescribing TKIs due to this being “better than doing nothing”. The ERG’s clinical experts noted that TKI administration might be based on anecdotal evidence (e.g. a case report of regorafenib benefit in a single patient).

The above explanations, whilst speculative, are indicative of where there are uncertainties in clinical practice. The CS does not explicitly state whether any prior TKIs administered before patients enrolled in the NAVIGATOR study had been employed in the adjuvant setting (i.e. prior to unresectable or metastatic disease diagnosis). However, we believe this to be unlikely since: (i) There is a consensus that *PDGFRA* D842V-mutated GISTs should not be treated with any adjuvant therapy[6] and our clinical expert advisors concurred. (ii) The company had excluded adjuvant TKI use in their analysis of BLU-285-1002 and presumably would have done the same for NAVIGATOR had any of the enrolled patients been known to have received prior TKIs in an adjuvant setting.

ERG conclusion

One single-arm prospective study on avapritinib and two retrospective chart review/survey studies on comparator TKIs (i.e. ECM) are relevant to this appraisal. The populations available for analysis have relatively small sample sizes (N=19 to N=58). The participants in the avapritinib study received more frequent prior TKI use than would be expected in UK clinical practice, despite TKIs being ineffective in the *PDGFRA* D842V subgroup. The TKI use in the ECM studies is unclear as it was not reported for the relevant subgroup of patients who had unresectable/metastatic GIST and the *PDGFRA* D842V mutation. The rationale for why patients with the *PDGFRA* D842V mutation group in these studies received TKIs is not discussed.

3.2.2 Risk of bias assessment

The company assessed the avapritinib and comparator studies using the Downs and Black checklist for non-randomised studies.[16] This checklist contains 27 questions which assess four aspects (domains) of study quality: reporting, internal validity (bias and confounding), power, and external validity. The checklist has been validated by its authors for internal consistency, test-retest and inter-rater reliability, and criterion validity [16] and evaluated independently.[32] We are unclear how frequently the Downs and Black checklist has been used for evaluating non-randomised studies in NICE Technology Appraisals.

The company report results of the assessment as yes/no answers to each question in CS Appendix Table 19. One question about power has not been answered (question 27) (see section 3.2.4 for discussion of statistical power). As noted by Deeks et al. in a review of quality assessment tools,[32] most of the questions in the Downs and Black checklist relate to reporting rather than validity. We have therefore focused on those questions concerning internal validity (bias and confounding) (questions 14 to 26) for the present appraisal.

A comparison of the company's and ERG's assessments of the NAVIGATOR, BLU-285-1002 and Cassier studies for questions 14 to 26 of the Downs and Black checklist is shown in Appendix 2. However, we encountered several problems whilst applying the checklist to the included studies, as explained in Appendix 2.

The key issues identified from the ERG's validity assessment, which apply to all the measured outcomes, are:

- The studies were all single-arm studies which may be at risk of bias (selection bias, performance bias, and/or confounding) since factors other than the intended intervention might explain the outcome (such factors can be controlled for in comparative studies but not in single-arm studies).
- The comparator studies were retrospective chart reviews which carries an additional risk of selection bias arising through the possibility of selective ascertainment (i.e. "cherry picking") of cases and/or results.
- The studies had relatively small sample sizes. Whilst small sample sizes may not necessarily introduce bias (i.e. systematic error) they would increase uncertainty in estimates of effects through lack of precision.

An appropriate way to reduce the risk of selection bias in the evidence synthesis would be to ensure that active treatment and comparator groups of the single-arm studies are as well-matched as possible on participant characteristics when conducting an indirect treatment comparison (ITC). We assessed the risks of bias in the company's approach to their ITC, as described in section 3.4.5 below.

ERG conclusion

The included studies are inherently at risk of bias due to their single-arm designs and, in the case of the comparator studies, their retrospective designs.

3.2.3 Outcomes assessment

The CS provides information for the outcomes of the NAVIGATOR study across CS Tables 4, 5, 6 and 8. All outcomes specified in the scope and decision problem, except for HRQoL, are reported.

Appendix 4 of this report provides further description of the primary, secondary and exploratory outcomes of the NAVIGATOR study that are reported in the CS.

The CS uses outcomes commonly reported in cancer drug appraisals: overall response rate (ORR), overall survival (OS), progression-free survival (PFS), duration of response (DoR), disease control rate (DCR) and clinical benefit rate (CBR). Additional time to event endpoints that have been used are Time to Response and Time on Treatment (ToT) (for definitions see Appendix 4). PFS can be used as a surrogate for OS, yet neither PFS nor OS data are mature in the NAVIGATOR study. Therefore, ORR is appropriate as the primary outcome (supported by DoR as a secondary outcome), in the NAVIGATOR study. ORR is useful for clinical effectiveness assessment in single-arm trials where there is no available therapy, requires a smaller population, and can be assessed earlier than overall survival data.[33,34]

Table 6 in the CS reports that the outcomes are based on tumour status assessed centrally, with measurements for ORR, DoR, PFS, and CBR based on the Modified Response Evaluation Criteria in Solid Tumors (mRECIST) version 1.1 which is a standard for measuring treatment response based on tumour shrinkage. According to the CSR, in order to minimise bias, assessment of the primary outcome (i.e. ORR) was carried out by two independent reviewers concurrently who were blinded to the results of the other reviewer, when adjudication was performed the third reviewer was blinded to the identities of the first two reviewers but not to their analyses.

The following outcomes inform the economic model. These are based on the most recent January 2020 data cut, except for adverse effects of treatment which are based on the November 2018 data cut:

- Adverse effects of treatment, primary outcome
- ToT, primary outcome
- PFS, secondary outcome
- OS, exploratory outcome.

The remaining outcomes, which do not inform the economic model, are based on data from the November 2018 data cut:

- ORR, primary outcome
- DoR, secondary outcome
- DCR, part of ORR
- CBR, secondary outcome
- Time to treatment, exploratory outcome

The CS does not report an HRQOL outcome. HRQoL is an outcome in the NICE scope and CS Table 1 indicates that it would be addressed in the CS. However, CS section B.3.4.1 states that no HRQoL data were collected in the NAVIGATOR study. Data for HRQoL in the company's economic model are sourced from the published literature.

ERG conclusion

All included outcomes are clinically relevant and match the scope and decision problem, except for HRQoL which was not assessed in the pivotal NAVIGATOR study. Whilst the outcomes used in the economic model are appropriate and all use the latest data cut, the survival data remain immature.

3.2.4 Approach to study statistics

The statistical approaches for each outcome, except for ToT, are defined in CS Table 8.

The statistical analysis plan (SAP) for NAVIGATOR was provided in response to clarification question A1. In addition to the information in CS Table 8 the SAP states that descriptive statistics will be provided for ToT.

Data are immature. Median OS and DoR have not been reached, and so results for OS, PFS and ORR should be treated with caution. CS section B.2.11 reports that follow up is ongoing for survival. No details of any further potential data cuts are provided.

The ERG believe that the appropriate statistical methods have been applied for analysing each outcome. OS, PFS, and DoR were analysed using Kaplan-Meier methods, with variance tested using Greenwood's formula, a common Kaplan-Meier estimator.[35] ORR and CBR were estimated using frequency, percentage and 95% confidence intervals based on the exact binomial distribution.

The NAVIGATOR CSR[17] states that no formal adjustments for possible covariate effects were planned. However, CS Table 8 describes adjustments for ORR (the CS presents a list of covariates used to fit a logistic regression), PFS (used estimated hazard ratios of confounding factors), and OS (stratified Cox regression analysis using mutation type as a stratification factor; CSR page 84).

3.2.4.1 Sample size and power calculation

CS Table 8 reports that a sample size of 31 patients would be required for 90% power to test the null hypothesis of ORR ≤ 10% versus the alternative hypothesis of ORR ≥ 35% using an exact binomial test, and assuming a two-sided Type 1 error rate of 0.05. As the sample of patients with the *PDGFRA* D842V mutation in the study 56, the ERG is satisfied that the study is adequately powered for this particular hypothesis test.

3.2.4.2 Analysis populations

The clinical effectiveness analysis population of NAVIGATOR is defined as patients with the *PDGFRA* D842V mutation (N=56). This is a pre-specified subgroup (Group 2) of the safety population of the NAVIGATOR study (N=237).

The safety analysis population of NAVIGATOR includes all patients in the study with unresectable or metastatic GIST with any mutation, not limited to *PDGFRA* D842V (N=237).

The NAVIGATOR SAP states that all primary analyses will be conducted and presented by starting daily dose (grouped as <300mg, 300mg, 400mg, 300/400mg and 'all doses'). The 'all doses' group is the company's preferred analysis population, as reported for the clinical effectiveness results in CS section B.2.6. The ERG and our clinical experts agree that dose pooling is appropriate (see section 3.2.1).

3.2.4.3 Subgroup analyses

According to the NAVIGATOR SAP, comparisons of the different avapritinib dose groups were pre-specified, albeit descriptively without formal statistical testing being mentioned. CS Appendix L presents descriptive comparisons between the 300mg QD, 400mg QD and all-doses groups for overall survival (CS Appendix Table 51; CS Appendix Figure 12), progression-free survival (CS Appendix Table 52; CS Appendix Figure 13), overall response rate (CS Appendix Table 53), duration of response (CS Appendix Table 54), and time to response (CS Appendix Table 55). Results for the <300mg group are not reported in the CS but can be found in the CSR.[17]

Comparisons of adverse event frequencies were also made between these dose groups (CS Appendix F), as discussed in section 3.2.6 below.

3.2.4.4 Missing data

The CS does not explicitly discuss missing data in the NAVIGATOR study. The CSR states that, in general, no imputation was performed for missing data points (CSR section 11.3.2). However, the CS reports sample sizes alongside the clinical effectiveness outcomes which suggest that for most outcomes all available study participants were included in analyses.

Sensitivity analyses were carried out for DoR and PFS (CS Table 8). For DoR, FDA[37] censoring rules were used in the primary analysis and a sensitivity analysis was carried out using EMA[38] censoring rules. Detailed FDA[37] and EMA[38] censoring rules for PFS and DoR are reported in Table 4 of the NAVIGATOR SAP. The company clarified at the factual error check stage that PFS censoring followed EMA rules. The CS states that if a patient had not had an event, PFS was censored at the date of last valid assessment that was stable or better (CS Table 8). The CS reports DoR using the results of the sensitivity analysis (the EMA rules); however, only one less patient was censored by these rules and the Kaplan-Meier estimates remained the same (CSR Tables 14.2.2.1.2 and 14.2.2.2.2).

ERG conclusion

The ERG are satisfied that the company's approach to statistics is generally appropriate: the study was adequately powered and the latest available data were used to inform the survival statistics.

3.2.5 Clinical effectiveness results

Clinical effectiveness results are reported for NAVIGATOR in CS section B.2.6 and CS Appendix L; for BLU-285-1002 in CS Tables 9 and 11, CS Appendix Tables 14, 16 and 17, and CS Appendix N.4; and for the Cassier study in CS Tables 15 to 17 and CS Appendix P.

Below we present a summary of results from NAVIGATOR for the primary outcome of ORR, related response outcomes (including DCR, CBR and DoR) and time-to-event outcomes used in the economic model (OS, PFS, time to response, time on treatment), alongside those from the BLU-285-1002 and Cassier studies where available. We note that:

- Radiographic tumour reductions are reported in the CS but are not in the decision problem nor used in the economic model and are therefore not commented upon here.

- Results appear consistent for each outcome across the dose subgroups, although the sample sizes are relatively small (CS Appendix L).
- Data for time-to-event outcomes (OS, PFS, DoR) are immature and therefore have increased uncertainty relative to a mature data set.

3.2.5.1 Overall Response rate (ORR)

The ORR and related response outcomes are shown in Table 6. No patients in the ECM studies achieved a response, compared to [redacted] % in the avapritinib study.

Table 6 Overall response rate in the avapritinib and ECM studies

	Avapritinib	ECM		
	NAVIGATOR All doses group (N=56)	BLU-285-1002 Unresectable/ metastatic group (N=19)		Cassier et al 2012 *PDGFRA* subgroup (N=32)
ORR outcome		2nd line n=19	3rd line n=16	
ORR, n (%) [95% CI]	[redacted] 300mg dose [redacted]	[redacted]	[redacted]	0 (0)
Complete response	[redacted]	[redacted]	[redacted]	0 (0)
Partial response	[redacted]	[redacted]	[redacted]	0 (0)
Stable disease	[redacted]	[redacted]	[redacted]	10 (31)
Progressive disease	[redacted]	[redacted]	[redacted]	21 (66)
Other	[redacted]	[redacted]	[redacted]	1 (3) [a]
CBR, n (%) [95% CI]	[redacted]	Not reported		Not reported
DCR, n (%) [95% CI]	[redacted]	Not reported		Not reported

Source: NAVIGATOR: CS section B.2.6.3; BLU-285-1002: CSR; Cassier: study publication.
CBR: clinical benefit rate; DCR: disease control rate (see Appendix 4 for definitions)
[a] Includes not evaluable and not assessed (in Cassier study 1 patient died before first assessment)

3.2.5.2 Duration of Response (DoR)

The available data for duration of response in NAVIGATOR are shown in Table 7.

Table 7 Duration of response in the NAVIGATOR study

DoR, Kaplan-Meier estimates	
Median (months) (95% CI)	All doses: [redacted]
	300mg dose [redacted]
3 months, % (95% CI)	[redacted]
6 months, % (95% CI)	[redacted]
9 months, % (95% CI)	[redacted]
12 months, % (95% CI)	[redacted]
18 months, % (95% CI)	[redacted]
24 months, % (95% CI)	[redacted]
Source: CS section B.2.6.4	

3.2.5.3 Overall Survival (OS)

Median estimates of overall survival in NAVIGATOR are shown in Table 8 (from CS Table 9). The NAVIGATOR OS Kaplan-Meier curve is provided in CS Figure 3 (not reproduced here). At 42 months [redacted]% of patients were still alive, and that had not changed from the 30-month time point. Median OS is [redacted] at the latest (January 2020) data cut. In contrast, median OS was [redacted] months and [redacted] months for second- and third-line therapy respectively in BLU-285-1002, and 14.7 months in the Cassier study.

Table 8 Overall survival in the avapritinib and ECM studies

	Avapritinib	**ECM**	
OS, Kaplan-MOSeier estimates	**NAVIGATOR All doses group (N=56)**	**BLU-285-1002 Unresectable/ metastatic group (N=19)**	**Cassier et al 2012 *PDGFRA* subgroup (N=32)**
Median follow-up, months	[redacted]	Not reported	45.3 [a]
Median OS, months (95% CI)	[redacted]	[redacted]	14.7 (not reported)
6 months, %	[redacted]	Not reported	Not reported
12 months, %	[redacted]	Not reported	Not reported
18 months, %	[redacted]	Not reported	Not reported
24 months, %	[redacted]	Not reported	Not reported
30 months, %	[redacted]	Not reported	Not reported
36 months, %	[redacted]	Not reported	Not reported
42 months, %	[redacted]	Not reported	Not reported

Source: NAVIGATOR: CS section 2.6.1; BLU-285-1002: CS Appendix Table 17; Cassier: study publication.
[a] median follow-up for surviving patients

Although no subgroup analyses were planned for UK NAVIGATOR patients within the *PDGFRA* D842V mutation subgroup, the CS notes that [redacted] of the [redacted]) UK patients in the study were still alive at the time of the January 2020 data cut, with a median follow-up of [redacted] months. This is relevant to the appraisal and the results are in line with the rest of the study population.

3.2.5.4 Progression Free Survival (PFS)

Median estimates of progression-free survival are shown in Table 9 (from CS Table 10). The NAVIGATOR PFS Kaplan-Meier curve is provided in CS Figure 4 (not reproduced here). Median PFS was [redacted] at the latest (January 2020) data cut compared with only [redacted] to [redacted]months in the ECM studies. In NAVIGATOR [redacted]% of patients were alive and progression free at 42 months. We note that the reported duration of PFS was longer for patients receiving third- line than for those receiving second-line therapy in BLU-285-1002, although these estimates are uncertain (sample sizes are small and confidence intervals wide).

Table 9 Progression-free survival in the avapritinib and ECM studies

	Avapritinib	**ECM**	
PFS, Kaplan-Meier estimates	**NAVIGATOR All doses group (N=56)**	**BLU-285-1002 Unresectable/ metastatic group (N=19)**	**Cassier et al 2012 *PDGFRA* subgroup (N=32)**
Median, months (95% CI)	[redacted]	[redacted]	2.8 (2.4 to 3.2)
6 months, %	[redacted]	Not reported	8 patients (25%) had PFS longer than 6 months (range 6.4 to 50.8 months)
12 months, %	[redacted]	Not reported	
18 months, %	[redacted]	Not reported	
24 months, %	[redacted]	Not reported	
30 months, %	[redacted]	Not reported	
36 months, %	[redacted]	Not reported	
42 months, %	[redacted]	Not reported	
Source: NAVIGATOR: CS section B.2.6.2; BLU-285-1002: CS Appendix Table 17; Cassier: study publication.			

3.2.5.5 Time to Response

The median time to response was [redacted] days in the all-doses group of the NAVIGATOR study. Time to response was not reported for the ECM comparator studies.

3.2.5.6 Time on treatment (ToT)

Median estimates of time on treatment are shown in Table 10 for the NAVIGATOR study, for both the *PDGFRA* D842V population and the safety population. Time on treatment was not reported in the BLU-285-1002 and Cassier studies.

Table 10 Time on treatment in the NAVIGATOR study

ToT, Kaplan-Meier estimates	Analysis population		
	***PDGFRA* D842V group (N=56) January 2020 data cut**	***PDGFRA* D842V group (N=56) November 2018 data cut**	**Safety population (N=237) November 2018 data cut**
Median, months (95% CI)	[redacted]	[redacted]	[redacted]
6 months, %	[redacted]	Not reported	Not reported
12 months, %	[redacted]	Not reported	Not reported
18 months, %	[redacted]	Not reported	Not reported
24 months, %	[redacted]	Not reported	Not reported
30 months, %	[redacted]	Not reported	Not reported
36 months, %	[redacted]	Not reported	Not reported
42 months, %	[redacted]	Not reported	Not reported

Source: CS Tables 21 to 23.
[a] reported in the CS in weeks, converted to months by ERG

3.2.5.7 HRQoL outcomes

CS section B.3.4.1 states that no HRQoL data were collected in the NAVIGATOR study, and CS section B.2.13.2 discusses the lack of HRQoL data for the specific *PDGFRA* D842V mutation population of patients with unresectable of metastatic GIST as a limitation of the evidence base overall. Sources of HRQoL data for the company's economic analysis were taken from the published literature (see section 4.2.7 below).

