CONFIDENTIAL UNTIL PUBLISHED
Evidence Review Group's Report

Entrectinib for treating NTRK fusion-positive solid tumours

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

Authors Robert Hodgson, Research Fellow, CRD

Alexis Llewellyn, Research Fellow, CRD

Peter Murphy, Research Fellow, CRD

Lindsay Claxton, Research Fellow, CRD

Lucy Beresford, NIHR Training Fellow, CRD

Matthew Walton, Research Fellow, CRD

David Glynn, Research Fellow, CHE

Kath Wright, Information Service Manager

Sofia Dias, Professor of Health Technology Assessment

Correspondence to Sofia Dias, Professor of Health Technology Assessment, CRD, University of York, York YO10 5DD

Date completed Date completed (29/07/2019)

Source of funding

This report was commissioned by the NIHR HTA Programme as project number NIHR128763

Declared competing interests of the authors

None

Acknowledgements

We thank Dr Robert Metcalf, Honorary Consultant in Medical Oncology at The Christie NHS Foundation Trust, Professor Christian Ottensmeier, Chair in Experimental Cancer Medicine at University Hospital Southampton, Dr Helene Schlecht, Principal Clinical Scientist at Manchester University NHS Foundation Trust, and Marta Pereira, Clinical Scientist at Manchester University NHS Foundation Trust for their expert advice throughout this project. We also thank Professor Steve Palmer of the Centre for Health Economics, for his advice on the project.

Rider on responsibility for report

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

This report should be referenced as follows:

Hodgson R, Llewellyn A, Murphy P, Claxton L, Beresford L, Walton M, Glynn D, Wright K, Dias S. Entrectinib for treating *NTRK* fusion-positive solid tumours: A Single Technology Appraisal. CRD and CHE, University of York, Technology Assessment Group, 2019.

Contributions of authors

Robert Hodgson, Lindsay Claxton, Peter Murphy, David Glynn and Lucy Beresford wrote the cost effectiveness sections and conducted the ERG economic analyses. Alexis Llewellyn, Matthew Walton and Lucy Beresford wrote the clinical effectiveness sections of the report. Kath Wright wrote the sections on the search strategies. Robert Hodgson provided advice, commented on drafts of the report and took overall responsibility for the cost effectiveness sections. Sofia Dias provided advice, commented on drafts of the report and took overall responsibility for the report.

Note on the text

All commercial-in-confidence (CIC) data have been highlighted in [redacted], all academic-in-confidence (AIC) data are highlighted in [redacted]

Copyright statement

Table of Contents

Table of Tables

Table of Figures

List of abbreviations

ADR	Adverse events reactions
AE	Adverse events
ALK	Anaplastic lymphoma kinase
BHM	Bayesian hierarchical model
BSC	Best supportive care
CCOD	Clinical data cut-off date
CDF	Cancer Drugs Fund
CEA	Cost-effectiveness analysis
CI	Confidence interval
CNS	Central nervous system
CR	Complete response
CRC	Colorectal cancer
CrI	Credible interval
CS	Company submission
CSR	Clinical study report
DNA	Deoxyribonucleic Acid
DoR	Duration of response
DSA	Deterministic sensitivity analysis
EEA	Efficacy evaluable analysis dataset
EMA	European Medicines Agency
eMIT	Electronic market information tool
ERG	Evidence review group
ESMO	European Society for Medical Oncology
FISH	Fluorescence in situ hybridisation
HRQoL	Health related quality of Life
HTA	Health Technology Assessment
ICER	Incremental cost-effectiveness ratio
IHC	Immunohistochemistry
IPD	Individual participant data
ITT	Intention to treat
IV	Intravenous
KM	Kaplan Meier
LYG	Life years
MAIC	Matched adjusted indirect comparison
MASC	Mammary-analogue secretory cancer
NGS	Next generation sequencing

NICE	National Institute for Health and Care Excellence
NNS	Number needed to screen
NSCLC	Non-small cell lung cancer
NTRK	Neurotrophic tyrosine receptor kinase
ORR	Objective response rate
OS	Overall survival
PAS	Patient access scheme
PFS	Progression free survival
PR	Partial response
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSM	Partitioned survival model
QALY	Quality adjusted life-year
RCT	Randomised Controlled Trial
RNA	Ribonucleic acid
ROS1	Proto-oncogene tyrosine-protein kinase ROS
RT-PCR	Real-time polymerase chain reaction
SLD	Sum of longest tumour diameter
SLR	Systematic literature review
SOC	Standard of care
Trk	Tropomyosin receptor kinase
TTO	Time trade off
WGS	Whole-genome sequencing
WTP	Willingness-to-pay

1 Summary

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

Entrectinib is a potent inhibitor of tropomyosin receptor kinases A, B, and C, encoded by the neurotrophic tyrosine receptor kinase genes *NTRK1, NTRK2,* and *NTRK3,* anaplastic lymphoma kinase (*ALK*) and *ROS* proto-oncogene 1 receptor tyrosine kinase (*ROS1*). The recommended dose for entrectinib is 600 mg orally once daily for adults, and 300 mg/m^2 orally, once daily for paediatric patients who have the ability to swallow whole capsules. Entrectinib is currently awaiting European marketing authorisation.