3.3 Safety outcomes

The company's economic model includes Grade 3-4 adverse events with incidence of greater than 2% in either arm. The avapritinib arm also includes any comparator adverse events with greater than 2% incidence (CS section B.3.3.5) (see section 4.2.7.4 below).

3.3.1 Current submission

Adverse events are reported in CS section B.2.10 for the entire safety population of the NAVIGATOR study (N=237), not limited by the type of mutation or starting dose of avapritinib (therefore including patients without the *PDGFRA* D842V mutation, treated at fourth line, following failure of imatinib, sunitinib and regorafenib). The company argue, and the ERG's clinical experts agreed, that the full NAVIGATOR population provides the maximum amount of data available on the safety of avapritinib and is appropriate as a reference since there is no evidence to suggest that presence or absence of the *PDGFRA* D842V mutation would affect the frequency of adverse events.

The company report adverse events separately for the 300mg and 400mg QD dose subgroups of the entire NAVIGATOR study population (N= and N= respectively) and the combined 300mg+400mg group (N=) in CS Appendix F. Frequencies of adverse events are similar when the combined 300mg+400mg QD subgroup (N=) is compared against the all-doses group (N=237) (CS Appendix Table 23).

Overall, % of all patients in NAVIGATOR who received avapritinib (N=237) had experienced at least one adverse event at the November 2018 data cut (median duration of treatment weeks). The most frequent adverse events were nausea (%), fatigue (%), and anaemia (%), with a wide range of other adverse events occurring at lower frequencies (CS Table 25). Adverse events of Grade 3 or above occurred in % of patients, whilst serious adverse events occurred in %. Adverse events leading to avapritinib discontinuation, dose interruption or dose reduction occurred, respectively, in %; % and % of patients. Overall, % of patients died within 30 days of receiving their last dose of avapritinib (CS Appendix Table 22), mostly as a result of progressive disease.

The overall high frequency of any adverse events is consistent with those in the comparator TKIs (imatinib 98%, sunitinib 94%, regorafenib 100%) (CS Table 29). However, the frequency of adverse events of Grade 3 or more was higher in the NAVIGATOR all-doses

population ([redacted]%) than among patients receiving imatinib (52.4%). A caveat is that it is difficult to directly compare rates of different types of adverse events across the TKIs due to differences in how the events were defined and reported.

The company identified cognitive effects and intracranial bleeding as adverse events of special interest among patients receiving avapritinib. Cognitive effects were experienced by [redacted]% of patients overall, with the most frequent being memory impairment ([redacted]%), cognitive disorder ([redacted]%), confusional state ([redacted]%) and encephalopathy ([redacted]%). Intracranial bleeding occurred in three patients ([redacted]%) (CS Table 27).

The majority of cognitive adverse events were Grade 1 or Grade 2. The only cognitive adverse event of Grade 3 or more reported with an incidence of ≥2% (as measured in the 300mg/400mg dose group) was confusional state ([redacted]%).

The company conducted a post-hoc descriptive analysis of cognitive effects to clarify the safety and tolerability of the 300mg QD avapritinib dose in relation to these adverse events (CS section B.2.10.7.1). This included the population of patients who received the 300mg QD dose of avapritinib in the NAVIGATOR study and the ongoing VOYAGER trial (total N=184) (VOYAGER is described in section 3.2.1 above). The analysis demonstrated similar frequencies of cognitive effects to those seen in the all-dose group in NAVIGATOR (CS Table 28). Cognitive effects led to dose interruption in [redacted]% of patients and dose reduction in [redacted]%. The post-hoc safety analyses reports that Grade 3 or above adverse events for cognitive impairment were: confusional state (1%), cognitive disorder (<1%), encephalopathy (<1%), and memory impairment (0%). No patients in the post-hoc safety analysis experienced adverse events for any cognitive effects of Grade 4 or above.

3.3.2 FDA safety assessment

The US Food and Drug Administration (FDA) have conducted a detailed assessment of the safety of avapritinib based on data submitted by the company.[30] The primary safety population in the FDA assessment was defined as all patients in NAVIGATOR who received avapritinib doses of 300mg or 400mg QD (N=[redacted]). Additional safety data were examined from the phase 3 VOYAGER trial of patients with advanced GIST (BLU-285-1303) and a dose-finding study of avapritinib use in patients with advanced systemic mastocytosis (EXPLORER; BLU-285-2101). As would be expected, the FDA and ERG reached similar conclusions on the safety of avapritinib. The key conclusions from the FDA assessment[30] were:

- The size of the safety database is adequate to provide a reasonable estimate of adverse reactions that may be observed with avapritinib, and the duration of treatment is adequate to allow assessment of adverse reactions over time.
- The proposed 300mg QD dosage has a manageable safety profile.
- The 300mg QD dose appears to be better tolerated than 400mg QD; specifically, a higher incidence of Grade 3+ adverse events (82% versus 67%), adverse events leading to dose reduction (66% versus 41%) and cognitive adverse events of special interest (48% versus 35%) occurred in the 400mg QD starting dose group compared to 300mg QD.
- The frequency of some treatment-emergent adverse events varied with age; however, due to the single-arm study design it is not possible to conclusively say whether the differences were due to age alone.
- The frequency of some treatment-emergent adverse events varied with race; however, due to dominance of the data by people with white race, it is unclear whether adverse event frequencies do differ consistently between racial groups.
- Intracranial bleeding is a rare but significant adverse event likely related to avapritinib.
- Central nervous system (CNS) effects occurred in [redacted]% of patients, of which [redacted]% were Grade 3 or Grade 4. Avapritinib was permanently discontinued due to CNS effects in [redacted]% of patients. Cognitive impairment was more frequent in patients aged over 65 years.
- The pharmacokinetics of avapritinib in people with severe hepatic impairment requires investigation.

ERG conclusion

Avapritinib has a manageable safety profile which has some broad similarities with the safety profiles of the comparator TKIs, although comparisons are difficult due to differences in how adverse events are defined and recorded. Avapritinib is uniquely associated with cognitive effects which, in clinical studies, required dose interruption and/or reduction in [redacted]% of patients.

3.4 Critique of studies included in the indirect treatment comparison

3.4.1 Rationale for the ITC

The comparison of interest is avapritinib versus established clinical management (ECM), where ECM reflects the use of TKI therapy and/or best supportive care.

As noted above in section 3.2.1, a single-arm study on avapritinib and two single-arm studies on various ECM comparators were available, but no studies have directly compared avapritinib against ECM. An indirect treatment comparison was therefore necessary.

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

As noted in section 3.2.1 above, the company identified seven studies that provide data on patients with the *PDGFRA* D428V mutation who received avapritinib or ECM. Four of these studies had very small sample sizes (N=3 to N=12) and were not considered in detail by the company. The remaining three studies on avapritinib (NAVIGATOR, N=56) and on ECM (BLU-285-1002, N=22 and the Cassier study, N=32) were selected for inclusion in indirect comparisons. The ERG agree with the company that these studies are the most appropriate for the ITC.

3.4.3 Clinical heterogeneity assessment

Table 4 and Table 5 in section 3.2.1 above compare the study and patient characteristics, respectively, across the three studies. Heterogeneity among the studies is evident as follows:

- There are differences between NAVIGATOR and BLU-285-1002 in terms of age, race, geographical region, tumour size, and prior therapies.
- Prior TKI therapy received is not reported in the ECM studies for the relevant *PDGFRA* D842V mutation subgroup (Table 5 above).
- ECOG performance status was recorded as missing for all but one of the BLU-285-1002 patients.
- Patient-level baseline characteristics in the Cassier study were unavailable for the *PDGFRA* D842V mutation subgroup (32/58=55%), precluding a qualitative assessment of heterogeneity.

An important aspect of heterogeneity assessment is to establish whether the studies differ on key prognostic factors. This is discussed in section 3.5.2 below.

3.4.4 Similarity of treatment effects

- The company's ITC approach covers OS and PFS outcomes which appear comparable across the studies.
- The proportions of patients receiving each line of treatment differed between the ECM comparator studies (CS Appendix Figures 30 and 31).

- The Cassier study second-line cohort appears to fare notably worse on PFS than the BLU-285-1002 third-line cohort (CS Appendix Figures 30 & 31). As such, the base case adjusted ITC (NAVIGATOR versus BLU-285-1002) could be viewed as more conservative than the unadjusted scenario ITC (NAVIGATOR versus the Cassier study).

3.4.5 Risk of bias assessment for RCTs included in the ITC

- As noted above in section 3.2.2, the included studies are inherently at risk of selection bias due to their single-arm designs and, in the case of the comparator studies, also their retrospective methods.
- The bias risk arising from the lack of a comparator group in each study may be reduced if studies can be well-matched on all key prognostic factors and effect modifiers in an ITC (although "perfect" matching is considered very difficult if not impossible to achieve).
- The inherent bias risk arising from the comparator studies being retrospective (i.e. possibility of selective "cherry picking" of cases or results) cannot be reduced using ITC methods.
- As summarised in Table 11, both the adjusted and unadjusted indirect comparisons are at risk of bias, with the unadjusted comparison with the Cassier study being particularly at high risk of bias due to the lack of any matching of covariates. These comparisons are illustrative, since the risk of bias cannot be quantified.

Table 11 Overview of bias risk for studies included in the ITC

Bias source	NAVIGATOR versus BLU-285-1002	NAVIGATOR versus Cassier
Inherent risk of bias due to single-arm design [a]	Yes in both studies but possibly reduced by matching in ITC	Yes in both studies but cannot be reduced (no matching)
Inherent risk of bias due to retrospective methods [b]	Yes in BLU-285-1002 – cannot be reduced by ITC methods	Yes in Cassier study – cannot be reduced by ITC methods
[a] this covers several domains of bias e.g. selection bias, performance bias and confounding which single-arm studies are prone to [b] bias due to selective ascertainment of cases and/or results		

ERG conclusion

The company employed ITC as the method of data synthesis, which is appropriate given the lack of any comparative studies. One avapritinib study and two ECM

studies are eligible for comparison. All three studies have inherent risks of bias arising from their single-arm designs and, in the case of the two ECM studies, also due to their retrospective ascertainment of patient records. The studies exhibit heterogeneity in several baseline characteristics. Some baseline characteristics of the ECM studies, including prior TKI use, are not fully clear due to not being reported specifically for the PDGFRA D842V mutation subgroup of patients with unresectable or metastatic GIST.

3.5 Critique of the ITC methods

3.5.1 Overview of the company's ITC approach

The CS reports that the following ITCs were conducted:

- An adjusted ITC using propensity score weighting was conducted to compare NAVIGATOR and BLU-285-1002. This is an appropriate methodology for this comparison since the company had access to individual patient-level data (IPD) for both studies.
- An unadjusted (naïve) ITC was conducted to compare NAVIGATOR and the Cassier study (CS Appendix Tables 14 to 17). The CS states that an adjusted comparison was not feasible since the Cassier study publication reported inadequate information on baseline characteristics of the patients in the *PDGFRA* D842V subgroup to enable statistical matching of the studies (as noted above in Table 5, some baseline characteristics are missing and those that are reported are for the whole study group, not specifically for the *PDGFRA* D842V subgroup).

The company favoured the adjusted comparison as their primary comparison and the unadjusted comparison as a sensitivity analysis. We agree that the company's approach is appropriate, given the heterogeneity among studies noted above in section **Error! Reference source not found.**, since an adjusted ITC is preferable for reducing imbalances in prognostic factors and effect modifiers and, hence, for minimising the risk of bias arising from the comparison.

We also agree with the company that an adjusted comparison of the NAVIGATOR and Cassier studies would not be feasible and therefore only a naïve comparison could be made. A possible advantage of including the Cassier study in a naïve comparison would be that the cohort was Europe-based whilst the BLU-285-1002 study consisted solely of US patients, and therefore the company's combination of adjusted and naïve ITCs in their primary comparison and sensitivity analyses respectively would make best use of the available

comparator data. However, the following limitations of the naïve comparison should be kept in mind:

- Relatively few patient characteristics can be compared between NAVIGATOR and the Cassier study (Table 5);
- Some of those characteristics that can be compared are heterogeneous across the studies (Table 5);
- The Cassier study does not report any baseline characteristics for the *PDGFRA* D842V subgroup (Table 5). Clinical experts advising the ERG suggested that patient characteristics would be unlikely to differ between the overall unresectable/metastatic population and those with the PDGFRA D842V mutation in the Cassier study. However, the ERG is uncertain whether prior TKI use would have been homogeneous, given that TKI clinical effectiveness differs between these population groups.
- The Cassier study (as with BLU-285-1002) is at risk of selection bias due retrospective data collection.

3.5.2 Data inputs to the adjusted ITC

In CS Appendix D, page 26, the company note that "some factors potentially associated with treatment outcomes were identified". The ERG, concerned that the list of prognostic factors may be incomplete, requested clarification of the evidence in support of prognostic factors (clarification questions A2 and A3). The company responded that no established evidence for prognostic factors exists given the small numbers of patients with the *PDGFRA* D842V mutation, although they suspect that ECOG performance status could be related to outcomes as is often the case in oncological indications. The company reported they thus used a comprehensive approach that included all available potentially relevant prognostic factors in the ITC (CS Appendix D Table 8). The ERG's experts agreed the list to be comprehensive.

However, the propensity score weighting did not include tumour size nor the specific previous TKIs. No explanation is given in the CS for why these covariates have not been included in the adjusted ITC analysis. The company clarified (at the factual inaccuracy check stage) that tumour size was only measured at the time of diagnosis, not at the time of initial treatment for unresectable or metastatic disease; and that inclusion of the specific prior TKIs received was not feasible due to the small sample size. It is unclear whether these covariates could be prognostic or what the effect of including/excluding them from the propensity score weighting exercise would be. Nevertheless, despite this weakness, the

ERG agree that adjusting for some of the prognostic factors is preferable to a naïve indirect comparison adjusting for none.

The continuous variables age and disease duration were dichotomised in the propensity score weighting (see Table 5). It was unclear to the ERG why this was done (clarification question A4). In response, the company conducted a series of sensitivity analysis using different cut-offs and continuous variables; none of these analyses improved model fit.

The company excluded race and ECOG performance status from the propensity score weighting analysis due to missing values. In response to clarification question A7, the company conducted a sensitivity analysis including race in the analysis which had little impact on OS or PFS. Furthermore, the ERG's experts were unaware of any prognostic effect of race.

3.5.3 Statistical methods for the adjusted ITC

For the adjusted ITC the company used an inverse probability weighting (IPW) method. Using IPW, outcomes are weighted by the inverse of the propensity score which is the probability of a patient with a given covariate set being assigned to a treatment. Avapritinib (i.e. the NAVIGATOR cohort) was selected as the reference treatment so patients' propensity score weights were estimated as the probability that a patient belongs to the ECM cohort (i.e. BLU-285-1002). Thus, ECM patients with a higher propensity score, and hence a lower IPW, had a lower probability of belonging to the NAVIGATOR cohort, and vice versa. The IPW method was preferred by the company over an alternative possible method, propensity score matching, as setting a matching threshold (caliper) could have led to the exclusion some patients in an already small dataset (clarification response A10). A logit regression including all prognostic factors was used to estimate the propensity score.

The company presented Kaplan Meier curves and median OS and PFS estimates for each indirect treatment comparison but did not report relative treatment effects in terms of hazard ratios (HR) (presumably because these were not required for the economic model). The ERG requested the company to report HRs for the comparisons of NAVIGATOR against BLU-285-1002 (IPW adjusted ITC) and NAVIGATOR against the Cassier study (unadjusted ITC) in clarification question A6. These results are provided in section 3.5.5 below.

CS Appendix N.3 shows that two patients from BLU-285-1002 had relatively high inverse propensity score weights (and therefore low propensity scores), and therefore may have had

a disproportionate impact on the analysis. The ERG asked the company to repeat the analysis removing these two patients in clarification question A9. The company ran this analysis but qualified this by stating that they did not consider this a valid analysis. The ERG agrees since this removes the two patients most resembling NAVIGATOR from BLU-285-1002. Nevertheless, it does show the analysis is sensitive to the inclusion of these two patients and illustrates uncertainty due to the small sample size (i.e. the analysis would be less sensitive to the inclusion of these two patients if the sample size was larger).

The ERG also queried whether the IPW exercise had been wholly successful (clarification question A10), since differences in the mean values of certain patient characteristics between NAVIGATOR and BLU-285-1002 were greater post-IPW than pre-IPW (CS Appendix N.2, Table 58). The largest differences in pre-IPW means were for age, ethnicity, and total number of TKIs. Although there is still some misalignment of these characteristics (including total number of TKIs) post-IPW, the company argued that the overall effect is of an improvement in terms of balanced patient characteristics post-IPW. Further scenario analyses around the inclusion of covariates provided in clarification response A10 support the company's conclusion. Based on this, the ERG accept the company's argument.

NICE DSU Technical Support Document 17 recommends that sensitivity analyses are conducted using different matching methods and using different covariate sets and functional form for continuous covariates (e.g. polynomials, interactions) to test the stability of the results. No such analyses are presented in the CS, hence this was queried by the ERG in clarification question A11. The company subsequently presented a series of scenario analyses using a backwards stepwise selection process and a probit model to calculate the propensity score. Backward selection is an automated regression process which starts with the model including all covariates then sequentially removes the least predictive covariate until all remaining covariates are statistically significant. (The level of statistical significance used by the company is not reported.) The most parsimonious stepwise model included age and number of TKIs (clarification response Table 7). The ERG acknowledge that visually there is little difference in the Kaplan Meier curves between the parsimonious model and full covariable model for OS or PFS and between the logit and probit models (clarification response Figures 8 and 9).

There is some variation in the unadjusted and IPW-adjusted Kaplan Meier curves, resulting in a downward shift of the OS curve for ECM (clarification response Figure 8). PFS was similar in the adjusted and unadjusted analyses (clarification response Figure 9).

It should be noted that the outputs of the ITC are not used directly in the economic model. The OS and PFS inputs to the economic model are taken from the extrapolation of the IPW Kaplan Meier curves.

The analysis was conducted in Stata 13. In response to clarification question A5, the company provided the data and Stata code. The ERG checked the code and confirmed that the models had been correctly applied.