The NICE scope reflects the anticipated licence, which presents entrectinib as a treatment option for [redacted]

The ERG found that the intervention and outcomes presented in the company submission (CS) evidence match the NICE scope. The comparators selected by the manufacturer were all therapeutic options offered in established management without entrectinib, as defined in the NICE scope. The ERG is concerned that the population presented in the evidence submitted does not match the NICE final scope. Only a small subset of tumour types known to harbour *NTRK1*/2/3 fusions were represented in the CS and only one *NTRK2* patient was included. A significant proportion ([redacted]) of trial patients received entrectinib as first line systemic therapy, including for several tumour types where the company placed entrectinib in subsequent lines of therapy. The high proportion of patients receiving entrectinib in earlier lines of therapy across tumour types may mean that survival benefits are overestimated.

1.2 Summary of clinical effectiveness evidence submitted by the company

The efficacy evidence in the CS was supported by four uncontrolled basket trials that included a total of 66 efficacy evaluable patients with metastatic or locally advanced *NTRK* fusion-positive solid tumours, including seven paediatric patients. Most of the efficacy evidence came from an *NTRK* positive subgroup of an uncontrolled phase 2 basket trial. Clinical efficacy for ten tumour types across 54 patients were included in the company's submission: sarcoma, non-small cell lung cancer (NSCLC), Mammary analogue secretory carcinoma (MASC), breast, thyroid, colorectal cancer (CRC), neuroendocrine tumours, pancreatic cancer, gynaecological cancers and cholangiocarcinoma. Following an ERG request, response data for [redacted] further patients across [redacted] were provided. Each tumour type was represented by between one and 13 patients in the whole *NTRK* population.

At the latest clinical data cut-off date (CCOD) provided [redacted], the objective response rate (ORR) was [redacted]; complete response was reported in [redacted], and partial response in [redacted]. Median duration of response was [redacted] in responders [redacted] and the Kaplan-Meier (KM) estimated median progression free survival (PFS) was [redacted]. At CCOD [redacted], [redacted] had died and the Kaplan-Meier estimated median overall survival (OS) was[redacted]. Following a request from the ERG, the company provided responder analyses as well as individual patient-level

response data for 66 *NTRK* positive patients by tumour type and line of therapy, but not for PFS and OS. The company's Kaplan-Meier curves from responder analyses showed that the OS benefit observed in responders ceased approximately at [redacted], at which point the two survival curves cross.

Health-related quality of life outcomes were reported. The safety population included 355 patients across four trials, of which 68 had an *NTRK* fusion. [redacted]. AEs leading to discontinuation of entrectinib were reported in [redacted] of the safety population.

In the absence of a control group in the trial evidence, the company adopted a pragmatic approach to identify PFS and OS comparator data for established management without entrectinib, by searching NICE pathways to identify NICE approved comparators for each of the tumour types represented in the CS efficacy evidence. Median PFS and OS from each tumour type were averaged and then pooled to calculate mean overall PFS and OS across all tumour types, weighted by the prevalence of each tumour type within the trial population.

The ERG found that the population included in the comparator trials is unlikely to match the entrectinib efficacy population, notably due to the unknown prevalence of *NTRK* fusions in most of the comparator evidence, and mismatches in the lines of therapy previously received with the treatment pathway in practice. In the base case analysis no attempts were made to adjust for differences in population characteristics between the entrectinib and comparator trial populations; comparisons were naïve and did not account for any potentially important prognostic factors. The ERG found that the methods used to identify, select and combine comparator data are inappropriate, and that the comparator data used to inform the company model is highly unreliable.

1.3 Summary of the ERG's critique of clinical effectiveness evidence submitted

Overall, the trial evidence showed a clinically meaningful overall response rate across tumour types. However, there is considerable uncertainty regarding the extent to which the response observed translates into clinically meaningful survival benefits. The ERG identified a number of important issues, particularly due to the significant immaturity of the PFS and OS data. Despite substantial censoring and the small number of patients at risk in the tails of the Kaplan-Meier curve, the crossing of OS curves between entrectinib responders and non-responders is of some concern.

The ERG were concerned that the large number of tumour types not represented in the trial, the previously discussed issues concerning trial power, and the naïve comparisons with somewhat arbitrary comparator data meant that the evidence submitted in the CS may not have allowed the company to meaningfully address the decision problem.

The ERG explored heterogeneity in response rates between the 13 tumour types included in the EEA dataset using a Bayesian hierarchical model, which assumes the response probabilities are similar

across tumour types, rather than identical (the company's preferred assumption). The ERG's analyses found that overall response rates obtained were similar to those observed when equal response probabilities are assumed, although there was considerable uncertainty in the level of heterogeneity of response rates across tumour types. Based upon this analysis, the response probability for an unrepresented tumour type could range from [redacted] Therefore, the possibility that some tumour types could have response rates that differ significantly from the pooled estimate of [redacted] cannot be excluded.

1.4 Summary of cost effectiveness submitted evidence by the company

The company's economic submission included a systematic review of published evidence on the cost-effectiveness, health-related quality of life, and resource use associated with entrectinib in the treatment of patients with *NTRK* fusion–positive solid tumours. No studies were, however, found to meet the review inclusion criteria and as such, no published evidence was identified on the cost-effectiveness, health-related quality of life, and resource use associated with entrectinib.

The CS presented a *de novo* cohort cost-effectiveness model to estimate the cost-effectiveness of entrectinib compared with established practice in a population of adult and paediatric patients with *NTRK* fusion-positive solid tumours. Established practice consisted of a composite comparator represented through a weighted average of comparators from the tumour types represented in the integrated analysis for entrectinib. Cost-effectiveness was assessed over a lifetime time horizon of 30 years with a 3.5% discount rate applied to both costs and quality adjusted life years (QALYs). No other discount rates were explored in the CS.