The ERG were able to replicate the CS results for OS and PFS in terms of Kaplan Meier curves and survival at key time points, but we were unable to replicate the hazard ratio for PFS reported in clarification response A3. Fitting a Cox proportional hazards model to the IPW-adjusted Kaplan Meier OS model resulted in a HR of 4.42 (95% CI 2.09, 9.34) which approximates the company's results but the PFS HR of 11.81 (95% CI 4.79, 29.15) is less favourable to ECM than the company's analysis. The reason for the discrepancy is unclear but, in any case, the HRs are not used in the economic model.

3.5.4 Results from the adjusted ITC

3.5.4.1 Overall survival

- The IPW-adjusted Kaplan Meier curves and median survival estimates for NAVIGATOR and BLU-285-1002 are reproduced in Figure 2 and Table 12 below.
- The OS hazard ratio for ECM versus avapritinib (clarification response A6) is [redacted]

[redacted]

Source: Reproduction of CS Figure 6

Figure 2 IPW-adjusted Kaplan–Meier curves for OS in the NAVIGATOR study (avapritinib) and BLU-285-1002 (ECM)

Table 12 IPW-adjusted Kaplan–Meier survival estimates of OS at key timepoints in the NAVIGATOR study (avapritinib) and BLU-285-1002 (ECM)

Kaplan–Meier survival estimates	NAVIGATOR	BLU-285-1002
Median, months	[redacted]	[redacted]
6 months	[redacted]	[redacted]
12 months	[redacted]	[redacted]
18 months	[redacted]	[redacted]
24 months	[redacted]	[redacted]
Source: Reproduced from CS Table 17		

3.5.4.2 Progression-free survival

- The IPW-adjusted Kaplan Meier curves and median PFS estimates for NAVIGATOR and BLU-285-1002 are reproduced in Figure 3 and Table 13 below.
- The PFS hazard ratio for ECM versus avapritinib (clarification response A6) is [redacted]

[redacted]

Source: Reproduction of CS Figure 7

Figure 3 IPW-adjusted Kaplan–Meier curves for PFS in the NAVIGATOR study (avapritinib) and BLU-285-1002 (ECM)

Table 13 IPW-adjusted Kaplan–Meier survival estimates of PFS at key timepoints in the NAVIGATOR study (avapritinib) and BLU-285-1002 (ECM)

Kaplan–Meier PFS estimates	NAVIGATOR	BLU-285-1002
Median, months	[redacted]	[redacted]
6 months	[redacted]	[redacted]
12 months	[redacted]	[redacted]
18 months	[redacted]	[redacted]
24 months	[redacted]	[redacted]
Source: Reproduction of CS Table 19		

3.5.5 Comparison of results from the adjusted and unadjusted ITCs

- Kaplan Meier curves, their extrapolation, and survival estimates over time for the Cassier study are presented in CS Appendix P. Hazard ratios for OS and PFS were provided by the company at the ERG's request. These are reproduced in Table 14 below, alongside those from the adjusted ITC analysis.
- The HRs from the adjusted and naïve analyses are in broad agreement. However, we caution that both analyses are subject to uncertainty, particularly the unadjusted comparison (see section 3.5.1 above). The results are based on relatively small sample sizes, meaning that confidence intervals for the HRs are relatively wide. Furthermore, the ITC results are at risk of bias, as summarised in Table 11 above, which adds further uncertainty that is not captured within the confidence intervals.

Table 14 Hazard ratios for median OS and PFS from the adjusted and naïve indirect comparisons

Outcome	Hazard ratio, ECM versus avapritinib (95% CI)	
	BLU-285-1002 versus NAVIGATOR, adjusted ITC	**Cassier study versus NAVIGATOR, unadjusted (naïve) ITC [a]**
Overall survival	[redacted]	[redacted]
Progression-free survival	[redacted]	[redacted]

Source: Clarification response A6
[a] The ERG assume these results are for the Cassier study PDGFRA D842V subgroup rather than for the whole unresectable/metastatic GIST population but this is not stated in the clarification response

3.5.6 Summary of the ERG's critique of the indirect comparisons

- The methodology followed by the company is appropriate given the data limitations
- The methodology has been described and applied correctly
- A thorough set of sensitivity analyses was conducted by the company
- However, three potentially relevant covariates (performance status score, specific prior TKIs received, and tumour size) could not be included in the model due to limitations of the data. The effect of this is unclear.
- The IPW analysis has been effective but remains uncertain given the choice of prognostic factors and small size of the PGDFRA D842V mutation population
- The outputs of the ITC are not used directly in the economic model. The OS and PFS inputs to the economic model are taken from the extrapolation of the IPW Kaplan Meier curves.

3.6 Conclusions of the clinical effectiveness section

The ERG's critique of the company's synthesis of clinical effectiveness evidence has identified a number of issues, as summarised in Table 15. As indicated in the table, some of these issues cannot easily be resolved unless further clinical effectiveness evidence becomes available, whilst other issues have been resolved or partly resolved.

Table 15 Key clinical effectiveness issues identified by the ERG

Issue	Where discussed	ERG comments
The clinical pathway is unclear and differs between the submitted evidence and expected UK clinical practice	Sections 2.2.3.1 and 3.2.1.3	Unclear whether resolvable by wider clinical consultation to reduce uncertainty around UK clinical practice or company clarification on the rationale for TKI use in the clinical studies. Prior TKI use in PDGFRA D842V patients would be expected to be lower than that seen in the company's studies given that TKIs lack clinical effectiveness in this group.
OS, PFS and DoR outcomes are immature	Sections 3.2.4 and 3.2.5Clinical effectiveness results	Not resolvable until the pivotal NAVIGATOR study is completed (or a more recent data cut provided)
Clinical evidence is based on small sample sizes	Section 3.2.1	Not resolvable without collecting further data – difficult in this small population subgroup
ECM comparator studies were retrospective and hence at risk of selection bias	Sections 3.2.2 and 3.4.5	Not resolvable without conducting further, prospective, protocol-based, studies (or retrospective studies with random sampling and blinding) – difficult in this small population subgroup
There is a lack of head to head comparative evidence.	Sections 3.2.1 and 3.4.1	This is partly resolvable by ITC, albeit with key uncertainties
Unclear whether all prognostic factors were accounted for in adjusted ITC (NAVIGATOR versus BLU-285-1002)	Section 3.5.2	Performance status score, tumour size and specific prior TKIs received could not be included as covariates in the analysis due to data limitations. It is unclear whether these would be influential as prognostic factors. Not resolvable without collecting further data – difficult in this small population subgroup
Adjusted ITC not feasible for NAVIGATOR versus Cassier	Section 3.5.1	The naïve ITC should be considered weaker than the adjusted ITC (increased risk of bias from lack of matching) so should be interpreted with caution. This issue might be resolvable if further data or

comparison and naïve ITC highly uncertain		clarification could be obtained from the Cassier study authors. However, although data from the Cassier study inform an economic scenario analysis, these are taken directly from the study rather than from the ITC (section 4.2.6 below)
Lack of HRQoL data for avapritinib	Section 3.2.5.7	Prospective data collection would be preferred. Resolved for now by using literature based HRQoL estimates for the economic model (section 4.2.7) and some HRQoL data from VOYAGER study in an ERG scenario (section 6.2)

4 COST EFFECTIVENESS

4.1 ERG comment on the company's review of cost-effectiveness evidence

The company conducted a systematic literature review of cost-effectiveness studies published from January 2009 until December 2019 for patients with unresectable or metastatic *PDGFRA* D842V-mutated GIST (CS section B.3.1 and CS Appendix G).

The following electronic databases were searched: MEDLINE® In-Process, Embase® and MEDLINE, EconLit®, Centre for Reviews and Dissemination York for Health Technology Assessment Database and National Health Service Economic Evaluation Database. In addition, the company searched conferences to identify relevant abstracts and key international HTA databases to identify relevant HTA evaluations.

The company applied inclusion and exclusion criteria to select relevant economic evaluation studies, which are listed in CS Appendix Table 30. The company's review did not identify any relevant cost-effectiveness studies assessing patients with unresectable or metastatic *PDGFRA* D842V-mutated GIST. Therefore, studies for the general unresectable or metastatic GIST population were considered for inclusion, based on the assumption that quality of life and resource use are similar among patients, regardless of their mutational status. Twenty-one publications for the general GIST population were identified (CS Appendix Figure 5 and CS Table 31). Of these studies, three were previous NICE Technology Appraisals for avapritinib comparators: TA86/TA209 for imatinib,[11,39] TA179 for sunitinib[12] and TA488 for regorafenib.[13] The remaining 18 studies were from international healthcare settings and/or reported the same data used in previous NICE appraisals. The main characteristics of the three previous NICE appraisals are summarised in CS Table 31.

The ERG updated the company's search and one additional study met the inclusion criteria.[40] This study assessed the clinical effectiveness, safety and cost-effectiveness of different sunitinib doses in unresectable or metastatic GIST and different axitinib doses in metastatic renal cell carcinoma. The resource use and cost expenditures were obtained from a Dutch perspective and the authors did not report health state utility values.

ERG conclusion

The ERG consider the company's review of cost-effectiveness evidence adequate and comprehensive, albeit a few months out of date. The company's review did not identify any relevant cost-effectiveness studies assessing patients with unresectable or

metastatic *PDGFRA* D842V-mutated GIST. The additional study found by the ERG[40] does not present any further relevant information for the current appraisal.

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

4.2.1 NICE reference case checklist

Table 16 shows that the company's economic evaluation adheres to the NICE reference case requirements.

Table 16 NICE reference case checklist

Element of health technology assessment	Reference case	ERG comment on company's submission
Perspective on outcomes	All direct health effects, whether for patients or, when relevant, carers	Yes
Perspective on costs	NHS and PSS	Yes
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
Synthesis of evidence on health effects	Based on systematic review	Yes, discussed in CS Appendix D and Appendix H
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.	Yes, EQ-5D data collected from previous NICE appraisals
Source of data for measurement of health-related quality of life	Reported directly by patients and/or carers	Yes, utility values collected from patients in TA179 and TA488. Unclear if utility values were collected from patients in TA86/209

Source of preference data for valuation of changes in health-related quality of life	Representative sample of the UK population	Yes
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Yes
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
PSS, personal social services; QALYs, quality-adjusted life years; EQ-5D, EuroQoL-5D instrument for measuring utilities		

4.2.2 Model structure

4.2.2.1 Overview of the model structure

The company constructed a cohort partitioned survival cost effectiveness model, described in CS section B.3.2.6 and illustrated in CS Figure 12, reproduced in Figure 4 below. There are five health states for patients treated with first line, second line, and third line therapies and also progressed disease (PD) and death. The model has monthly cycles and a lifetime horizon (40 years).

A cohort of patients with unresectable or metastatic *PDGFRA* D842V-mutated GIST enters the model in either the avapritinib or first line treatment for ECM health states. They move to further lines of treatment according to the progression rates described in more detail in section 4.2.6.

Patients in the avapritinib arm who progress receive BSC and will not subsequently be treated with TKIs. In order to incorporate separate health utilities, BSC is modelled as three health states: SOC1/SOC2/PD. Patients in the SOC1 and SOC2 health states have the same probability of progression and death as those in the ECM arm for second and third-line respectively.

Patients in the ECM arm receive imatinib as first-line, sunitinib as second-line and regorafenib as third-line therapy. After failing third-line therapy, they will receive BSC (i.e. no further TKIs).

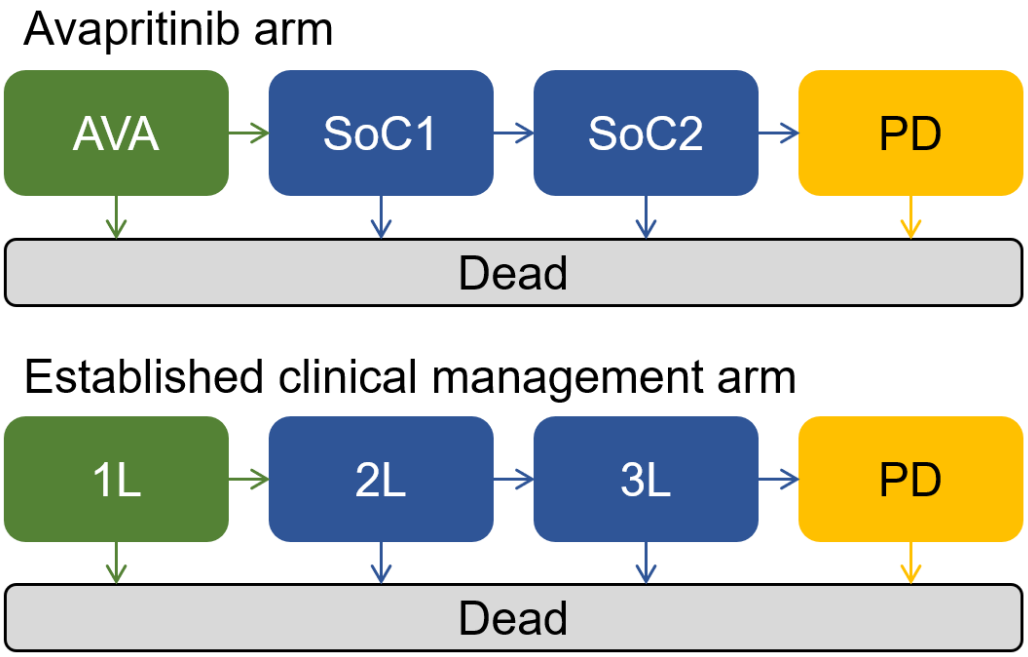


Figure 4 Structure of cost effectiveness model (reproduced from CS Figure 12)

The progression rates and death rates are taken from the NAVIGATOR study for avapritinib and BLU-285-1002 for ECM and are discussed in more detail in section 4.2.6.

The CS states "that the structure of the cost-effectiveness model is similar to the approaches used in previous NICE Technology Appraisals in unresectable or metastatic GIST" (TA86, TA179, TA488) .[11-13,39] The company note that those appraisals focused on one line of treatment (first-line, second-line and third-line GIST treatment respectively), whereas this appraisal considers the whole treatment pathway.

4.2.2.2 ERG critique of the model assumptions

The CS includes a table of modelling assumptions (CS Table 60). The ERG have added our views of these assumptions in Table 17 below.

Table 17 Company model assumptions (reproduced from CS Table 60)

Assumption	Justification	ERG comments
Clinical parameters and variables		

When a patient stops treatment with avapritinib, the benefit of avapritinib in terms of mortality is lost gradually.	• Clinical experts have suggested that the treatment effect is not lost immediately when a patient is discontinued from avapritinib and may continue for 60 months[41,42] • With no further information on the dynamics of the loss of this effect over time, linear interpolation was used	• Clinical experts advising the ERG have suggested that patients who discontinue avapritinib quickly have the same survival as those in the ECM arm.
Once a patient has lost the avapritinib treatment effect, it is appropriate to model their survival based on the ECM arm.	• Clinical experts confirmed in a survey that the overall survival of patients with unresectable or metastatic *PDGFRA* D842V-mutated GIST does not significantly differ as a result of treatment with imatinib, sunitinib, regorafenib or BSC	• We agree
The rate of further disease progression in patients with progressed disease in the avapritinib and ECM arms is the same.	• Clinical experts confirmed in a survey that the overall survival of patients with unresectable or metastatic *PDGFRA* D842V-mutated GIST does not significantly differ as a result of treatment with imatinib, sunitinib, regorafenib or best supportive care	• We agree
Health-related quality of life		
Health-state utility values from previous GIST appraisals are appropriate for decision making in this indication.	• No data are available to capture the specific HRQL of patients with this mutation • 100% of clinical experts consulted suggested that these values are representative • The progressive disease health-state utility value from TA179 was explored in a scenario analysis.	• We agree, however the utilities used for first-line are implausibly high (section 4.2.7).
Cost and health care resource use		
Excluding the costs of TKIs and management of adverse events, the cost of treating patients with metastatic or unresectable *PDGFRA* D842V-mutated GIST is the same as treating patients with general GIST.	• Excluding adverse events, there is no evidence to suggest that disease management costs will differ	• We agree
The use of branded pack costs for imatinib is appropriate.	• Generic imatinib is not currently approved by the EMA for use in GIST treatment. See CS section **Error! Reference source not found.**	• We agree

The first-line, second-line and third-line TKIs used in the treatment of patients with unresectable or metastatic GIST cost the equivalent to imatinib, sunitinib and regorafenib, respectively.	• In a survey of clinical experts, the majority of participants confirmed that, excluding patients who receive experimental therapies via clinical trials, compassionate use programmes or other means, patients with unresectable or metastatic *PDGFRA* D842V-mutated GIST in England and Wales are treated with imatinib, sunitinib and regorafenib, with most indicating that these would be used as first-, second- and third-line therapies, respectively, despite the lack of efficacy of these treatments • The mix of first-line therapies received by patients in the BLU-285-1002 study was explored in a scenario analysis	• Clinical advice to the ERG suggested that patients would not receive these therapies due to the lack of efficacy.

ERG conclusion

The three-state partitioned survival model is a standard modelling approach and has been applied in previous NICE appraisals for treatments for GIST. The company have adapted this approach to incorporate more lines of treatment. We consider that the model structure and partitioned survival approach is appropriate.

4.2.3 Population

The population included in the cost-effectiveness model is adult patients with unresectable or metastatic GIST with the *PDGFRA* D842V mutation. The population used in the economic model reflects the marketing authorisation and is in line with the NICE scope.

The ERG note that patients in the NAVIGATOR study had previously received TKIs, whereas the modelled population is assumed to have not previously received TKIs. The CS reports (CS Table 7) that "[redacted]% of patients had received prior treatment with imatinib (first line), [redacted]% had received prior treatment with sunitinib (second line) and [redacted]% had received prior treatment with regorafenib (third line). [redacted] patients ([redacted]%) had not received any prior TKI therapy". Although not explicit in the CS, we believe that all TKI use reported in NAVIGATOR would have been for unresectable or metastatic disease (i.e. not including adjuvant therapy), as explained in section 3.2.1.3 above.

The frequency of prior imatinib, sunitinib and regorafenib use (Table 5) was higher in the NAVIGATOR and BLU-285-1002 studies than would be expected in UK clinical practice. The company also provided a scenario analysis that used the study by Cassier et al[24,25] for the effectiveness of the ECM arm.

ERG conclusion

There is an inconsistency in the patient population between the economic model and the NAVIGATOR and BLU-285-1002 studies with regard to prior TKI use. The prior TKI use in these studies does not reflect UK clinical practice.