The model structure is based on a partitioned survival model (PSM) or "area under the curve" analysis comprising of three mutually exclusive health states: (i) PFS (progression free), (ii) progressive disease (PD; progression), and (iii) death. Within the PFS and PD health states, the model distinguished between patients who are receiving treatment and those who are not. The model predicted the total costs and QALYs separately for the entrectinib arm and the pooled comparator arm. The distribution of patients in each health state was determined by using estimates of PFS and OS.

For entrectinib, these distributions were based on KM data from the *NTRK* efficacy evaluable analysis set. In the comparator arm, estimates of mean OS and PFS for each tumour type were modelled to estimate time in each health state. These estimates of time in state were then used to estimate total costs and QALYs for each tumour type. Total costs and QALYs for the comparator arm were then estimated as weighted averages using the distribution of tumours in the integrated analysis of entrectinib.

The OS and PFS extrapolations for entrectinib were based on the integrated analysis which pooled data from three trials: ALKA, STARTRK-1, and STARTRK-2. The integrated analysis set included 54 patients across 13 different tumour types, but excluded 6 patients with primary CNS and a paediatric patient. The data-cut off used in the economic model was the 31st of May 2018, later updated to the [redacted] cut off at points for clarification. To extrapolate the observed OS and PFS data, the company fitted a number of standard parametric models. The models selected for the company's base-case analysis were extrapolated exponential OS and PFS survival functions.

Comparator OS and PFS data for each tumour type was generated from multiple NICE Technology Appraisals (TAs), which were then weighted by the distribution of tumour types in the integrated efficacy analysis. The OS and PFS data were extrapolated assuming an exponential survival function. As the company extracted only median OS and PFS values and not KM data, no other survival functions were considered.

The estimates used in the company's base-case analysis for health-related quality of life of patients in the PFS and progressive disease health states for entrectinib were based on EQ-5D-3L data collected in the STARTRK-2 study. Due to the small sample size and associated uncertainty, the post-progression utility from the integrated efficacy analysis was not used in the economic analysis. The company therefore assumed that utilities in the PD health state was equal to that of established management. The utilities used for established management were taken from the relevant NICE TAs identified in the clinical effectiveness section. The utilities for each tumour type were weighted according to the distribution of tumour types in the integrated efficacy analysis. In contrast with the approach taken for the comparator efficacy, where a range of estimates for each tumour type were pooled, the utility values extracted for each tumour type were obtained from a single selected TA.

Resource use and costs included: drug acquisition and administration costs, monitoring costs, costs related to health states and adverse events, the cost of subsequent treatments and screening costs. Patient access scheme (PAS) discounts are available for entrectinib, nintedanib, nab-paclitaxel, trifluridine/tipiracil, everolimus, eribulin and trabectedin. For the purpose of simplicity, the company grouped interventions into three classes: oral, simple intravenous (IV) and complex IV and used these costs to estimate drug administration costs as well as the progression-free health state costs based on the interventions comprising established management. For estimation of the screening costs, the company used a hierarchical approach to testing assuming immunohistochemistry (IHC) followed by next generation sequencing (NGS) for the majority of tumour types.

The company found entrectinib to be more costly (cost difference of [redacted]) and more effective ([redacted] QALYs gain) compared with established management. The deterministic base case incremental cost-effectiveness ratio (ICER) was £52,609 per QALY and the mean probabilistic ICER was £52,052 per

QALY. These results do not include PAS discounts available for nintedanib, nab-paclitaxel, trifluridine/tipiracil, everolimus, eribulin and trabectedin. The majority of the QALYs gained were generated as a result of additional life years. The company reported that the most influential parameters in the one-way sensitivity analysis included the comparator OS estimates and the screening costs.

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

The ERG highlights that there are significant number of issues that contributed to uncertainty in the cost-effectiveness results presented by the company.

The focus of the company's submission was on a single answer to indicate the cost-effectiveness of entrectinib in the population covered by the marketing authorisation. The general view of the ERG is that optimised decisions are preferable and while the ERG acknowledges the challenges presented by the current decision problem, the company could have gone further in justifying the use of a single ICER. In particular, the ERG considers that the company could have explored further the variability in the treatment effect across tumour types, as well as further considering how variability in testing costs impact on the tumour-type specific ICER. The ERG notes the possibility for heterogeneity in the treatment effect across tumour types, as well as across other clinical characteristics such as age (paediatric vs adults), fusion type and position in the treatment pathway, which were not accounted for.

The ERG has several concerns about the representativeness of the modelled population, which was based on the integrated efficacy analysis. These include concerns about the distribution of tumour types modelled, which appear to over represent some tumour types, while under-representing others. Further, the modelled population includes only the 13 tumour types represented in the trials, while there is evidence to suggest that *NTRK* fusions occur in at least another 11 tumour types representing a minimum of 20% of the eligible population. The omission of these patients has a number of implications for the model and potentially impacts upon a number of the inputs used to model established management including, comparator effectiveness, comparator treatment cost, testing costs, and health state utilities. The ERG is also concerned that the analysed integrated efficacy data set excluded available evidence on patients with primary CNS tumour as well as a number of paediatric patients,

[redacted]

There are also significant uncertainties regarding whether the appropriate comparators have been modelled. The anticipated marketing authorisation for entrectinib allows entrectinib to be used and multiple points in the treatment pathway, meaning there is significant uncertainty regarding the

patient group in which entrectinib may be used in practice. It is therefore unclear whether the modelled comparators represent current NHS practice. Further, because the model only considers 13 tumour types and not all tumour types in which *NTRK* fusions may occur, there are a number of relevant comparators not covered by the model. The model therefore implicitly assumes that the modelled population is representative of the eligible population which appears to be unlikely given available evidence on the distribution of tumour types with *NTRK* fusions.

The ERG highlights that the observed data for entrectinib was immature with median OS not yet met. As such, there is significant uncertainty regarding the longer-term survival benefits of entrectinib. The company base-case fits an exponential function to the available KM data, selected from a range of standard parametric functions on the basis that the exponential function has the best statistical fit to the observed data. The ERG considers the exponential function to represent a potentially plausible extrapolation of OS, but is concerned that it implies that post-progression survival is significantly longer than pre-progression survival. The ERG questions the clinical plausibility of this given that entrectinib therapy is discontinued on progression and that only [redacted] of patients received any subsequent therapy. The ERG's preference is therefore for the Weibull function, which produces a more reasonable balance between pre- and post-progression survival while also having good statistical fit to the observed data.

Because the available effectiveness evidence for entrectinib was from single arm studies, it was necessary to generate an appropriate comparator dataset. The company does this by using previous NICE TAs as a source of effectiveness data, which are then weighted by the distribution of tumour types in the integrated efficacy analysis. While the ERG considers the broad approach adopted by the company to be reasonable, there are significant challenges associated with implementing this approach successfully, as well as further issues resulting from the company's execution of this approach.

The ERG's principal concerns regarding the company's approach to generating a comparator is that it relies on an unadjusted naïve comparison between the weighted comparator and the integrated efficacy analysis with significant scope for confounding bias. The ERG in particular notes that a significant proportion of the patients in the integrated efficacy population (37.0%) received entrectinib as a first-line systemic therapy, while the comparator dataset draws predominantly from patients in later lines of therapy. Further, the use of NICE TAs as a source of effectiveness evidence means that comparator effectiveness data is being drawn from a population who are primarily *NTRK* fusion negative. This is problematic because there is evidence to suggest that *NTRK* fusions are prognostic, with variable impact upon prognosis depending upon tumour type.

Because of these significant concerns about confounding bias and the challenges of generating a truly comparable comparator data set, the ERG considers that the company should have also considered other approaches to generating a comparator data set to further explore the uncertainties associated with generating a comparator data set. For example, the company could have utilised two alternative methods outlined in Hatswell *et al.*[1], which would have provided alternative estimates of comparator effectiveness and could have been used to validate the company's base-case.

The ERG also has substantive concerns regarding the companies approach to modelling *NTRK* fusion testing. The ERG in particular is concerned that the company appears to have included extensive testing costs in the comparator arm of the model. The ERG considers that the focus of modelled testing costs should be on the incremental testing costs associated with identifying *NTRK* fusion positive patients.

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

1.6.1 Strengths

ORR rates were clinically meaningful and objective response was observed across all tumour types and lines of therapy included. Clinical efficacy evidence included 13 tumour types in mostly metastatic patients, a paediatric population, and several cancers expected to harbour a larger proportion of patients who may be eligible for entrectinib according to the anticipated marketing population.

1.7 Weaknesses and areas of uncertainty

The main weaknesses and areas of uncertainty identified by the ERG include:

Uncertainty surrounding the homogeneity of the treatment effect

The ERG considers the company's assumption that all tumour types will have identical response rates when treated with entrectinib to be very strong and subject to considerable uncertainty. Analyses presented by the ERG suggest that there is heterogeneity in response and that response rates in tumour types not represented in the trial data could vary considerably from what has been presented.

Uncertainty surrounding the relevant patient population

Significant uncertainties exist regarding the position of entrectinib in the patient pathway. The anticipated marketing authorisation for entrectinib allows patients to be treated when there is [redacted] is ambiguous and is likely to be influenced by subjective assessments of the response rates and adverse event burden associated with existing options.

The choice of comparator regimens

Because of the significant uncertainties surrounding the position of entrectinib in the treatment pathway, it is not clear whether the comparators considered reflect current established management in the treated population.

The uncertainty surrounding the extrapolation of OS for entrectinib

The ERG notes that significant uncertainties remain regarding the extrapolated OS estimates for entrectinib. While the ERG considers that the company's approach based on an exponential function provides reasonable estimates of long-term survival there are concerns about what this implies regarding the split between pre- and post-progression survival.

Uncertainty surrounding the costs of identifying patients the NTRK fusions

Current testing for the majority of tumour types does not routinely include testing for *NTRK* and the rarity of the *NTRK* fusions means that the number needed to screen (NNS) to identify a single *NTRK* fusion positive patient is often high. Testing costs therefore represent a substantial proportion of the incremental costs associated with implementing entrectinib.

A number of plausible testing strategies exist that could be implemented, should entrectinib be approved for use in the NHS, with a range of advantages in terms of the costs and diagnostic performance. There are also significant uncertainties around who will receive testing and when testing will be implemented across tumour types, as knowledge on the tumour types which harbour *NTRK* fusions is current incomplete.