4.2.4 Intervention and comparators

The economic model compares the cost effectiveness of avapritinib versus ECM (consisting of first-line line imatinib, second-line line sunitinib and third-line regorafenib, followed by BSC). The CS states that the TKI treatments in the ECM arm have a lack of efficacy in this population with very low overall response rates and this was confirmed by the company's clinical experts. Clinical experts advising the ERG agreed and commented that for this reason few patients would receive TKIs in clinical practice. They estimated 20% of patients would receive imatinib and fewer than 10% of patients would receive sunitinb and regorafenib (although these estimates are uncertain). The ERG base case (discussed in section 6) assumes fewer patients receive these treatments in the ECM arm as suggested by our clinical experts.

ERG conclusion

The intervention (avapritinib) and the comparator (ECM comprising of first-line imatinib, second-line sunitinib and third-line regorafenib, followed by BSC) match the decision problem. However, in clinical practice not all patients would receive all the TKI treatments in ECM.

4.2.5 Perspective, time horizon and discounting

In the company's economic analysis, direct health effects of treatments are modelled and costs are estimated from the perspective of the NHS and Personal Social Services (PSS). Costs and outcomes are discounted at 3.5% in the base case (as recommended by NICE guidance[43]) and 0% discount is applied in scenario analyses.

In the base case, costs and QALYs are estimated over a lifetime time horizon (40 years). The cost-effectiveness results for alternative time horizons of 6 and 10 years are considered in scenario analyses and the model results are sensitive to changes in the time horizon.

The ERG agree that the lifetime time horizon adopted by the company in their base case is appropriate and in line with NICE guidelines.[43] The discounting rates and perspective used in the economic analysis are also consistent with NICE guidelines.

4.2.6 Treatment effectiveness and extrapolation

The CS notes that survival data in the NAVIGATOR are immature (median OS not reached) and therefore extrapolation is necessary to model mean survival. OS, PFS and ToT use the datasets from NAVIGATOR IPW for avapritinib and BLU-285-1002 for ECM. The company also provided a scenario analysis that used the study by Cassier et al[24,25] for the effectiveness of the ECM arm. ECM consisted of TKIs and BSC. The method for IPW is discussed above in section 3.5.3**Error! Reference source not found.**. The extrapolations for OS, PFS and ToT are discussed in more detail in the following sub-sections.

The company considered the visual fit of the survival extrapolation against the Kaplan-Meier data; the plausibility of the log-cumulative hazard plots, conditional probability plots, and the 1-, 2- ,5- and 10-year survival estimates; and the goodness-of-fit statistics, i.e. Akaike information criterion (AIC) and Bayesian Information Criterion (BIC). Base case survival results were validated against the opinions of UK clinical experts. The survival curves used in the model for each arm are shown in Table 18.

Table 18 Survival models used in the company base case for PFS, OS and ToT

Treatment arm	PFS	OS	ToT
Avapritinib	Weibull	Log-normal	Gompertz
ECM	1L Weibull; 2L Log-logistic; 3L Gompertz	Weibull	Uses PFS

4.2.6.1 Overall survival for patients receiving avapritinib

Modelling approach – censoring OS data for discontinuation

The OS for the avapritinib arm is estimated by combining survival estimates for patients who are on treatment and those who have discontinued treatment.

For those on treatment, the company fitted survival models to the OS IPW adjusted Kaplan-Meier data from NAVIGATOR, with events censored for discontinuation (see section 3.5.3 for more discussion of the IPW adjustment). The survival of patients who have discontinued treatment is modelled based upon the survival of patients who had received ECM, adjusted for the time since avapritinib discontinuation.

The ERG consider that the approach taken to estimate OS for avapritinib by combining the survival of those still on treatment and those who have discontinued treatment is reasonable;

however, we consider a more standard approach would be to fit survival curves to the Kaplan-Meier data for the whole population (CS Figure 6) and this approach would have a more complete dataset. For this reason, the ERG requested more explanation of the company's rationale behind their approach and a scenario using parametric curves fitted to Kaplan-Meier data without censoring for discontinuation (clarification response B1). The company provided more explanation on their approach for modelling OS. They state that this approach was taken to build a link between ToT and OS in order to allow a gradual loss of the treatment effect to be explicitly modelled. The company fitted parametric curves to the OS Kaplan-Meier data without including censoring for discontinuation (as requested by the ERG). The ICER for avapritinib versus ECM increased significantly for this scenario, compared to their base case assumption. Whilst fitting OS to the uncensored Kaplan-Meier data is preferable to the ERG, we have continued to use the company model in the ERG base case but corrected the OS extrapolation by varying the treatment waning duration.

Assumption of treatment waning

The company assume that the treatment effects of avapritinib persist after treatment discontinuation, with a gradual loss of treatment effect over 60 months after discontinuation. The CS states that clinical experts supported this assumption, but does not provide a rationale (e.g. whether the assumption reflects avapritinib's mechanism of action) and does not provide survival data over a long enough time period to validate this assumption. Based on the advice of our clinical experts, we do not consider the company's assumption of persistence of treatment benefits for avapritinib for five years to be appropriate. Our experts' view is that the risk of death for patients discontinuing avapritinib would rapidly increase to a similar risk as the ECM arm. We explore the impact of this assumption on the cost-effectiveness results in the ERG scenario analyses (discussed in section 6).

Curve fitting

The Kaplan-Meier plot for those patients on treatment is shown in CS Figure 13. The visual fit of the survival models against the Kaplan-Meier IPW adjusted plot is shown in CS Figures 14-16 and the statistical fit is shown in CS Table 34. The CS states that "given the low number of events in the Kaplan-Meier data, it is difficult to evaluate the fit of the parametric models". The log-normal model is used in the base case and this was supported by the company's clinical experts. We have a few concerns with the company's approach, as discussed below.

Firstly, NICE DSU guidance 14[44] states that the same distribution would be appropriate for both treatment arms. We therefore suggest the Weibull (which is used in the ECM arm) is a better survival model to use for avapritinib OS. Changing the distribution used for OS from the log-normal to the Weibull has a minimal effect on the cost effectiveness results. The ERG uses the Weibull model for avapritinib OS in the ERG base case in section 6.

Second, whilst the model fit for patients on treatment appears reasonable against the IPW-adjusted Kaplan-Meier data censored for discontinuation (CS Figures 14-16), the modelled OS for avapritinib differs from the IPW-adjusted Kaplan-Meier curve (CS Figure 6). In response to clarification, the company compared the OS Kaplan-Meier plot without censoring for discontinuation with the modelled OS (Clarification response Figure 14). They acknowledge that the modelled OS deviates from the OS Kaplan-Meier data (see Figure 7). The company, however, suggest that this is because NAVIGATOR is expected to underestimate the survival outcomes that would be observed in clinical practice.

We note that the discrepancy between the modelled and observed OS is largely because of the assumption that treatment effects persist beyond treatment discontinuation. As noted above, we do not support this assumption. We ran the model with differing waning durations and concluded that a waning duration of 1 month gives a close fit to the observed OS data. Therefore in the ERG base case (section 6) we have reduced the waning duration to 1 month and varied the waning duration in scenario analyses. Figure 5 shows the modelled OS compared to the observed data and the ERG's suggested approach with a waning duration of 1 month and Weibull OS distribution for avapritinib.

[redacted]

Figure 5 Avapritinib OS estimates for the company base case compared with KM data and the ERG's suggested approach

4.2.6.2 Overall survival for patients receiving ECM

The company use the IPW-adjusted BLU-285-1002 OS dataset for ECM. CS Figures 17 to 19 show the OS Kaplan-Meier data from BLU-285-1002 compared to the parametric models fitted. The AIC and BIC statistics are shown in CS Table 36. The CS states that "the Weibull parametric curve is applied at base case in the model because it has the best statistical fit as well as good visual fit to the observed data and in the long term." Table 19 shows the modelled OS for ECM compared to the observed Kaplan-Meier data. CS Table 35 shows the survival estimates for the other distributions.

Table 19 Modelled OS compared to IPW-adjusted KM survival data

Time	BLU-285-1002[a]	Company base case model
6 months	[redacted]	[redacted]
12 months	[redacted]	[redacted]
18 months	[redacted]	[redacted]
24 months	[redacted]	[redacted]
5 years	[redacted]	[redacted]
10 years	[redacted]	[redacted]

[a] CS Table 17

The ERG agree the Weibull is an appropriate distribution for OS in the ECM arm and provides a good fit to the observed data. We also agree that the exponential shows a reasonable visual and statistical fit to the observed data. We have used the exponential model for OS in the ECM arm in our ERG sensitivity analyses (section 6).

4.2.6.3 Progression-free survival for patients receiving avapritinib

The best AIC and BIC statistics for IPW-adjusted PFS for avapritinib were for the Weibull and exponential models. CS Figures 20 to 22 show the Kaplan-Meier data from NAVIGATOR compared to the parametric models fitted. The AIC and BIC statistics are shown in CS Table 40. The CS states that both of these showed reasonable statistical fits. The Weibull model was used because "the probability of progression is not expected to increase with time for patients treated with avapritinib". The exponential is presented as a scenario analysis (CS Table 64). Table 20 shows the modelled PFS for avapritinib compared to the observed data. We agree with the approach taken by the company and that there is a reasonable fit to the observed data.

Table 20 Modelled PFS compared to IPW-adjusted KM survival data

Time	NAVIGATOR IPW [a]	Company base case model
6 months	[redacted]	[redacted]
12 months	[redacted]	[redacted]
18 months	[redacted]	[redacted]
24 months	[redacted]	[redacted]
[a] CS Table 19		

4.2.6.4 Progression-free survival for patients receiving ECM

First-line imatinib treatment

CS Figures 25 to 27 show the IPW-adjusted PFS Kaplan-Meier data from the BLU-285-1002 study for first-line imatinib treatment compared to the parametric models fitted. The AIC and BIC statistics are shown in CS Table 43. The CS states that each model presents a similar fit to the PFS Kaplan-Meier data The best AIC and BIC statistics were for the Weibull and exponential models. The Weibull model was used as it "had the best statistical fit as well as good visual fit to the clinical data." The exponential is presented as a scenario analysis (CS Table 64). Table 21 shows the modelled PFS for the first-line TKI within ECM compared to the observed data. We agree with the approach taken by the company and that there is a reasonable fit to the observed data.

Table 21 Modelled PFS compared to IPW-adjusted KM survival data

Time	IPW BLU-285-1002 [a]	Company base case model
6 months	[redacted]	[redacted]
12 months	[redacted]	[redacted]
18 months	[redacted]	[redacted]
24 months	[redacted]	[redacted]
[a] CS Table 19		

Second-line sunitinib treatment

Figures 21 to 23 in CS Appendix O show the PFS Kaplan-Meier data from the BLU-285-1002 study for second-line sunitinib treatment compared to the parametric models fitted. The AIC and BIC statistics are shown in Table 62 in CS Appendix O. The CS states that each model shows a similar visual fit to the Kaplan-Meier data and the company's clinical expert suggested the log-logistic distribution was the most realistic and this was used in the company's base case. The exponential and Weibull distributions had the best statistical fit.

We disagree with the choice of the log-logistic distribution and suggest that the Weibull would be a better model to use as it is consistent with the first-line model and provides a better statistical fit. We have used the Weibull model for second-line sunitinib PFS in our ERG base case (see section 6).

Third-line regorafenib treatment

Figures 24 to 57 in CS Appendix O show the PFS Kaplan-Meier data from the BLU-285-1002 study for third-line treatment compared to the parametric models fitted. The AIC and BIC statistics are shown in Table 65 in CS Appendix O. The exponential and Weibull had the best statistical fit. The Gompertz distribution is used in the company's base case "because a decreasing probability of progression over time was considered clinically plausible, and the clinical expert consulted suggested that the resulting overall survival estimates (which rely on the choice of PFS curves for the ECM arm) were appropriate". We disagree with the choice of the Gompertz distribution and suggest that the Weibull would be a better model to use as it is consistent with the first-line model and provides a better statistical fit. We have used the Weibull model for third-line regorafenib PFS in our ERG base case (see section 6).

4.2.6.5 Time on treatment

ToT for avapritinib was based on the NAVIGATOR IPW data (CS Figure 28). For ECM, the model uses PFS for a proxy for ToT (clarification response B7).

CS Figures 29 to 30 show the visual fit of the parametric models to the observed data for avapritinib. The AIC and BIC statistics are shown in CS Table 46. The exponential and Weibull distributions had the best statistical fit. However, the Gompertz distribution is used as the "company's clinical expert suggested that the model using Gompertz extrapolation of avapritinib provided clinically plausible results." We disagree with the choice of the Gompertz distribution for ToT for avapritinib. We suggest that the Weibull would be a better model to use as it is consistent with the model used for PFS and provides a better statistical fit to the observed data.

The company provide an explanation for their approach to the calculation of ToT in clarification response B9: "The ToT used censors for death and progression in order to isolate the probability of discontinuing from avapritinib treatment. This allows it to be used independently of the probabilities of death and progression of disease." Therefore the company have adjusted ToT by scaling using the percentage remaining alive who are on

treatment by OS. The equation in column CN of the “Markov Avapritinib” model sheet is provided below for clarity:

$$E(c(AVA)_t) = \%\ alive \times \%\ on\ trt \mid alive \times 1\ month\ cost\ of\ AVA$$

As with OS, we believe it would be more straightforward and transparent to fit parametric curves to the observed Kaplan-Meier for ToT. A more complicated approach is more likely to introduce errors or bias. The estimates of ToT in the economic model do not show a good fit to the observed ToT in the NAVIGATOR IPW data (see **Error! Reference source not found.**) and we believe that the cost of avapritinib has been underestimated.

In order to produce a better fit with the observed ToT and for consistency with ECM, in the ERG base case (section 6) we set ToT for avapritinib to be equal to PFS. The ToT using the ERG approach compared with the company base case and the IPW adjusted Kaplan-Meier data are shown in Figure 6.

[redacted]

Figure 6 Modelled ToT for avapritinib compared to IPW adjusted KM data

4.2.6.6 Adverse events

Adverse events for the avapritinib arm are included in the economic model for Grade 3+ all-cause adverse events with incidence ≥ 2%, as measured in the 300/400 mg dose group (CS Table 47). Adverse event data for avapritinib are from all patients in NAVIGATOR study, not restricted to those with the *PDGFRA* D842V mutation. The CS states “there is no clinical

evidence to suggest that AEs would occur more frequently in patients with or without the *PDGFRA* D842V mutation; therefore, given the ultra-orphan nature of the condition, it was considered more appropriate to include evidence for the maximum number of patients to provide a clear safety profile for avapritinib." The ERG's clinical experts agreed with this assumption.

The most frequent Grade 3 adverse event was anaemia (%), with the remaining Grade 3 events occurring in < 7% of patients. Adverse events for the first-, second-, and third-line components of ECM are shown in CS Table 48.

Adverse events were incorporated by using the cycle probability of each adverse event. In the base case, the impact of adverse events was incorporated by estimating weighted average disutilities and costs per patient, as described below in sections 4.2.7.3 and 4.2.8.3.

ERG conclusion

- The methodology used to extrapolate OS, PFS and ToT for the economic model is generally appropriate and consistent with NICE recommended methodology.
- We disagree with the company's assumption that treatment effects persist up to five years. Including this assumption leads to a large discrepancy between the OS estimates in the model and the observed OS data.
- We disagree with the choice of the log-normal model for OS for the avapritinib arm and prefer using the Weibull model.
- We prefer the Weibull model for PFS for the ECM arm for the second-line and third-line treatments. The estimates of ToT in the economic model provide a poor fit to observed ToT. To produce a better fit with the observed ToT and for consistency with ECM, we assume ToT for avapritinib to be equal to PFS. We view that the Weibull is a better model for ToT for avapritinib.

4.2.7 Health related quality of life

4.2.7.1 Systematic literature review for utilities

The company conducted a systematic literature review from database inception until January 2020 aiming to identify health related quality of life (HRQoL) studies on patients with unresectable or metastatic *PDGFRA* D842V-mutated GIST (CS section B.3.4.3 and CS Appendix H). The company searched the same databases and applied the same

methodology used to identify cost-effectiveness studies (section 4.1). The inclusion and exclusion criteria used in the review are detailed in CS Appendix Table 36.

The systematic review did not identify any relevant HRQoL studies assessing patients with unresectable or metastatic *PDGFRA* D842V-mutated GIST but identified 18 studies assessing the overall unresectable or metastatic GIST population (CS Appendix Figure 6 and CS Appendix Table 37). Of these, three were for NICE Technology Appraisals for avapritinib comparators[11-13,39] and the remaining studies were not conducted from a UK perspective and/or reported the same utility data used in the NICE appraisals. The three NICE Technology Appraisals[11-13,39] informed the health state utility values used in the company's submission (section 4.2.7.2 below).

The ERG identified two additional studies presenting relevant utility evidence which were not included by the company.[45,46] However, both of them reported the same utilities as the company.

4.2.7.2 Health state utility values

Health state utility values were informed by previous NICE appraisals for imatinib, sunitinib and regorafenib in the treatment of unresectable or metastatic GIST.[11-13,39] In TA86/TA209 (imatinib), three clinicians answered a questionnaire to map patients' ECOG performance from the B222 trial[29] to EQ-5D. In TA179 (sunitinib) and TA488 (regorafenib), EQ-5D data were collected from patients in the A6181004[47] and GRID trials,[48] respectively. Utility values from the three previous appraisals were derived using UK preference scores. The ERG notes that the health state utility values collected from TA179 and TA488 are consistent with the NICE reference case[43] but we are unclear whether utilities from TA86/TA209 were derived directly from patients.

Previous NICE Appraisals for imatinib,[11,39] sunitinib,[12] and regorafenib[13] noted the following:
TA86/TA209: The Committee considered the utility value for the progression-free imatinib/sunitinib arm (0.935) questionable and implausibly high.
TA179: The ERG considered the source of utility values appropriate. The TA179 ERG raised some uncertainty about the utility of 0.577 for patients in progressive disease, but the ICER was insensitive to this. The NICE Appraisal Committee had the same uncertainty as the ERG regarding the progressive disease utility but agreed that the ICER was rather insensitive to variations in the utility values and therefore considered the utility values adequate.

TA488: The ERG considered the source and instrument used to measure utility values appropriate and used the same utility values as the company in their base case. The NICE Appraisal Committee accepted the company and the ERG base case utility values.

Table 22Table 22 presents the health state utility values used in the company's base case analysis in the present appraisal. We note that this set of utilities was assessed by five clinical experts advising the company, who reportedly agreed these values to be reflective of patients with unresectable or metastatic GIST in UK clinical practice. To assess the impact of utilities on the overall cost effectiveness results, the company also conducted a scenario analysis using an alternative progressive disease utility value, from TA86/209[11,39] and TA179[12] (CS Table 64). The ICER increased less than £1,000 for this analysis.

Table 22 Health state utility values used in company's base case analysis

Health state	Utility value, mean (SE)	Source	ERG comment
AVA/1L	0.935 (0.094)	TA86/TA209[11,39]	Utility value from the progression free health state for the imatinib / sunitinib arm of TA86/TA209.
SOC1/2L	0.781 (0.078)	TA179[12]	Utility value from the progression free health state from the best supportive care arm of TA179.
SOC2/3L	0.767 (0.077)	TA488[13]	Utility value from the progression free health state from the entire cohort, measured at baseline and not split by arm, of the GRID trial[48] [a]
Progressive Disease	0.647 (0.065)	TA488[13]	Utility value from the progressive disease health state from the entire cohort of the GRID trial[48] [a]

Table reproduced from CS Table 50
AVA: avapritinib; 1L: first line; 2L: second line; 3L: third line; SOC1/SOC2: standard of care
[a] In the GRID trial, EQ-5D-3L was administered to collect HRQoL data for PFS and PD health states of patients receiving either regorafenib 160mg or placebo, plus BSC. The GRID trial reports utilities for the entire cohort and also split by treatment arm.

We note that the utility value for the first line health state (AVA/1L) is higher than the UK population norm for this group. According to Ara and Brazier,[49] the utility of this age group is 0.822. In line with the NICE Appraisal Committee's assessment of TA86/TA209 (see section 4.2.7.2), the experts who provided clinical advice to the ERG also considered the first line

utility value an overestimation, stating that patients in this health state would at best experience a quality of life equal to the general population (but likely to be lower). We therefore used the utility value of 0.822 for the first-line health state in our base case analysis (see section 6).

We also note that the utility data informing SOC2/3L were collected at baseline from the entire cohort of the GRID trial,[48] pooling the regorafenib and placebo arms. We recognise that measuring HRQoL at baseline does not capture the effect of regorafenib or placebo in the management of the disease. However, we agree that the utility value used by the company for SOC2/3L is acceptable.

In general, the ERG agree with company's approach to estimate health state utility values (except for the utility value for first-line PFS) and consider that they were informed by appropriate sources.

4.2.7.3 HRQoL data from the VOYAGER study

In response to ERG clarification question B6, the company provided HRQoL data from the VOYAGER study as supplementary evidence. The VOYAGER study is an ongoing open-label, randomized, phase 3 study of avapritinib versus regorafenib in patients with locally advanced, unresectable or metastatic GIST previously treated with imatinib and one or two other TKIs. The ERG note that the HRQoL was collected as an exploratory endpoint using the EQ-5D-5L questionnaire, but no other details of HRQoL data collection were reported by the company. EQ-5D-5L data for the mutated subgroup (*PDGFRA* D842V) is based on a limited number of patients ([redacted] for avapritinib and [redacted] for regorafenib), which the company argue to be not representative of the whole unresectable or metastatic *PDGFRA* D842V population and we agree. Nonetheless, the ERG consider the HRQoL data from the VOYAGER study relevant since it is the most recently available data coming from a large sample of patients with unresectable or metastatic GIST ([redacted] patients treated at third-line and [redacted] at fourth-line), although mostly without the *PDGFRA* D842V mutation, and this is the only study collecting quality of life data from patients treated with avapritinib. Therefore, we conducted a scenario analysis using the utility values for the overall GIST population from the VOYAGER study (see section 6). Table 23 presents the mean baseline utilities for the overall GIST population from the VOYAGER study - third line utility values from the VOYAGER study ([redacted]) were used in SoC2/3L health state and the fourth line ([redacted]) in the PD health state. The company reported the baseline values, stating that they are the most appropriate for the economic model because "there were no trends identified in utility values

over time, especially when removing time points with less than 5 patients measured" and, contrarily to baseline, follow-up values can introduce bias "as patients progress and drop out of the analysis, leaving behind the healthiest patients."

Table 23: Mean baseline utilities for avapritinib and regorafenib treated patients at third or fourth line (ITT population)

	Patients treated at third line			Patients treated at fourth line		
	Avapritinib n=198	**Regorafenib n=187**	**Total n=385**	**Avapritinib n=30**	**Regorafenib n=32**	**Total n=62**
Utilities	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Source: Reproduced from Table 3 of Supplementary health-related quality of life data from the VOYAGER study.

ERG conclusion

Whilst the VOYAGER data are relevant (as the only study that provides HRQoL data directly for avapritinib), they are based on a very small sample size for the PGDFRA D842V mutation subgroup. Therefore, we consider it appropriate to include the utility data obtained from this study in a scenario analysis (as shown in section 6) rather than in our preferred base case.

4.2.7.4 Adverse event utility decrements

The company included utility decrements associated with adverse events of grade 3-4 (CS Table 49) and assumed that adverse events of grade 1-2 have no disutility. The most common adverse events of grade 3-4 (>20%) are anaemia (reported by [redacted]% of patients receiving avapritinib), dermatitis/rash and hypertension (reported by [redacted]% and [redacted]% of patients receiving regorafenib, respectively) (CS Table 47 and Table 48).

Utility decrements are based on published articles and previous NICE appraisals. The ERG were unable to reconcile the utility decrement of hypertension with the corresponding source.[50] In clarification response B10, the company provided more details on the source of this utility decrement, which is Table 75 of the review of TA176 and partial review of TA240 for first line treatment of metastatic colorectal cancer.[51] We note that the cited table reports the adverse event utility decrements used in the Merck Serono model, in which hypertension was informed by Doyle et al.[50] However, we are unable to find a utility decrement for hypertension in the study of Doyle et al.[50] We identified a study reporting an alternative utility decrement value (0.153) for this adverse event.[52] This utility decrement is, however,

considerably higher than the value used in the company's submission (0.069) and we note that hypertension mainly occurs in patients receiving regorafenib. We believe that the alternative value will likely have a small impact on the model results, and that the utility decrement from the company's submission provides a more conservative approach (i.e. likely to favour the comparator).

When a suitable source to estimate a utility decrement was not found for an adverse event, the company assumed utility decrements from similar conditions or, when none were available, the maximum of the other utility decrements for the adverse events was assumed. This is a conservative assumption, because patients treated with avapritinib experienced a higher rate of all adverse events to which the maximum utility decrement was applied except two (abnormal liver function and haemorrhage).

A considerable number of patients receiving avapritinib experienced cognitive effects ([redacted]%) - memory impairment, cognitive disorder, confusional state and encephalopathy. However, the company only report a utility decrement for the confusional state, in which the maximum utility decrement was used (0.200), because this is a Grade 3-4 adverse event. The ERG note that the other cognitive adverse events are mild and/or occur in few patients (<2%), therefore are unlikely to influence HRQoL. We consider that the company have appropriately explored the impact of this special interest group of adverse events in HRQoL of patients with unresectable or metastatic *PDGFRA* D842V-mutated GIST.

The mean duration of adverse events in the model is seven days (CS section B.3.4.4), informed by a previous NICE appraisal (TA349)[53] and by the study of Freeman et al.[54] The clinical experts advising the ERG agreed that this period is appropriate.

4.2.7.5 Age-related utility decrements

The company account for utility decrements related to age by using the algorithm provided by Ara and Brazier.[49] We agree that this is appropriate and it is recommended by NICE DSU Technical Support Document 12.[55]

ERG conclusion

The company's review of HRQoL evidence is robust and relevant to the decision problem. The approach to estimate health state utility values, adverse events and age-related utility decrements is appropriate and consistent with the NICE reference case. However, we consider the utility values for first line treatment to be implausible

and prefer to use the UK population norm for the first-line health state. Although the ERG is not aware of the details of HRQoL data collection in the VOYAGER study, we consider that this study reports relevant HRQoL data for the unresectable or metastatic GIST population. Therefore, we conducted a scenario analysis using the utility values for the overall GIST population from the VOYAGER study for the SoC2/3L and PD health states.

4.2.8 Resources and costs

The costs included in the economic model consist of drug acquisition for first-line and subsequent treatments, health state management cost, costs for managing adverse events and terminal care costs incurred at the end of life.

The company conducted a comprehensive literature search to identify costs and resources used in the treatment and management of unresectable or metastatic GIST. The original search was completed on 5th December 2019. The search was limited to those studies published after 2009. Details of the search strategy and eligibility criteria are shown in CS Appendix I. The same study selection methodology was applied as the systematic literature review of published cost-effectiveness studies (section 4.1 above and CS Appendix G.2). The searches identified one study by Schoffski et al.[56] in which the core population was patients with unresectable or metastatic GIST with a *PDGFRA* D842V mutation. Schöffski et al. evaluated the financial impact of imatinib palliative therapy in metastatic GIST patients in Belgium, and the potential cost saving by a tailored use of imatinib based on genotyping. A further 20 studies for patients with unresectable or metastatic GIST were identified (CS Appendix I Table 45).

The resources used in the company's model were largely based upon those used in the Technology Appraisal for regorafenib for unresectable or metastatic GIST (TA488).[13]

4.2.8.1 Drug acquisition

The acquisition cost per pack for each drug is taken from the Monthly Index of Medical Specialties (MIMS).[57] Intended dosages were adjusted by the dose intensity observed in the treatments' trials. However, the relative dose intensity for imatinib was not reported in TA86 so the company assumed that there are no dose reductions or escalations for imatinib patients.

Avapritinib is an oral treatment and is licensed at 300mg QD. It is available at doses of 100 and 200 mg but all doses have the same cost. The list price of avapritinib is £26,666.67 for

30 tablets. Avapritinib is supplied to the NHS with a confidential patient access scheme (PAS) respectively.

The dosing, frequency and unit costs of the drugs are shown in Table 24 (CS Tables 51 to 55). Sunitinib has a PAS where the first six weeks of a treatment cycle are free and regorafenib has a confidential PAS. The company has reported all analyses using the list price of the comparator treatments and the PAS price for avapritinib. The ERG have replicated the company's analyses using the comparator treatment PAS prices in a separate confidential appendix to this report.

Table 24 Dosing, frequency and unit costs per administration for intervention and comparator

Drug	Daily dose	Number of capsules per pack	Pack price	Dose intensity	Cost per model cycle (1 month)
Avapritinib	300 mg	30	£26,666.67	[redacted]	[redacted]
Imatinib	400 mg	30	£1,133.41	100%	£1,149.94
Sunitinib	50 mg	28	£3,138.80	97%	£2,206.45
Regorafenib	40 mg	84	£3,744.00	87%	£3,540.84

As discussed in section 4.2.4, clinical experts advising the ERG commented that not all patients would receive all treatments in the ECM arm and that very few patients would receive second-line sunitinib and third-line regorafenib (<10%). The ERG base case (section 6) therefore assumes fewer patients receive these treatments (20% of patients receive imatinib, and 10% of patients receive sunitinib and regorafenib) in the ECM arm. For further details, refer to section 6.

4.2.8.2 Health state unit costs

The resource use and unit costs of the progression-free and progressed disease health states are shown in CS Tables 56 and 57. The CS states that health care resources for unresectable or metastatic GIST with the *PDGRA* D842V mutation are unlikely to significantly differ from general unresectable or metastatic GIST population. The health care resources were taken from the regorafenib Technology Appraisal (TA488) [13] and cost values were taken from NHS reference costs (2018-19).[58] The resource use in TA488 were based upon a survey conducted in 2013 involving 15 physicians in England and Wales. The

health care resource costs consist of one-off costs of tests taken by a proportion of patients before treatment and regular resource use per patient including pain management. Resources consist of CT and MRI scans, liver function and blood tests and outpatient visits. The one-off costs were £575.08 and £456.47 for patients in the progression-free and progressed disease health states respectively. The regular resource use costs were £217.86 and £247.32 per model cycle for patients in the progression-free and progressed disease health states, respectively.

Clinical experts advising the ERG agreed that there would be no difference in resource use between patients with this mutation and the general unresectable or metastatic GIST population and that the resource use estimates in the model are appropriate. The ERG note that the cost for an outpatient care visit in NHS reference costs was £194.17 rather than £190.64 used in the CS and company model. We obtained this value from NHS reference costs 2018/19, CL tab, currency code WF01A, service code 370, service description: Medical oncology.This was corrected in the ERG analyses (section 6).

The ERG's clinical experts also suggested that some patients with progressed disease would have fewer investigations, such as patients on palliative care. Patients receiving palliative care may transfer from hospital to hospice, so would be followed up less intensively. The experts suggested that around two thirds would continue to have regular follow-up investigations and around one third would not. We have changed resource use for these patients in the ERG analyses to the values suggested by our clinical experts (section 6).

4.2.8.3 Cost of terminal care

The company's model includes a cost of end-of-life care of £9,144.20 based upon TA488 (CS Table 59) and inflated to 2018/19 prices using the Personal Social Services Research Unit (PSSRU) Unit Costs of Health and Social Care.[59] The original cost was taken from a study by Abel et al.[60] that presents end-of-life costs for a cohort of hospice patients in South West England. The study estimated that 16% of patients die in hospital and 84% die outside of hospital. The ERG's clinical advisors agreed with these estimates.

4.2.8.4 Adverse event costs

The model includes the costs of managing grade 3+ adverse events. For each model cycle, the cost of the adverse event was multiplied by its probability by cycle and the proportion of patients on treatment. The unit costs used for the management of adverse events were

taken from the latest NHS reference costs 2018/19.[58] The unit costs of the management of adverse events are shown in CS Table 58. For several of the adverse events there was no detail of the HRG code used and the ERG requested further information on these. The company provide further information on these in clarification response B5.

ERG conclusion

Fewer patients would receive TKIs in the ECM arm than assumed in the company's model. The ERG have concerns on the dose intensity used for the comparator treatments. We consider the dose intensity should be similar between the TKIs and avapritinib.

5 COST EFFECTIVENESS RESULTS

5.1 Company's cost effectiveness results

The company present their base case results for avapritinib versus ECM in CS Section 4.3.8. These results, reproduced below in Table 25, show that avapritinib provides a mean QALY gain of [redacted] for an additional mean cost of [redacted] : giving an incremental cost-effectiveness ratio of £49,996 per QALY.

The cost-effectiveness results presented in the CS include a confidential PAS price for avapritinib but do not include existing PAS discounts for the comparators (sunitinib and regorafenib for the ECM arm). The results including all agreed PAS discounts for comparators as well as the company's proposed price discount for avapritinib are presented in a confidential addendum to this ERG report.

Table 25 Cost effectiveness: Company's base case (discounted)

Technologies	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£/QALY)
ECM	[redacted]	[redacted]	-	-	-
Avapritinib	[redacted]	[redacted]	[redacted]	[redacted]	£49,996
Source: CS Table 61					

5.2 Company's sensitivity analyses

5.2.1 Deterministic sensitivity analyses

The company summarise the parameters and ranges included in the deterministic sensitivity analysis (DSA) in the CS Appendix M. The DSA are presented as a tornado plot in CS Figure 33. The plot shows that the baseline patient age, discount rates and health state utility values are key drivers of the model results. Other parameters such as management costs and HRQoL parameters for the general population also influence the results, but to a lesser extent. The DSA did not include parameters related to clinical effectiveness, except for proportion of deaths in the pre-progression stage for the Cassier ECM study. The company argue that their base case clinical effectiveness estimates are likely to be underestimates as the data from NAVIGATOR constituted a mixture of patients at different treatment lines whereas, in clinical practice, avapritinib would be used as a first-line therapy. To examine the impact of the clinical effectiveness parameters on the overall model results as well as a range of other parameters, the ERG conducted a range of sensitivity analyses, details of which are discussed in section 6.

5.2.2 Scenario analyses

The company conducted a range of scenario analyses to analyse the impact of key variables on the model outcomes in CS Table 64.

Whilst most of the scenarios did not have a significant impact on the cost effectiveness results, the use of shorter time horizons (i.e. 6 years and 10 years) had the greatest impact on model results. We extended the range of scenario analyses in ERG additional analyses, described below (see section 6).

5.2.3 Probabilistic sensitivity analysis

The company conducted a probabilistic sensitivity analysis (PSA) on their base-case model to assess parameter uncertainty. In summary, the company assigned the normal distribution for efficacy parameters; the beta distribution for health state utilities and adverse event disutilities; and the gamma distribution for the costs and duration of adverse events. Probabilistic results are presented in CS Table 63; scatter plots in CS Figure 31; and the cost effectiveness acceptability curve (CEAC) in CS Figure 32. The company report that the PSA results are close to the deterministic results. The CS states that at a willingness-to-pay threshold of £50,000 per QALY, avapritinib had 42.4% probability of being cost-effective compared to ECM.

Whilst we consider that the company assigned appropriate distributions to the model parameters; the ERG were unable to replicate the PSA simulations. The company provided a revised version of their model with corrections in response to clarification question B4 to appropriately reflect uncertainty over the input parameters. The ERG were able to replicate the PSA in this updated version of the model.

5.3 Model validation and face validity checks

5.3.1 Company validation of their model

The company describe their approach to model validation in CS section B.3.10. They state that they conducted clinical validation of the survival estimates produced by the cost-effectiveness model, health states utilities, current and avapritinib treatment pathways, and healthcare resources used.

The key conclusions that the company drew from the validation exercise are as follows:

- Two clinical experts independently agreed with the company's estimates of PFS and OS;
- Five clinical experts agreed with the health state utilities used by the company for their base case;
- In the company's survey of clinical expects, most of the experts are reported to agree with the treatment sequences used in the company's model for the ECM arm for patients with unresectable or metastatic PDGFRA D842V-mutated GIST. Nonetheless, the company conducted a scenario analysis where they include the drug costs of additional TKIs in BLU-285-1002 study, which are not currently approved for the treatment of GIST patients in England and Wales. Further details on the mix of TKIs in BLU-285-1002 study is presented in CS Appendix Q. Including the additional TKIs did not have a significant impact on the cost effectiveness results.
- There was a lack of clinical consensus on the healthcare resources used at different treatment lines. To address this issue, the company conducted a scenario analysis where values suggested by the clinical survey results (further details in CS Appendix R) were used. This scenario increased the ICER for avapritinib versus ECM by approximately [redacted] compared to the company's base case ICER.