Uncertainty surrounding broader infrastructure and training requirements

The provision of entrectinib on the NHS is likely to substantially increase the number of patients requiring molecular testing. The ERG considers that important uncertainties remain concerning whether the additional resource/cost implications for the NHS have been fully quantified. The ERG notes that particular consideration should be given to whether there are additional infrastructure or training requirements for the NHS which have not been captured.

1.8 Summary of exploratory and sensitivity analyses undertaken by the ERG

The key uncertainties addressed by the ERG scenario analyses relate to:

- The testing costs associated with the implementation of entrectinib;
- The population modelled and the distribution of eligible patients across tumour types;

- Unit costs associated with the chemotherapy regimens that constitute established management;
- Drug wastage associated with entrectinib.

Further to the above, the company presented additional analysis as part of the points for clarification response which incorporated the latest [redacted] data cut; incorporated available effectiveness evidence available for patients with primary CNS tumours as well as several paediatric patients; and made alternative assumptions about the duration of subsequent therapies received by entrectinib patients.

The results of these scenario analyses including the ERG's base-case are summarised in Table 1. Due to time constraints, deterministic ICERs are presented throughout.

The ERG alternative base-case analysis incorporated a number of alternative assumptions, a number of which were also explored by the company in scenario analyses. The changes made by the ERG include:

- Inclusion of children and primary CNS tumours in the population;
- Weibull distribution for extrapolation of entrectinib OS and PFS;
- Inclusion of marginal testing costs only;
- Confirmatory RNA-based NGS test after whole genome sequencing (WGS) test, and removal of NGS testing costs for lung cancer patients;
- Testing costs estimated using the number needed to screen based on the whole *NTRK* population;
- WGS test to identify *NTRK* tumours in paediatric patients,
- Second-line therapy following discontinuation of entrectinib, limited to 6 month duration;
- electronic market information tool (eMIT) costs for therapies in the established management arm;
- Inclusion of drug wastage of entrectinib.

Under the ERG's alternative set of assumptions, the ICER for entrectinib versus established care is £77,109 per QALY.

Table 1 Summary of ERG exploratory analyses

	Incremental costs	**Incremental QALYs**	**ICER (£/QALY)**
Base case	[redacted]	[redacted]	£52,609
Scenario 1: Alternative distribution of tumour types	[redacted]	[redacted]	£69,747
Scenario 2: Remove testing costs in established management arm	[redacted]	[redacted]	£63,329

Scenario 3: Remove lung cancer cost of testing	[redacted]	[redacted]	£59,465
Scenario 4: Confirmatory RNA-based NGS in WGS patients	[redacted]	[redacted]	£64,608
Scenario 5: Prevalence of *NTRK* fusions (tumour types represented in the trial)	[redacted]	[redacted]	£56,914
Scenario 6: Prevalence of *NTRK* fusions (based on the whole *NTRK* population)	[redacted]	[redacted]	£65,981
Scenario 7: Cumulative impact of 2, 3, 5, 7	[redacted]	[redacted]	£64,115
Scenario 8: No testing costs	[redacted]	[redacted]	£36,914
Scenario 9: WGS for identifying *NTRK* tumours in paediatric patients *	[redacted]	[redacted]	£48,860
Scenario 10: eMIT costs for therapies in the established management arm	[redacted]	[redacted]	£52,081
Scenario 11: With drug wastage	[redacted]	[redacted]	£55,357
ERG alternative base-case analysis **	[redacted]	[redacted]	**£77,109**

* These results should be compared to the analysis including primary CNS and paediatric patients, see Table 50.
** These results have been updated by the ERG following the factual accuracy check to include the change made in Scenario 9

The ERG also presented a further scenario analysis using the ERG's base assumptions in which an alternative model structure was used where PFS and OS were determined according the ORR. This method used the survival of non-responder patients to estimate survival predictions in the established management arm. The entrectinib arm was based on a weighted average of responder and non-responder survival predictions, which allowed for the exploration of cost-effectiveness in different tumour types by varying the response rate used to estimate the weighted average. The ICER for the pooled group was £95,705 per QALY. When varied by tumour type, the ICERs ranged from £57,451 per QALY for sarcoma patients, to £128,663 for thyroid tumours.

In further exploratory analysis using the response-based model, the ERG also presents an example of how a response-based model can be used estimate the value of heterogeneity and the population net health effect, so as to potentially permit optimised decisions that would limit the provision of entrectinib to those patients in which it is most cost-effective. Using the tumour type CRC as an example, an 'optimised' recommendation which excludes CRC might result in an additional 12.99 QALYs per year to the health system.

2 Background

2.1 Critique of company's description of underlying health problem

The present appraisal concerns the treatment of locally advanced or metastatic solid tumours exhibiting gene fusions involving neurotrophic tyrosine receptor kinase (*NTRK*) genes 1, 2, and 3 in any solid tumour. This is the first time a technology has been appraised for a histology-independent indication, with treatment determined by the presence of a specific type of genomic alteration, rather than the location of the tumour.

The CS describes that advances in techniques used to identify particular gene fusions have enabled the development of therapies directed specifically at the molecular targets responsible for the growth of cancer cells, and that *NTRK* gene fusions are 'clinically actionable' drivers of solid tumour formation and development across a wide variety of sites. The underlying health condition considered in this appraisal is therefore defined with respect to the presence of *NTRK* fusions and not tumour type. As such, in contrast with other NICE appraisals of cancer therapies (where the indication considered is a single tumour type), this appraisal considers any solid tumour exhibiting the *NTRK1*, *2* or *3* gene fusions.