To check for face validity, the company compared the modelled outcomes with the clinical data censored for discontinuation or death, rather than full dataset (CS Appendix J). We reproduce these results in Table 26 below. We note that the PFS and OS estimates from the raw data (sources are in Table 26) are slightly higher than the modelled estimates. The company defends this by citing that i) OS for avapritinib is estimated using ToT, the raw data for avapritinib and raw data for ECM; and ii) this would be true when the raw avapritinib Kaplan-Meier data fails to reach median (as can be seen at different years from baseline).

Table 26 Comparison of the modelled outcomes with clinical data as reported by the company

Treatment	Data (source)	Year 1	Year 2	Year 3	Year 4	Year 5
PFS						
Avapritinib	Raw data (NAVIGATOR, IPW adjusted censored for death)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Modelled data	87%	67%	48%	31%	18%
ECM	Raw data (BLU-285-1002, IPW adjusted censored for death)	6%	6%	0%	0%	0%
	Modelled data	3%	0%	0%	0%	0%
OS						
Avapritinib	Raw data (NAVIGATOR, IPW adjusted censored for discontinuation)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Modelled data	99%	95%	89%	81%	70%
ECM	Raw data (BLU-285-1002 IPW adjusted)	48%	34%	19%	17%	14%
	Modelled data	45%	29%	20%	14%	11%
Source: CS Appendix J Table 46						

In response to clarification question B2, the company compared the modelled outcome of OS with the observed data from the NAVIGATOR IPW analysis and the BLU-285-1002 study, applying various durations of treatment waning (reproduced in Figure 7 below). The modelled OS curve is an overestimate compared to the OS Kaplan-Meier data from NAVIGATOR. The company argue that better survival would be expected in clinical practice when avapritinib is used first line, rather than when used at a later line of treatment as in NAVIGATOR (clarification response B1). We view that the company's argument may not be applicable for patients who have had several prior TKIs as in the case of participants in the NAVIGATOR study where they received more frequent prior TKIs than would be expected in UK clinical practice. Therefore, we treat the modelled overestimation of OS with caution and explore this further in our scenario analyses in section 6.

[redacted]

Source: reproduced from Figure 14 in clarification response B2

Figure 7 Comparison of modelled OS outcomes against clinical data

ERG conclusion

The company conducted appropriate internal validity and face validity checks. Whilst there are no previous technology appraisals in GIST population with PDGFRA D842V mutation, we view it reasonable to compare the model results in the current appraisal against three previous appraisals in GIST (including an update) i.e. TA86/TA209; TA179 and TA488, to provide some means of cross-validation.

5.3.2 ERG validation of the company's model

The ERG checked the economic model for transparency and validity. We conducted a range of tests to verify model inputs, calculations and outputs:

- Cross-checked all parameter inputs against values reported in the CS and cited sources;
- Checked the individual equations within the model;
- A range of extreme value and logic tests to check the plausibility of changes in results when parameters are changed;
- Recoded sections within the Markov calculations for the ECM arm to check model calculations;
- Checked all model outputs against results cited in the CS, including the base case, PSA, DSA and manually ran all the scenarios.

The company model was generally well-implemented with no substantive errors in parameter inputs or coding, except two issues as discussed in section 5.3.3.

Face validity checks

We consulted our clinical experts to validate the company's assumptions relating i) the impact of survival benefit from treatment waning; and ii) likely survival estimates observed in clinical practice for patients with GIST. The observations are summarised in Table 27. We conducted scenario analyses to address our experts' opinions; see Section 6.

Table 27 Comparison of the model assumptions with ERG clinical experts' opinions

Aspect	Company assumption	ERG clinical experts' opinion
Treatment waning	After patients discontinue treatment with avapritinib, the company assume that there will be a gradual loss of treatment effect, rather than losing all survival benefits immediately.	Patients' responses to TKIs are variable. While gradual waning may be true for some patients, others may exhibit rapid progression. Our experts viewed the overall company estimate of five years for treatment waning to be too long. For patients who stopped treatments due to progression, it was highly likely that they would rapidly revert to the rate of death that would be expected in untreated patients. Generally, most patients not on treatment would die within 12-18 months.
OS estimates for the ECM arm from the company's base case	5 years: 10.7% 10 year: 3.1%	5 years: Between 5% and 11% 10 years: 0%

Cross validity checks against previous Technology Appraisals

We compared the modelled QALY estimates from the current appraisal with three previous NICE technology appraisals (including an TA update) for treatments for patients with GIST (TA86/TA209; TA179 and TA488). Despite methodological differences between the models, they provide some means of cross-validation. We note that the QALY and life year estimates from the current appraisal are [redacted] than the other available lifetime model: e.g. the QALY estimate from TA488 was 0.969 for the comparator treatment (BSC) compared to [redacted] for ECM in the current appraisal (Table 28).

Table 28 Comparison of modelled outcomes

Source (time horizon)	QALYs	QALYs	Life Years
Current appraisal (Lifetime-40 years)	Avapritinib	[redacted]	[redacted]
	ECM	[redacted]	[redacted]

Source (time horizon)	QALYs	QALYs	Life Years
TA86/TA209 (10 years) [a]	Path 1: BSC	2.397	4.154
	Path 7: Sunitinib	2.411	3.716
	Path 4: Imatinib 600 mg	4.256	5.211
	Path 3: Imatinib 600 mg, followed by sunitinib	4.286	5.032
	Path 6: Imatinib 800mg	3.635	4.506
	Path 5: Imatinib 800 mg, followed by sunitinib	3.659	4.336
	Path 2: Imatinib 600-800 mg, followed by sunitinib	4.803	5.278
TA179 (6 years)	Sunitinib	1.23	1.98
	BSC	0.73	1.21
TA488 (40 years) [b]	Regorafenib	1.733	NA
	Placebo + BSC	0.969	NA

[a] The 7 strategies represent 7 model pathways: Path 1- patients receive BSC; Path 2- treatment with escalated doses of imatinib (600 and 800mg/day) followed by sunitinib; Path 3- treatment with escalated dose of imatinib 600mg/day followed by sunitinib; Path 4- treatment with imatinib 600mg with no treatment switching; Path 5- treatment with escalated dose of imatinib 800mg/day followed by sunitinib; Path 6- treatment with imatinib 800mg; and Path 7- treatment with sunitinib

[b] Based on company's revised base case using 2017 data cut and GRID trial treatment duration;[48]
NA: Not Publicly Available; BSC: Best Supportive Care

5.3.3 ERG corrections to the company model

As previously stated, the company model was generally well-implemented, with no substantive errors in parameter inputs. The ERG, however, identified two errors in the model which are summarised below in Table 29.

Table 29 ERG corrections to the company model

Issue	Model aspect	Issue	ERG correction
1	Estimation of outpatient care visit from NHS Reference costs	The company model used an estimate of £190.17	The corrected value is £194.17 (see below Table 30).
2	PSA simulations	We were unable to replicate the company's PSA simulations.	The company addressed this issue in their response to clarification questions (for further details see clarification response B4). The ERG could replicate the PSA simulations using the revised code provided by the company.

The ERG addressed Issue 1 by re-running the analysis with the corrected value of £194.17 for an outpatient care visit. The overall impact of this change is small i.e. an increase in the base case ICER from £49,996 (company's base case) to £50,033 per QALY (see Table 30).

Table 30 Cost effectiveness results from ERG correction of Issue 1 (discounted)

Therapy	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£/QALY)
ECM	[redacted]	[redacted]	-	-	-
Avapritinib	[redacted]	[redacted]	[redacted]	[redacted]	£50,033

We re-ran all the company's scenario analyses (presented in CS Section B.3.8.3 Table 64) with the corrected model and the results are presented in Table 31 below. The cost-effectiveness results in these scenarios are similar to those from the company's scenario analyses with the ICERs increasing minimally (approximately £100) across each of these scenarios.

Table 31 Results of the company's scenario analysis using the ERG corrected company model (discounted)

Scenario		Total costs	Total QALYs	ICER (£/QALY)
Base case (ERG corrected)	Avapritinib	[redacted]	[redacted]	
	ECM	[redacted]	[redacted]	£50,033
No discounting	Avapritinib	[redacted]	[redacted]	
	ECM	[redacted]	[redacted]	[redacted]
No discounting for outcomes	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
No discounting for costs	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Time horizon: 6 years	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Time horizon: 10 years	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
ECM TKIs from BLU-285-1002	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Incomplete loss of treatment benefit: 10%	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Incomplete loss of treatment benefit: 20%	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Post-avapritinib progression rate slowed by 10%	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Post-avapritinib progression rate slowed by 20%	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
End of life costs from Round et al.	Avapritinib	[redacted]	[redacted]	[redacted]

	ECM	[redacted]	[redacted]	[redacted]
PD utility from sunitinib TA	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Palliative surgery, radiotherapy, hospitalizations from clinical Survey	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Cassier et al.survival for the comparator arm	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Overall survival: Avapritinib - log-logistic	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Progression-free survival: Avapritinib - exponential	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Progression-free survival: ECM - exponential	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]

5.3.4 ERG summary of key issues and additional analyses

A full summary of ERG observations on key aspects of the company's economic model is presented in Table 32.

Table 32 ERG observations of the key aspects of the company's economic model

Issues	Features of the company model	ERG comments	ERG analysis	Priority issues to consider
Modelled decision problem				
Population	The modelled patient population is described in CS section 3.2.1.2	The model population is appropriate for the scope and the anticipated marketing authorisation. However, patients in the model are assumed to have no previous TKIs unlike patients in the NAVIGATOR and BLU-285-1002 studies. Secondly, the prior TKI use in these studies does not reflect the UK clinical practice.		HIGH (remains unresolved)
Intervention & comparators	• Intervention: Avapritinib (1st line), SoC (2nd line) and SoC (3rd line) • Comparator: ECM which comprises of imatinib (1st line), sunitinib (2nd line) and regorafenib (3rd line)	The intervention and comparators align with the NICE scope, although in clinical practice not all patients would receive all the comparator treatments. Based on the ERG's clinical advice, few patients would receive 2nd and 3rd line treatments due to lack of efficacy in this patient population combined with drug toxicity.	***ERG base case:*** For the ECM arm, 20% of patients receive 1st line imatinib, followed by 10% of patients receiving 2nd line sunitinib and 3rd line regorafenib. ***ERG scenario:*** For the ECM arm, 0% receives 1l, 2L and 3L.	HIGH
Assumptions about treatment				
Dose intensity	***Company base case:*** • Avapritinib: [redacted] • Imatinib: 100% • Sunitinib: 97% • Regorafenib: 87%	Our clinical experts view that the dose intensity is similar amongst the TKIs.	***ERG base case:*** Dose intensity of imatinib, sunitinib and regorafenib same as that of avapritinib	LOW
Model structure and framework				
Model type	Cohort partitioned survival model (CS Figure 12).	The overall model structure is appropriate, consistent with previous TA models in GIST and accurately implemented.		
Cycle length	1 month	The ERG agree with this assumption		

Half cycle correction	A half cycle correction was applied by using the mean number of patients in each health state at the beginning and end of each cycle to calculate costs and QALYs	Consistent with NICE methods guidance.		
Time horizon	40 years (patients enter the model at [redacted] years of age)	Consistent with a lifetime horizon and NICE guidance		
Clinical parameters				
Overall survival	***Treatment waning effect:*** The company's base case model assumes a gradual movement of OS hazard from the avapritinib arm to that of the ECM arm upon discontinuation from avapritinib treatment, meaning a gradual loss of treatment effect over a period of 5 years (60 model cycles).	The ERG's clinical experts advise that the company's assumption is not reflective of clinical practice (see section 4.2.6 for further details).	***ERG base case:*** Duration of treatment waning effect is 1 month ***ERG scenarios:*** Treatment waning ranging between 1 and 24 months	HIGH
	OS survival data: OS for avapritinib is based on patients who are still receiving avapritinib, i.e. the OS for this arm is censored for discontinuation	The model OS has a poor fit to the OS data in NAVIGATOR. This is corrected by changing the assumption of treatment waning.		
	Extrapolation (company's base case): • Avapritinib: Log-normal • ECM: Weibull	We agree that the Weibull is an appropriate fit for the ECM OS curve. To align with the NICE DSU methods guide,[44] both treatment arms should preferably use the same distribution (see section 4.2.6).	**ERG base case:** • Avapritinib: Weibull • ECM: Weibull **ERG scenario:** • Avapritinib: Exponential • ECM: Exponential	MEDIUM
Progression free survival	***Extrapolation:*** • Avapritinib: Weibull (base case) • ECM: - 1L: Weibull (base case); exponential (scenario) - 2L: Log-logistic (base case) - 3L: Gompertz (base case)	We agree with the extrapolation methods, but consider that the Weibull model provides a better fit to the observed data and would be more consistent with that used for 1L (see section 4.2.6).	***ERG base case:*** • Avapritinib: Weibull • ECM: - 1L: Weibull - 2L: Weibull - 3L: Weibull	MEDIUM
Time on Treatment	For the avapritinib arm, the ToT is censored for discontinuation	The method used to estimate ToT does not provide a close fit to the ToT KM data. For consistency with the ECM arm, we have set ToT equal to PFS for avapritinib (see section 4.2.6).	***ERG base case:*** We have set ToT equal to PFS for avapritinib.	HIGH

	The ToT is extrapolated using a Gompertz curve	We note that a Weibull curve provides a better statistical fit to the observed data and is consistent with the model used for PFS.	***ERG base case*:** Weibull	MEDIUM
Mortality	The model uses general population all-cause mortality rates adjusted for age and gender from UK Life tables (ONS data for 2015-2017). The excess mortality for GIST is obtained from the OS estimates.	The data for all-cause mortality are slightly out-of-date. For completeness, the ERG consider it appropriate to use the ONS data for 2016-2018.	**ERG base case:** Using ONS mortality data for the year 2016-2018.	LOW
Utilities				
Health state utilities	Company base case model estimates: • PFS 1L: 0.935 • PFS 2L: 0.781 • PFS 3L: 0.767 • PD: 0.647 Company scenario analysis: • PFS 1L: 0.935 • PFS 2L: 0.781 • PFS 3L: 0.767 • PD: 0.577	The utility for PFS 1L is implausible as this value is higher than the general population utility of 0.822. The estimates for 2L ,3L and PD appear to be appropriate	***ERG base case**:* • PFS 1L: 0.822 • PFS 2L: 0.781 • PFS 3L: 0.767 • PD: 0.647 ***ERG scenario:*** • PFS 1L: 0.822 • PFS 2L: 0.781 • PFS 3L: 0.767 • PD: 0.577	HIGH
	HRQoL data for this population are collected using EQ-5D-EL and EORTC as part of the VOYAGER study comparing avapritinib versus regorafenib.	Due to uncertainty in HRQoL estimates in this patient population, these data may provide some useful information, and hence we ran a scenario analysis with VOYAGER utility estimates.	***VOYAGER scenario:*** • PFS 3L: 0.782 • PD: 0.727	MEDIUM
	General population utility by age and gender from Ara and Brazier (2010).	The ERG agree with this assumption		
Costs and resource use				
Resources used for estimating the health state costs	All patients with progressed disease continue to receive investigations.	Patients with progressive disease on palliative care may have fewer investigations (e.g. scans and blood tests may not be necessary). Patients receiving palliative care may transfer from hospital to hospice, so would be followed up less intensively.	***ERG base case:*** Two-thirds would continue to have regular follow-up investigations and around one-third would not.	LOW
1L: first-line; 2L: second-line; 3L: third-line				

6 EVIDENCE REVIEW GROUP'S ADDITIONAL ANALYSES

6.1 Exploratory and sensitivity analyses undertaken by the ERG

Based on the ERG critique of the company's model assumptions (as outlined in Table 32 above), we performed a range of additional scenario analyses (shown in Table 33 below) on the following model assumptions:

- Varying the proportion of patients receiving 1L, 2L and 3L treatments in the ECM arm;
- Same dose intensity for all the TKIs as avapritinib;
- Varying the duration of treatment waning for avapritinib;
- Extrapolating OS, PFS and ToT using different survival distributions;
- Assuming ToT for avapritinib arm as similar to PFS (i.e. same assumption as used in the ECM arm);
- Alternative sources for utilities and resource use; and
- Using updated all-cause mortality data (i.e ONS 2016-2018)

The scenarios analyses were performed on the ERG corrected company's model. We note:

- For the scenarios, the ICERs range from [redacted] (Scenario: ERG resource used for progressed disease) to [redacted] (Scenario: Duration of treatment waning at 1 month). The ICERs for avapritinib versus ECM remain above £50,000 except for one scenario (Scenario: Change in resource use for progressed disease).
- Assuming a duration of treatment waning of 1 month had the greatest impact on the cost-effectiveness results. The ICER for avapritinib versus ECM increased to [redacted] per QALY. Using PFS as a proxy for time on treatment for avapritinib has a significant impact on the cost-effectiveness results; the ICER for avapritinib versus ECM increased to [redacted] per QALY.
- Another scenario that significantly influences the cost effectiveness results is the extrapolation of overall survival using the exponential distribution for both treatment arms to [redacted] per QALY.
- Using utility values from the VOYAGER study did not have a significant impact on the cost effectiveness results; the ICER for avapritinib versus ECM decreased minimally to [redacted] per QALY compared to [redacted] per QALY in the base case.

Table 33 Additional analyses conducted by the ERG on the company's base case (ERG corrected)

Scenario	Treatment	Total costs	Total QALYs	ICER (£/QALY)
Corrected company base case	Avapritinib	[redacted]	[redacted]	
	ECM	[redacted]	[redacted]	£50,033
ECM: Proportion of patients receiving 1L (20%); 2L (10%) and 3L (10%)	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
ECM: Proportion of patients receiving 1L (0%); 2L (0%) and 3L (0%)	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Dose intensity: same for all the treatments at [redacted]	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Duration of treatment waning: 1 month	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Duration of treatment waning: 6 months	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Duration of treatment waning: 12 months	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Duration of treatment waning: 24 months	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Duration of treatment waning: 36 months	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Duration of treatment waning: 48 months	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Extrapolation of OS: Avapritinib (Weibull); ECM (Weibull)	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Extrapolation of OS: Avapritinib (Exponential); ECM (Exponential)	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Extrapolation of PFS for ECM: 1L (Weibull); 2L (Weibull); 3L (Weibull)	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Time on treatment for avapritinib: Using PFS as proxy for ToT	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Extrapolating time on treatment for avapritinib using Weibull	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
All cause mortality using ONS 2016-2018	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Utility for PFS 1L 0.822	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Utility for PD 0.577	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Utility from VOYAGER trial (PFS 3L: 0.782; PD: 0.727)	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Resources used for PD state based on ERG clinical advice	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]

6.2 ERG's preferred assumptions

Based on the ERG critique of the company's economic model discussed in section 5.3.4**Error! Reference source not found.**, we have identified seven key aspects of the company base case with which we disagree. Our preferred model assumptions are discussed below:

1. ***Proportion of patients receiving 1st line, 2nd line and 3rd line treatments in the ECM arm:*** Advice from our clinical experts suggest that due to lack of effectiveness and risk of toxicity, GIST patients with the *PDGFRA* D842V mutation would not usually receive TKIs. Therefore, we assume in our base case that only 20% of the patients in the ECM arm would receive first-line imatinib treatment; followed by 10% of patients receiving second-line sunitinib and third-line regorafenib treatments respectively (for further discussion, see section 4.2.4).
2. ***Dose intensity****:* We assume similar dose intensity for all the TKIs in the ECM arm to that of avapritinib, i.e. [redacted] (further discussion in section 4.2.8.1).
3. ***Duration of treatment waning:*** Advice from our clinical experts is that patients are not likely to have persistence of clinical benefit for avapritinib for 5 years (section 4.2.6). The ERG assume a treatment waning period of 1 month as our preferred assumption (for further discussion, see section 4.2.6).
4. ***Extrapolation of survival curves:*** For consistency between treatment arms we prefer the Weibull distribution for the avapritinib arm (same as that of the ECM arm) to estimate OS, aligning with the NICE DSU guideline.[44] Similarly, for the PFS we use Weibull distribution for the second and third lines of ECM as this provides a better fit to the observed data and is consistent with first-line PFS (for further discussion, see section 4.2.6).
5. ***Time on treatment:*** We consider that the company's approach to fit parametric curves to the Kaplan-Meier data censored for deaths does not provide a close fit to the Kaplan-Meier data. For consistency with the ECM arm, we use PFS as a proxy for time on treatment for avapritinib (for further explanation, see section 4.2.6). For our preferred base case, we use a Weibull distribution for ToT extrapolation for this arm. This is consistent with our preferred distribution for PFS (for further discussion, see section 4.2.6)
6. ***Utility:*** For the ERG preferred base case, we assume that the utility value for the first line health state (AVA/1L) is same as that of the general UK population for this age-group, which is 0.822. We agree with the company's estimates for the remaining health states (for further discussion, see section 4.2.7)
7. ***Resource use:*** To reflect clinical practice, we assume that two-thirds of the patients in the progressed health state would continue to have regular follow-up investigations (i.e. CT scans; MRI scan; full blood count; and liver function test) and about one-third would not (for further discussion, see section 4.2.8)

In addition to the above key issues, for completeness we have also updated the model with the latest all-cause mortality data available from ONS 2016-2018 estimates.

Results from the ERG preferred assumptions

We show the cumulative impact of applying the ERG preferred assumptions to the corrected company's base case in Table 34. Incorporating the ERG assumptions has a significant impact on the overall ICER for avapritinib versus ECM, increasing the ICER from [redacted] per QALY to [redacted] per QALY. We observe that:

- The change that has the biggest impact on the cost-effectiveness results is the assumption that treatment waning is for 1 month. Using PFS as proxy for time on treatment for avapritinib and using general population utility value for first-line PFS also cause an increase in the ICER for avapritinib versus ECM.
- Incorporating the remaining of the ERG assumptions influence the ICER for avapritinib versus ECM, but to a lesser extent.

Table 34 Cumulative cost-effectiveness results for ERG's preferred model assumptions

Parameter	Treatment	Total costs	Total QALYs	ICER (£/QALY)
Company base case (ERG corrected)	Avapritinib	[redacted]	[redacted]	
	ECM	[redacted]	[redacted]	£50,033
+ ECM: Proportion of patients receiving 1L (20%); 2L (10%) and 3L (10%)	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
+ Dose intensity: same for all the treatments at [redacted]	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
+ Duration of treatment waning: 1 month	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
+ Extrapolation of OS: Avapritinib (Weibull); ECM (Weibull)	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
+ Extrapolation of PFS for ECM: 1L (Weibull); 2L (Weibull); 3L (Weibull)	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
+ Time on treatment for avapritinib: same as PFS	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
+ Extrapolating time on treatment for avapritinib using Weibull	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
+ All-cause mortality using ONS 2016-2018	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
+ ERG preferred utilities (PFS 1L: 0.822; PFS 2L: 0.781; PFS 3L: 0767; PD: 0.647)	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
+ Resources used for PD state based on ERG clinical advice	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
ERG preferred base case	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]

Incorporating the ERG preferred assumptions lower the OS estimates for the avapritinib arm significantly, compared to the estimates obtained from the company's base case (see Table

35). Based on the ERG assumptions, the overall survival of GIST patients with PDGFRA D842V mutation at 1 year is [redacted]; and[redacted] at 5 years. The OS estimates for the ECM arm are unchanged from the company's base case as we agree with the company's assumptions in relation to the estimation of the ECM survival estimates (as discussed previously in section 4.2.6).

Table 35 Comparison of the OS estimates between company's base case and ERG base case

Time	OS from Company's base case		OS from ERG base case	
	Avapritinib	ECM	Avapritinib	ECM
1 year	[redacted]	45%	[redacted]	45%
2 years	[redacted]	29%	[redacted]	29%
3 years	[redacted]	20%	[redacted]	20%
4 years	[redacted]	14%	[redacted]	14%
5 years	[redacted]	11%	[redacted]	11%

We present a comparison of the Markov traces for the ERG base case and company's base case showing the proportion of the cohort in each health state over time in Figure 8 and Figure 9

Figure 9. The proportions of patients in the treatment state is lower in the ERG base case compared with the corrected company's base case.

i) Company base case

[redacted]

ii) ERG preferred base case

[redacted]

Figure 8 Comparison of Markov traces for avapritinib: proportion of cohort in each health state over time

i) Company base case

[redacted]

ii) ERG preferred base case

[redacted]

Figure 9 Comparison of Markov traces for ECM: proportion of cohort in each health state over time

6.3 Scenario analyses conducted on the ERG's preferred assumptions

We performed a range of scenario analyses on the ERG base case, as shown in Table 36. Briefly, we conducted these analyses to assess the impact of changing the following model assumptions on the overall cost effectiveness results. Most of these scenarios are replicated from the company's scenario analyses (as previously outlined in section 5.2.2)

- Varying patients' initial age;
- Different model time horizons;
- Variations in duration of treatment waning for avapritinib;
- Inclusion of drug costs of the additional TKIs in BLU-285-1002 study, which are not currently approved for the treatment of GIST patients in England and Wales;
- Variation in the percentage of incomplete loss of treatment benefit after discontinuation for the avapritinib arm;
- Variation in the post-progression rate for the avapritinib arm;
- Using alternative sources to inform model parameters such as End of Life costs (i.e. Round et al); resource use (based on clinical survey); and utilities (VOYAGER and previous NICE TA);
- Using the Cassier study as source for comparator clinical effectiveness; and
- Assigning different survival distributions to extrapolate OS and PFS.

We note:

- The ICERs for avapritinib versus ECM range from [redacted] per QALY (Scenario: treatment waning period of 24 months) to [redacted] per QALY (Scenario: extrapolating the OS curves using exponential distribution) with the ICER above £50,000 per QALY for all the scenarios.
- The scenarios that have the greatest impact on the cost-effectiveness results are using a shorter time horizon of 6 years (ICER of [redacted] per QALY) and extrapolating the OS curves using the exponential distribution (ICER of [redacted] per QALY);
- Duration of treatment effect influences the cost-effectiveness results. For example, assuming a treatment waning period of 6 months reduces the ICER significantly to [redacted] per QALY; [redacted] per QALY for 12 months and [redacted] per QALY for 24 months, respectively;
- Using the Cassier study to inform survival for the ECM arm increases the ICER for avapritinib versus ECM to [redacted] per QALY; an increase of [redacted] from the ERG base case ICER;

- The remaining scenarios (i.e. changing patient's age, inclusion of drug costs of the additional TKIs in BLU-285-1002 study, extrapolation of PFS using exponential and incorporating utility estimates from the VOYAGER study and previous NICE TA for sunitinb have a lesser impact on the ICER for avapritinib versus ECM.

Table 36 Scenario analyses using the ERG's preferred model assumptions

Scenario	Treatment	Total costs	Total QALYs	ICER (£/QALY)
ERG preferred model	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Initial age: 50 years	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Initial age: 70 years	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
No discounting	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Time horizon 6 years	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Time horizon 10 years	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Treatment waning: 6 months	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Treatment waning: 12 months	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Treatment waning: 24 months	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
ECM TKIs from BLU-285-1002	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Incomplete loss of treatment benefit after discontinuation for avapritinib arm: 10%	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Incomplete loss of treatment benefit after discontinuation for avapritinib arm: 20%	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Post-avapritinib progression rate slower: 10%	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Post-avapritinib progression rate slower: 20%	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
End of life costs from Round et al	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
PD utility from Sunitinib TA	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Palliative surgery, radiotherapy, hospitalisations from clinical survey	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
Cassier et al survival for comparator arm	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
OS Avapritinib: log logistic	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
OS Avapritinib and ECM: exponential	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]
PFS for avapritinib and ECM: exponential	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]

Scenario	Treatment	Total costs	Total QALYs	ICER (£/QALY)
Utility from VOYAGER trial (PFS 3L: 0.782; PD: 0.727)	Avapritinib	[redacted]	[redacted]	[redacted]
	ECM	[redacted]	[redacted]	[redacted]

6.4 ERG conclusions on cost effectiveness

The key issues in the cost effectiveness evidence are:

- company modelled outcomes provide a poor fit to observed OS Kaplan-Meier data for avapritinib,
- company modelled outcomes do not provide a close fit to observed ToT Kaplan-Meier data for avapritinib,
- health utility values for first-line treatment therapy appear to implausible,
- the economic model assumes that patients in the ECM arm all have first-line imatinib, second-line sunitinib and third-line regorafenib treatment which clinical experts advising the ERG said was not reflective of clinical practice,
- the survival models used, for OS and ToT, differed between treatment arms.

The ERG also disagree with the company with the dose intensity used for the comparator TKIs and the proportion of patients receiving investigations in the progressed health state, however these are minor issues.

The ERG base case, which corrects the aforementioned issues, with our preferred assumptions increased the ICER for avapritinib versus ECM to £[redacted] per QALY gained. The results were most sensitive to changes in the time horizon, the treatment waning duration and the survival models used for OS.

7 END OF LIFE

The CS contends that avapritinib should be considered as an end-of-life therapy. The evidence for this is presented in CS Table 30 (reproduced below in
Table 37). With ECM, life expectancy is expected to be below 24 months. Patients have a median OS as low as [redacted] months. In the base case for the economic model, mean OS for patients treated with ECM is 23.72 months.

Median OS was [redacted] in the NAVIGATOR study as the OS data are not yet mature. The economic model presented in CS section **Error! Reference source not found.**. shows that avapritinib would provide an additional [redacted] life-years for patients with unresectable or

metastatic GIST with the *PDGFRA* D842V mutation, compared with ECM (CS Appendix J.2.). We do not agree with all the assumptions made for the modelling of avapritinib. Nevertheless, even with the ERG's suggested changes, the additional life years for patients treated with avapritinib would be considerably more than an additional 3 months.

On the basis of the evidence presented in the CS, the ERG agree that avapritinib meets the requirements set by NICE to be considered as an end-of-life therapy.

Table 37 End-of-life criteria

Criterion	Data available	Reference in CS (section and page number)
The treatment is indicated for patients with a short life expectancy, normally less than 24 months	**Mean survival:** 23.72 months **Median survival estimates:** BLU-285-1002 IPW-adjusted: [redacted] months Cassier et al., 2012: 14.7 months **24 month survival estimates:** BLU-285-1002 IPW-adjusted: [redacted] % Cassier et al., 2012: NR[25]	**BLU-285-1002** Section **Error! Reference source not found.**; Page 54 and Section **Error! Reference source not found.**, Page 136 **Cassier et al., 2012** Appendix D.1; Pages 35–36
There is sufficient evidence to indicate the treatment offers an extension to life, normally of at least an additional 3 months, compared with current NHS treatment	**Median survival estimate:** [redacted] **24 month survival estimate:** [redacted] % **42 month survival estimate:** [redacted] % **Incremental LY gains:** [redacted]	Section **Error! Reference source not found.**, Pages 38–40 and Appendix J.2, Page 89
Source: Reproduced from CS Table 30 LY, life year		

8 References

1. Blueprint Medicines Corporation. Clinical survey, 2020.
2. Rossi CR, Mocellin S, Mencarelli R, et al. Gastrointestinal stromal tumors: From a surgical to a molecular approach. *International Journal of Cancer* 2003;107(2):171-76.
3. Starczewska Amelio JM, Cid Ruzafa J, Desai K, et al. Prevalence of gastrointestinal stromal tumour (GIST) in the United Kingdom at different therapeutic lines: an epidemiologic model. *BMC Cancer* 2014;14:364.
4. Nilsson B, Bumming P, Meis-Kindblom JM, et al. Gastrointestinal stromal tumors: the incidence, prevalence, clinical course, and prognostication in the preimatinib mesylate era--a population-based study in western Sweden. *Cancer* 2005;103(4):821-9.
5. Gatta G, van der Zwan JM, Casali PG, et al. Rare cancers are not so rare: the rare cancer burden in Europe. *Eur J Cancer* 2011;47(17):2493-511.
6. Judson I, Bulusu R, Seddon B, et al. UK clinical practice guidelines for the management of gastrointestinal stromal tumours (GIST). *Clin Sarcoma Res* 2017;7:6.
7. European Medicines Agency (EMA). *Avapritinib Orphan Designation EU/3/17/1889*. https://www.ema.europa.eu/en/medicines/human/orphan-designations/eu3171889 (accessed 11 March 2020).
8. Food and Drug Administration (FDA). *FDA approves avapritinib for gastrointestinal stromal tumor with a rare mutation*. https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-avapritinib-gastrointestinal-stromal-tumor-rare-mutation (accessed 13 March 2020).
9. Casali PG, Abecassis N, Aro HT, et al. Gastrointestinal stromal tumours: ESMO-EURACAN Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2018;29(Suppl 4):iv267.
10. von Mehren M, Randall RL, Benjamin RS, et al. Soft Tissue Sarcoma, Version 2.2018: Clinical Practice Guidelines in Oncology. *JNCCN* 2018;16(5):536-63.
11. National Institute for Health and Care Excellence (NICE). *Imatinib for the treatment of unresectable and/or metastatic gastro-intestinal stromal tumours (TA86)*. https://www.nice.org.uk/guidance/TA86 (accessed 05 March 2020).
12. National Institute for Health and Care Excellence (NICE). *Sunitinib for the treatment of gastrointestinal stromal tumours*. https://www.nice.org.uk/guidance/ta179 (accessed 05 March 2020).
13. National Institute for Health and Care Excellence (NICE). *Regorafenib for previously treated unresectable or metastatic gastrointestinal stromal tumours*. https://www.nice.org.uk/guidance/ta488 (accessed 05 March 2020).
14. Farag S, Somaiah N, Choi H, et al. Clinical characteristics and treatment outcome in a large multicentre observational cohort of PDGFRA exon 18 mutated gastrointestinal stromal tumour patients. *Eur J Cancer* 2017;76:76-83.
15. Heinrich MC, Corless CL, Demetri GD, et al. Kinase mutations and imatinib response in patients with metastatic gastrointestinal stromal tumor. *J Clin Oncol* 2003;21(23):4342-49.
16. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998;52(6):377-84.
17. Blueprint Medicines Corporation. (ed.) *BLU-285-1101 (Avapritinib). A Phase I study of BLU-285 in patients with gastrointestinal stromal tumors (GIST) and other relapsed and refractory solid tumors*. Data on File; 2019.
18. George S, Bauer S, Jones RL, et al. Correlation of ctDNA and response in patients (pts) with PDGFRa D842 GIST treated with avapritinib. *Ann. Oncol.* 2018;29:viii584.
19. Heinrich M, Jones R, Schoffski P, et al. Preliminary safety and activity in a first-in-human phase 1 study of BLU-285, a potent, highly-selective inhibitor of KIT and PDGFRa

activation loop mutants in advanced gastrointestinal stromal tumor (GIST). *Eur J Cancer* 2016;69:S4.
20. Heinrich MC, Jones RL, Mehren MV, et al. GIST: Imatinib and beyond-clinical activity of BLU-285 in advanced gastrointestinal stromal tumor (GIST). *J Clin Oncol.* 2017;35(15).
21. Heinrich MC, Jones RL, von Mehren M, et al. Clinical activity of avapritinib in≥ fourth-line (4L+) and PDGFRA Exon 18 gastrointestinal stromal tumors (GIST). *J. Clin. Oncol* 2019;37(15_suppl):11022.
22. Blueprint Medicines Corporation (ed.) *A retrospective natural history study of patients with PDGFRα D842 mutant Gastrointestinal stromal tumor (GIST) previously treated with a kinase Inhibitor*. Data on file; 2018.
23. Mehren MV, Heinrich MC, Shi H, et al. A retrospective natural history study of patients (pts) with PDGFR? D842V mutant advanced gastrointestinal stromal tumor (GIST) previously treated with a tyrosine kinase inhibitor (TKI). *J Clin Oncol.* 2018;36(15).
24. Biron P, Cassier PA, Fumagalli E, et al. Outcome of patients (pts) with PDGFRAD842V mutant gastrointestinal stromal tumor (GIST) treated with imatinib (IM) for advanced disease. *J Clin Oncol.* 2010;28(15).
25. Cassier PA, Fumagalli E, Rutkowski P, et al. Outcome of patients with platelet-derived growth factor receptor alpha-mutated gastrointestinal stromal tumors in the tyrosine kinase inhibitor era. *Clin Cancer Res* 2012;18(16):4458-64.
26. Rutkowski P, Bylina E, Klimczak A, et al. The outcome and predictive factors of sunitinib therapy in advanced gastrointestinal stromal tumors (GIST) after imatinib failure - one institution study. *BMC Cancer* 2012;12.
27. Yoo C, Ryu MH, Jo J, et al. Efficacy of Imatinib in Patients with Platelet-Derived Growth Factor Receptor Alpha-Mutated Gastrointestinal Stromal Tumors. *Cancer Res Treat* 2016;48(2):546-52.
28. Osuch C, Rutkowski P, Brzuszkiewicz K, et al. The outcome of targeted therapy in advanced gastrointestinal stromal tumors (GIST) with non-exon 11 KIT mutations. *Pol Przegl Chir* 2014;86(7):325-32.
29. Demetri GD, Mehren Mv, Blanke CD, et al. Efficacy and safety of imatinib mesylate in advanced gastrointestinal stromal tumors. *N Engl J Med.* 2002;347(7):472-80.
30. Center for Drug Evaluation and Research FaDAAF. NDA/BLA multi-disciplinary review and evaluation NDA 212608 AYVAKIT (avapritinib): CDER, FDA, 2019.
31. Blueprint Medicines Corporation. Indirect comparison for avapritinib in the treatment of gastrointestinal stromal tumors (GIST) with PDGFRA D842V mutation, 2020.
32. Deeks JJ, Dinnes J, D'Amico R, et al. Evaluating non-randomised intervention studies. Chapter 4: Evaluation of checklists and scales for assessing quality of non-randomised studies. *Health Technol Assess* 2003;7 (27):23-86.
33. Roever L. Endpoints in clinical trials: advantages and limitations. *Evidence Based Medicine and Practice* 2016;1:e111.
34. U.S. Department of Health and Human Services, Food and Drug Administration, Oncology Center of Excellence, et al. Clinical trial endpoints for the approval of cancer drugs and biologics: guidance for industry 2018.
35. Klein JP, and, Moeschberger ML. SURVIVAL ANALYSIS: Techniques for Censored and Truncated Data *Statistics for Biology and Health*. Second Edition ed: Springer; 2003.
36. Oxnard GR, Wilcox KH, Gonen M, et al. Response Rate as a Regulatory End Point in Single-Arm Studies of Advanced Solid Tumors. *JAMA Oncology* 2016;2(6):772-79.
37. Food and Drug Administration (FDA). *Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics: Guidance for Industry*. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-trial-endpoints-approval-cancer-drugs-and-biologics (accessed 9 April 2020).
38. European Medicines Agency (EMA). *Guideline on the evaluation of anticancer medicinal products in man*. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-evaluation-anticancer-medicinal-products-man-revision-5_en.pdf (accessed 7 February 2020).