The ERG considers the company's description of the underlying health problem to be appropriate and relevant to the decision problem under consideration. The company describes the role of the tropomysin receptor kinases (Trks) in the development and function of neurons in the central and peripheral nervous system. These receptor proteins can be expressed in a variety of tissue types and are involved in the regulation of function, proliferation, and survival of cells. *NTRK* gene fusions occur when the 3' region of *NTRK* gene is joined with the 5' sequence of a fusion partner gene by a chromosomal rearrangement event. This results in the over-production of a chimeric Trk protein which is permanently 'switched on', meaning cell survival and proliferation are decoupled from normal regulatory processes, which may lead to oncogenesis. The ATP-binding sites of the TrkA/B/C proteins share high structural similarity,[2] which entrectinib exploits to inhibit the activity of chimeric receptors to stop or reverse the growth of *NTRK* fusion-positive tumours.

The company suggests that the prognosis of patients with *NTRK*-fusion positive tumours is worse than those without this genomic alteration, and provides an example of a study in colorectal cancer patients in which shorter median overall survival (OS) is observed for patients with *NTRK*, *ALK*, or *ROS1* gene rearrangements.[3] However, this is a small study and does not report survival data by gene arrangement type, and therefore in the ERG's view cannot be considered conclusive. Furthermore, the ERG considers it more likely that the relative prognosis of patients with *NTRK* fusions will vary between cancer types, and that outlook could also plausibly vary by which of the three *NTRK* genes is

involved. This is supported by evidence from other cancer types. For example, in a study of patients with papillary thyroid cancers, prognosis was similarly found to be worse in patients with *NTRK* fusions when compared to those without,[3-5] while the presence of *NTRK* fusions in a mesoblastic nephroma patient population was associated with more favourable outcomes in another study.[6] From the evidence available, it also is unclear whether *NTRK* fusions are in themselves prognostic, or whether it is their association with other specific prognostic factors such as age and ECOG status that drives the observed differences in prognosis.

2.1.1 Prevalence of *NTRK* gene fusions

The CS estimates that *NTRK* gene fusions are present in 0.7% of all cancers, based on a weighting of literature prevalence estimates with figures observed in the entrectinib clinical trial; however, the ERG notes this is significantly higher than other figures reported in the literature sources referenced by the company. Excluding the estimate derived from the entrectinib trial, the prevalence of *NTRK* gene fusions is reported to be between 0.25% - 0.31% in the adult population[7-9] and 0.34% – 0.49% in the paediatric/adolescent population.[7, 8, 10, 11] The Foundation Medicine Inc. dataset cited by the company found [redacted]% of ~116,000 samples harboured an *NTRK* gene fusion. As the largest epidemiological study available, the ERG considers this figure the most reliable estimate of *NTRK* gene fusion prevalence.

This lower figure impacts upon the number of patients who would be eligible to receive entrectinib in clinical practice. The CS estimates that the number of patients eligible to receive entrectinib, i.e. those with *NTRK* fusion positive advanced or metastatic cancer, would be 648 per year in England. This figure is based on a number of assumptions: any cancer can harbour an *NTRK* gene fusion; fusions occur in 0.7% of all cancers; 34% of all cancers are advanced or metastatic; and that 90% of patients are fit enough for treatment. The ERG suggests this figure may be an overestimate, and provides an alternative estimate based on a different set of assumptions.

The company's population size estimate includes patients with any type of cancer, rather than just those with solid tumours as described in the anticipated marketing authorisation of entrectinib. Using the company's assumption of eligibility by cancer stage, but limiting the eligible population to solid tumours with a prevalence of *NTRK* of [redacted]%, the number of patients eligible for entrectinib in England is reduced to [redacted] individuals annually.

The ERG's estimate of patients eligible for entrectinib uses a bottom-up approach, where a total population size was calculated by using tumour-specific rates of *NTRK* gene fusions and disease incidence. The tumour types included in these calculations are presented in Table 2.

Table 2 Tumour types included in ERG population size calculations

Tumour Type	
MASC	Cervix
NSCLC (Adenocarcinoma & squamous cell carcinoma)	Soft tissue sarcoma
Breast cancer	Head and neck squamous cell carcinoma
Secretory breast carcinoma	Salivary gland (non MASC)
Papillary thyroid tumour	Sinonasal adenocarcinoma
Thyroid tumour	Gastro-oesophageal junction
Colon/colorectal	Prostate cancer
Melanoma	Renal cell carcinoma
Neuroendocrine	Low-grade glioma
Gastrointestinal stromal tumour	High grade glioma (inc. glioblastoma multiforme)
Cholangiocarcinoma	Paediatric high grade glioma
Pancreatic	Congenital mesoblastic nephroma
Appendix	Paediatric melanoma
Uterine	Infantile fibrosarcoma
Ovarian	Paediatric low grade glioma

The ERG considered it appropriate to limit the population to patients at the relevant stage of the treatment pathway for each tumour type (i.e. in line with the proposed positioning of entrectinib), thus yielding a more representative estimate of the population eligible for entrectinib in practice. Using these assumptions, the ERG estimate that 196 patients would become eligible for entrectinib every year in England. Clinical advisers to the ERG suggested that it is possible that *NTRK* fusions may present in any tumour type, so the ERG's estimate of the eligible population is likely to be conservative, as it does not account for cancers in which an *NTRK* fusion has not yet been identified.

Further details on the calculation of the size of the eligible population are presented in Appendix A.

2.2 Critique of company's overview of current service provision

2.2.1 Treatment pathways

The company states that there is currently no established treatment pathway for patients with *NTRK* fusion-positive tumours, with treatment guided by tumour type-specific care guidelines. The position at which *NTRK* fusion-positive cancer patients would be offered entrectinib is likely to vary by the availability of other effective treatments in each tumour. This is reflected in the anticipated marketing authorisation, which covers entrectinib as a treatment option for [redacted] The company's interpretation appears to position entrectinib as an alternative to standard chemotherapy when one or more other

options have been exhausted, or as a first-line option where there are no acceptable alternatives. However, what constitutes an '[redacted]' is ambiguous and may be affected by the availability of entrectinib itself. Clinical advice to the ERG suggested that 'acceptability' would be a subjective assessment of the response rates and adverse event burden associated with existing options, but the threshold at which a decision to offer entrectinib would be made is likely to vary between indications. Current availability of testing for targeted therapies in each indication is also likely to influence the positioning of entrectinib if *NTRK* fusion testing is added to existing screening processes; in those indications with early testing for other genetic oncodrivers it is likely that entrectinib will be used in place of other chemotherapy options. However, in their clarification response the company stated that they anticipate entrectinib to be used in later lines of treatment in the majority of cases, at the point where therapeutic options are very limited or exhausted altogether. The company also provided an outline of where they expect entrectinib to be offered within existing treatment algorithms for patients included in the integrated efficacy analysis, reproduced in Table 3.

Table 3 Proposed positioning of entrectinib for the treatment of *NTRK* fusion-positive, locally advanced or metastatic solid tumours (Reproduced from CS Table 6, Page 30)

Position of entrectinib in line of systemic therapy	
First-line*	**Second-line and beyond†**
MASC	NSCLC
Soft-tissue sarcoma	Breast
Pancreatic cancer	Thyroid cancer
Cholangiocarcinoma	Colorectal cancer
Gynaecological cancers	Neuroendocrine tumours

*Patients ineligible for curative surgery or radiotherapy with no immunotherapy or targeted therapy options
†Some patients may receive first-line entrectinib treatment if not eligible for targeted or immunotherapies

The ERG do not consider the company's definition of current treatment pathways and the anticipated positioning of entrectinib to sufficiently address the decision problem. Firstly, the as-yet undetermined timing of testing within each tumour type will inevitably define the eligible population. Secondly, the groupings of tumour types as presented by the company are too broad to accurately represent the diversity of cancer types and different treatment options available within each. For example, neuroendocrine and gynaecological cancers comprise numerous specific indications with differing prognoses and treatment options recommended by NICE. Furthermore, it is likely that entrectinib will be offered at different points in the respective treatment pathway of the tumour types covered by these umbrella terms.

2.2.2 *NTRK* fusion diagnostic pathways

2.2.2.1 Testing for *NTRK* gene fusions

A number of testing strategies are available for the identification of *NTRK* gene rearrangements across different tumour types. These include fluorescent in-situ hybridisation (FISH), immunohistochemistry (IHC), ribonucleic-acid (RNA)-based next generation sequencing (NGS) or reverse transcriptase polymerase chain reaction (RT-PCR).[12]

FISH testing is commonly used to detect chromosomal abnormalities such as *ALK* and *ROS1* gene rearrangements.[13, 14] FISH is used to identify a single specific gene fusion, so if a particular fusion is common in a tumour type, it can be efficient to use FISH. The NHS currently offers FISH for the detection of the highly prevalent *ETV6-NTRK3* gene fusion in MASC patients.[15] However, where the type of *NTRK* fusion is unknown, then three individual tests would need to be conducted in order to detect the presence of one of the three *NTRK* rearrangements.

Immunohistochemistry detects overexpression of Trk proteins, a subset of which may be the result of NTRK gene fusions. Unlike FISH, all three *NTRK* fusions can be tested in one IHC test by using a pan-*TRK* fusion panel. IHC is quick and inexpensive, and is currently used for a variety of gene rearrangements across tumour types in the NHS.[12]

Next generation sequencing methods that use rapid sequencing of RNA and DNA can be used to detect *NTRK* fusions. DNA-based NGS can be used to detect multiple chromosomal rearrangements from a single sample,[16] and is currently used as a diagnostic and prognostic method in oncology for a range of tumour types.[15] However, there are concerns that DNA-based NGS panels will not identify all *NTRK* fusions; for fusions where there is a large intron size, DNA-based NGS may be limited and may provide inaccurate results.[17] In research, DNA-based NGS panels to detect *NTRK* fusions have to be confirmed with RNA-sequencing or IHC.[17] RNA-based NGS can detect *NTRK* fusions independent of *NTRK* fusion type,[18] and is often seen as the 'gold standard' of testing for gene fusions if RNA quality is high.[17] NGS is substantially more resource intensive than FISH and IHC, with longer turnaround times and higher quality sample requirements.

More recently, hybrid DNA/RNA NGS panels have been developed, allowing DNA and RNA to be extracted and run simultaneously in one test.[12] The Oncomine Focus Fusion Assay, for example, screens for 161 cancer-associated gene rearrangements. Like FISH, knowledge of fusion partners is necessary in order to identify gene rearrangements. As these assays are in constant development, the addition of newly-identified mutations is relatively straightforward.[12]

2.2.2.2 Testing strategy options

There is currently no established strategy for detecting for *NTRK* fusions across tumour types in the NHS. For the tumours where WGS is currently unavailable, and with the exception of MASC, where patients receive the *ETV6-NTRK3* FISH test, *NTRK* fusions are not routinely tested for in solid tumours.