39. National Institute for Health and Care Excellence (NICE). *TA209: Final appraisal determination*. https://www.nice.org.uk/guidance/ta209/documents/gastrointestinal-stromal-tumours-unresectablemetastatic-imatinib-review-final-appraisal-determination3 (accessed 05 March 2020).
40. Centanni M, Friberg LE. Model-Based Biomarker Selection for Dose Individualization of Tyrosine-Kinase Inhibitors. *Front Pharmacol* 2020;11:316.
41. Blueprint Medicines Corporation. GIST Advisory Board, 2020.
42. Joseph CP, Abaricia SN, Angelis MA, et al. Avapritinib for the treatment of GIST: Analysis of efficacy, safety and patient management strategies at the recommended Phase 2 dose. *Connective Tissue Oncology Society Annual Meeting*. Tokyo: Japan, 2019.
43. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. *NICE, London* 2013.
44. Latimer N. *NICE DSU Technical Support Document 14: Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data.* http://nicedsu.org.uk/wp-content/uploads/2016/03/NICE-DSU-TSD-Survival-analysis.updated-March-2013.v2.pdf (accessed 05 March 2020).
45. Wilson J, Connock M, Song F, et al. Imatinib for the treatment of patients with unresectable and/or metastatic gastrointestinal stromal tumours: systematic review and economic evaluation. *Health Technol Assess* 2005;9(25):1-142.
46. Zuidema S, Desar IME, Erp NPv, et al. Optimizing the dose in patients treated with imatinib as first line treatment for gastrointestinal stromal tumours: A cost-effectiveness study. *British Journal of Clinical Pharmacology* 2019;85(9):1994-2001.
47. Demetri GD, Oosterom ATv, Garrett CR, et al. Efficacy and safety of sunitinib in patients with advanced gastrointestinal stromal tumour after failure of imatinib: a randomised controlled trial. *Lancet* 2006 368(9544):1329-38.
48. Poole CD, Connolly MP, Chang J, et al. Health utility of patients with advanced gastrointestinal stromal tumors (GIST) after failure of imatinib and sunitinib: findings from GRID, a randomized, double-blind, placebo-controlled phase III study of regorafenib versus placebo. *Gastric Cancer.* 2015;18(3):627-34.
49. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. *Value Health* 2010;13(5):509-18.
50. Doyle S, Lloyd A, Walker M. Health state utility scores in advanced non-small cell lung cancer. *Lung Cancer* 2008;62(3):374-80.
51. National Institute for Health and Care Excellence (NICE). *Cetuximab and panitumumab for previously untreated metastatic colorectal cancer*. https://www.nice.org.uk/Guidance/ta439 (accessed 05 March 2020).
52. Swinburn P, Lloyd A, Nathan P, et al. Elicitation of health state utilities in metastatic renal cell carcinoma. *Curr Med Res Opin* 2010;26(5):1091-6.
53. National Institute for Health and Care Excellence. Dexamethasone intravitreal implant for treating diabetic macular oedema TA349. 2015.
54. Freeman K, Connock M, Cummins E, et al. Fluorouracil plasma monitoring: systematic review and economic evaluation of the My5-FU assay for guiding dose adjustment in patients receiving fluorouracil chemotherapy by continuous infusion. *Health Technol Assess.* 2015;19(91):1-321.
55. Ara RW, A.J. *NICE DSU Technical Support Document 12: The use of health state utility values in decision models*.
56. Schöffski P, Wozniak A, Schöffski O, et al. Overcoming cost implications of mutational analysis in patients with gastrointestinal stromal tumors: a pragmatic approach. *Oncol Res Treat* 2016;39(12):811-16.
57. Monthly Index of Medical Specialties (MIMS). *MIMS Drug Database*. https://www.mims.co.uk/drugs/cancer/antineoplastics/sutent (accessed 27 March 2020).
58. Department of Health. NHS reference costs 2018-2019. *https://improvement.nhs.uk/resources/national-cost-collection/* 2020.

59. Curtis LB, A. Unit Costs of Health and Social Care 2019. https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2019/ 2019.
60. Abel J, Pring A, Rich A, et al. The impact of advance care planning of place of death, a hospice retrospective cohort study. *BMJ Support Palliat Care* 2013;3(2):168-73.

9 Appendices

Appendix 1 Therapies received by patients in the Cassier study

Therapy received (overall study population) [a]	Accumulated number of TKIs	Number of patients
First line (N=10)		
Imatinib	1	10
Second line (N=32)		
Imatinib dose increase	1	14
Sunitinib	2	11
Imatinib + sunitinib	2	1
Motesanib	2	1
Non-TKI therapy	1	5
Third line (N=16)		
Sunitinib after imatinib [b]	2	7
Sorafenib	3	3
Nilotinib	3	2
Imatinib after sunitinib	2	1
Imatinib + sirolimus	Unclear whether 1 or 2 [c]	1
Non-TKI therapy	Unclear whether 1 or 2 [c]	2

Source: Text in Cassier et al 2012.

[a] not reported separately for the *PDGFRA* D842V mutation subgroup

[b] ERG assumes this means second-line patients received increased-dose imatinib

[c] Not reported whether these patients had received sunitinib or increased-dose imatinib second line

Total number of TKIs	Number (%) of patients (N=58)
1	29 (50)
2	21 (36)
1 or 2 (unclear)	3 (5)
3	5 (9)

Appendix 2 Company and ERG assessments of study validity (questions 14-26 of the Downs and Black checklist)

The following table provides a comparison of the company's and ERG's assessments of the NAVIGATOR, BLU-285-1002 and Cassier studies for questions 14 to 26 of the Downs and Black checklist.[16] We encountered several problems whilst applying the checklist to the included studies, which are summarised below the table.

Question (as worded in the checklist and CS)	ERG interpretation (risk of bias: 'No'= high, 'Yes'=low) [a]	NAVIGATOR		BLU-285-1002		Cassier et al 2012	
		Company	ERG	Company	ERG	Company	ERG
14. Was an attempt made to blind study subjects to the intervention they have received?	'No' means lack of blinding could have introduced bias	No	Not applicable [b]	Not applicable	Not applicable [b]	No	Not applicable [b]
15. Was an attempt made to blind those measuring the main outcomes of the intervention?	'No' means lack of blinding could have introduced bias	No	Not applicable [b]	Not applicable	Not applicable [b]	No	Not applicable [b]
16. If any of the results of the study were based on 'data dredging', was this made clear?	'No' means there were unplanned analyses which could have introduced bias	No	Yes [c]	No	Yes [d]	Yes	Yes [d]
17. In trials and cohort studies, do the analyses adjust for different lengths of follow-up of patients?	'No' means imbalances in follow-up were present which could have introduced bias	Yes	Yes [e]	No	Yes [e]	Yes	Yes [e]
18. Were the statistical tests used to assess the main outcomes appropriate?	'No' means inappropriate statistical tests could have introduced bias	Yes	Yes	Yes	Not applicable [f]	Yes	Not applicable [f]

19. Was compliance with the intervention(s) reliable?	'No' means compliance was inadequate, which could have introduced bias	Yes	Yes [g]	No	Yes [g]	No	Yes [g]
20. Were the main outcome measures used accurate (valid and reliable)?	'No' means there were problems with outcomes which could have introduced bias	Yes	Yes	Yes	Yes	Yes	Yes
21. Were the patients in different intervention groups (trials and cohort studies) recruited from the same population?	'No' means different intervention subgroups (e.g. dose, mutation) were not recruited from the same population, which could have introduced bias	No	Yes	No	Yes	Yes	Yes
22. Were study subjects in different intervention groups (trials and cohort studies) recruited over the same period of time?	'No' means different intervention subgroups (e.g. dose, mutation) were not recruited at the same time, which could have introduced bias	No	Yes	Yes	Yes	Yes	Yes
23. Were study subjects randomized to intervention groups?	This question is not applicable to single-group studies	Not applicable	Not applicable	Not applicable	Not applicable	No	Not applicable
24. Was the randomized intervention assignment concealed from both patients and health care staff until recruitment was complete and irrevocable?	This question is not applicable to single-group studies	Not applicable	Not applicable	Not applicable	Not applicable	No	Not applicable
25. Was there adequate adjustment for confounding in	'No' means variables other than the planned	No	Primary outcome	No	No [i]	Yes	No [i]

the analyses from which the main findings were drawn?	intervention could have explained the results		Yes; other outcomes not reported [h]				
26. Were losses of patients to follow-up considered?	'No' means there were missing data that were not accounted for which could have introduced bias	Yes	Yes [j]	Yes	Yes [j]	No	Yes [j]

[a] the company do not explain how they interpreted the questions and do not provide any explanations for their judgements
[b] single-group study in which blinding would not be feasible
[c] dose-group comparisons were pre-specified in the statistical analysis plan
[d] no evidence of unplanned analyses
[e] KM analysis of time-to-event outcomes captures variation in follow up duration
[f] descriptive analysis without formal statistical testing (statistical tests were employed in the Cassier study but for comparisons between different mutation subgroups which are not relevant to the current appraisal)
[g] compliance appears reflective of that likely in clinical practice
[h] possible explanatory variables were included in logistic regression on the primary outcome (ORR) (CS Table 8)
[i] no exploration reported of whether factors other than the intended intervention could have explained the observed outcomes
[j] all participants were accounted for when reporting outcomes

The following problems were encountered by the ERG when assessing the company's validity assessment using the Downs and Black checklist:

- The company have not explained how they interpreted the questions in the Downs and Black checklist in relation to the included studies. We have stated the ERG's interpretation in the table above to reduce subjectivity; however, it might be that the company's interpretation was different to ours.
- The company have not provided a rationale for their yes/no answers in the checklist. Some company answers appear inconsistent but it is not possible to be sure since no explanation is given.
- Overall, the validity questions of the Downs and Black checklist do not appear to be well-suited for assessing the NAVIGATOR, BLU-285-1002 and Cassier studies since many of the questions appear to be not applicable. The studies were single-arm (so randomisation and blinding questions are not applicable) and comparative analyses were only conducted for NAVIGATOR (against a historical control)

(so several statistical analysis questions are not applicable to the BLU-285-1002 and Cassier studies). The company have not explained why the Downs and Black checklist was selected given that other tools could also have been considered for evaluating non-randomised studies.[32]

- The company compared the total checklist scores for each study (based on 1 for each yes answer and zero for each no answer for all 26 questions) (CS Appendix D.3). However, we caution that total scores should be disregarded as they conflate reporting, bias and other aspects of "quality" based on an implausible assumption that the questions about different aspects of reporting and validity have equal weight.

As stated in section 3.2.2 above, the key ERG conclusions are that the included studies are inherently at risk of bias due to their single-arm designs and, in the case of the comparator studies, their retrospective designs.

Appendix 3 ERG checklist for clinical effectiveness searches

Checklist criteria	Details	ERG comments
CS section(s)	Appendix D.1 (B.2.1 gives no details; refers to the appendix)	Not applicable
Dates covered	Databases: no limit – 05/11/2019 Conferences: most recent two years available	Over 5 months old, hence updated by ERG
Reporting	Clear outline of database sources, except not clear which websites were searched (perhaps refers to the conference proceedings). Bibliographic database search strategies are presented in CS Appendix Tables 1-3. CS Appendix Table 4 summarises results (number of hits reported does not match the numbers in CS Appendix Tables 1-3, but a minor issue as the hits reported in the PRISMA diagram match those in CS Appendix Tables 1-3). The CS Appendix tables are clearly labelled with the bibliographic database and database host. Not reported which systematic reviews and meta-analyses were reference-checked.	No major concerns
Search strategy overall	Very sensitive strategy overall: extensive use of alternative search terms/drug synonyms, broader subject headings used, searched for comparators outside of the NICE scope, reference checking of other systematic reviews (stated but no details).	No concerns
Strategy PICO and terms	Several additional pharmacological comparators were searched for, not just those in the scope, The company did not search on any terms that would express BSC, meaning that searches may have missed studies on BSC alone.	The ERG searched Medline for BSC-only studies and found none that matched the scope (studies were identified but did not separate locally advanced from unresectable/metastatic GIST.
Strategy subject headings	None missing. Exploded broader heading for 'Gastrointestinal tumors', rather than using the specific heading for 'Gastrointestinal stromal tumors'. Relied on automatic mapping of subject headings for MeSH on PubMed.	No concerns

Strategy free-text terms	Comprehensive.	No concerns
Strategy syntax	All correct.	No concerns
Strategy structure	Boolean and combinations of lines/concepts all correct.	No concerns
Sources	Searched core databases: Medline, Embase, Medline-in-process, Cochrane library – CDSR and CENTRAL 3 x general cancer conference proceedings	ERG checked for any specific sarcoma cancer conference proceedings and checked a trial registry – Clinical Trials.gov – nothing further found.
Limits	Language limit (i.e. English only) was applied at screening stage instead of at search stage.	No concerns
Filters	Published search filters not used. The concepts for study types in the strategy are well defined and include correct and relevant terms. Case studies/reports, etc., were removed (using NOT) It is appropriate to search for other clinical studies in addition to RCTs for this disease population.	No concerns
Translation	Medline and Embase searched together within Embase.com. Assume automatic mapping of subject headings. The other searches carried out are consistent across the databases.	No concerns
Missing studies	None.	No concerns

Appendix 4 Overview of outcomes assessed in the NAVIGATOR study

Primary outcomes	Outcome definition	Data cut	Specified in Decision Problem	Used in Economic Model	ERG comments
Overall Response Rate	The proportion of patients with a confirmed best response of CR or PR, where either was confirmed at a subsequent assessment without intervening progression. Standard used: mRECIST Version 1.1	Nov 2018	Yes	No	Includes discussion of Clinical Benefit Rate (a secondary outcome below) and Disease Control Rate.
Disease Control Rate	The proportion of patients with a confirmed best response of complete response (CR), partial response (PR), or stable disease.	Nov 2018	No	No	Outcome defined in text of CS in relation to ORR.
Adverse events	Type, frequency, severity, timing and relationship to the study drug of any adverse events, serious adverse events, and changes in vital signs, electrocardiogram tests and safety laboratory tests.	Nov 2018	Yes	Yes	
Time on Treatment (CS section B.2.10, Adverse reactions)	Calculated as: (treatment end date – treatment start date +1)/7 (CS Tables 22 and 23). The treatment start date was the first dose date of study drug, and the treatment end date was the last dose date of study drug or data cut-off date, whichever was earlier. Treatment duration (weeks) was summarized using descriptive statistics. See CSR section 9.7.1.10.	Jan 2020	Yes	Yes	In decision problem, but not in NICE scope. Helps to interpret the adverse events.
Secondary outcomes					
Duration of Response	The time from first documented response (CR/PR) to the date of first documented disease progression or death due to any cause, whichever occurred first. According to mRECIST 1.1. Sensitivity analysis was conducted using criteria in a study by Choi et al.	Nov 2018	Yes	No	Median DoR not reached for anticipated licensed dose. Not known to ERG if reached in Jan 2020 data cut.

Progression Free Survival	The time from the start of the treatment to the date of first documented disease progression or death due to any cause, whichever occurred first.	Jan 2020; Nov 2018	Yes	Yes	Time to event endpoints are difficult to interpret in non-randomised trials.
Clinical Benefit Rate	The proportion of patients with confirmed CR/PR/stable disease lasting four or more cycles from first dose date.	Nov 2018	No	No	See CS section B.2.6.3 – discussed as part of ORR.
Exploratory outcomes					
Overall Survival	The time from the start of treatment to the date of death. Last date known alive was defined as the last non-imputed date of any patient record prior to or on the data cut-off date in the clinical database	Jan 2020; Nov 2018	Yes	Yes	Median survival not reached.
Time to Response	The time from the start of treatment to the time the response criteria for CR/PR were first met.	Nov 2018	No	No	Not in protocol, not in the decision problem, and not used in the economic model. Included in SAP as helpful to interpret study results. ERG agree.
Tumour reduction	Central radiology assessment. Sum diameter of target lesions change from baseline. According to mRECIST Version 1.1	Nov 2018	No	No	A clinical expert advisor to the ERG noted that HRQoL-related issues, e.g. pain, bloating and fatigue, improve markedly in response to tumour shrinkage.

					However, this outcome is not in the NICE scope nor in the company's decision problem.