The company propose a two-tiered testing approach. First, IHC testing is used to detect the presence of an *NTRK* gene fusion. For individuals with a suspected *NTRK* fusion, confirmatory NGS will be used to identify the particular gene rearrangement. The CS does not include MASC patients and patients eligible for whole genome sequencing (paediatric tumours and sarcoma) in this testing strategy. However, the ERG was advised that current NHS WGS may not accurately detect *NTRK* fusions or other structural abnormalities, therefore confirmatory RNA-based NGS would still be required in these patients.

The company suggest the use the Roche Ventana pan-TRK assay for IHC testing, which detects Trk proteins A, B, and C, identifying any *NTRK* gene rearrangement in one test. The CS states that Ventana assay eliminates 89% of *NTRK* fusion-negative samples. However, the sensitivity of other pan-TRK IHC assays have been estimated to be as low as 55% for *NTRK3* fusions.[7] Thus, up to half of the individuals with an *NTRK3* rearrangement could incorrectly test negative. IHC also has limited predictive value in neural and smooth muscle tumours, where false positives occur due to the natural expression of Trk in these tissues.[12] The Oncomine Focus Fusion assay, a hybrid DNA/RNA assay recommended by the company, has a high specificity and sensitivity (both 100%), from the small sample available.[19]

There have been a variety of alternative testing algorithms proposed for the identification of *NTRK* fusions,[17] most of which suggest that the testing approach should vary depending on the prevalence of *NTRK* fusions, and the provision of genomic testing currently available.[12, 16]

The extent and purpose of current testing provision varies across tumour types, with some genomic and histological testing for specific genetic abnormalities already in place for specific cancer types.[20] Table 4 provides details of genomic and molecular testing currently available for tumour types with known *NTRK* fusions. There is some form of genetic or molecular testing available in the majority of tumour types with a known *NTRK* fusion. With the exception of gastrointestinal stromal tumours, NGS is not routinely provided to every patient with a particular tumour type. Eligibility for testing often depends on the histology and the sub-type of the tumour. For example, pre-surgery IHC is routinely offered for all individuals with invasive breast cancer at the time of diagnosis, and only women under 50 years old with triple negative breast cancer are eligible for *BRCA1* and *BRCA2* NGS testing.[21] The majority of NGS testing available on the NHS is currently DNA-based. There are

concerns that DNA-based NGS panels will not identify all *NTRK* fusions[17]. For the of cancer types where DNA-based NGS is currently offered,additional RNA fusion-based or RNA/DNA hybrid NGS will be required to confirm *NTRK* rearrangements.

Table 4 Current molecular and genetic testing for tumour types with *NTRK* fusions on the NHS

Tumour Type	**Frequency of *NTRK* fusion**	**Current Molecular Testing**
MASC	100.00%	FISH (*ETV6-NTRK3*)
NSCLC (Adenocarcinoma & squamous cell carcinoma)	[redacted]	IHC (22C3) Multi-target NGS panel (*EGFR*) [22]
Breast cancer	[redacted]	IHC (HER2)[21] Multi-target NGS panel: (Oncotype DX)[22]
Thyroid tumour	[redacted]	IHC for Papillary Thyroid Tumour[23] Multi-target NGS Panel (*BRAF*, *KRAS*, *NRAS*, *HRAS*)[23]
Colon/colorectal	[redacted]	IHC for Lynch Syndrome (hereditary CRC)[24] Multi-target NGS Panel (*BRAF*, *KRAS*, *NRAS*) [22]
Melanoma	[redacted]	Multi-target NGS Panel (*BRAF*, *NRAS*, *KIT*) [22]
Neuroendocrine	[redacted]	No Routine Testing Available
Gastrointestinal stromal tumour	[redacted]	IHC (*CD117*, *C134*, *DOG1*) [25] Multi-target NGS Panel (*KIT*, *PDGFRA*)[22]
Cholangiocarcinoma	[redacted]	No Routine Testing Available
Pancreatic	[redacted]	No Routine Testing Available
Appendix	[redacted]	No Routine Testing Available
Uterine	[redacted]	IHC (*EMA*, Ber-EP4, *PAX8*, *CK7*)[26] (REF) FISH (*EPC1-PHF1*)
Ovarian	[redacted]	IHC[27] Multi-target NGS panel (*BRAC1*, *BRAC2*)[22] Multi-target NGS panel (*SMARCA4*)[22]
Cervix	[redacted]	IHC[28]
Soft tissue sarcoma	[redacted]	Whole Genome Sequencing
Head and neck carcinoma	0.24%	IHC (HPV) [29] Multi-target NGS Panel – (*CDKN2A*, *EGFR*, *TP53*)[22]
Prostate cancer	[redacted]	IHC (PSA)[30]
Renal cell carcinoma	[redacted]	FISH/RT-PCR (*TFE3*)

High grade glioma (inc. glioblastoma multiforme)	[redacted]	IHC[31] Multi-target NGS (*IDH1*, *IDH2*, *ATRX*, *TERT*, *H3F3A*)[22] Multi-target NGS (*BRAF*), MGMT promotor hypermethylation[22]
Paediatric high grade glioma	5.30%	Whole Genome Sequencing
Congenital mesoblastic nephroma	60.70%	Whole Genome Sequencing
Paediatric melanoma	11.11%	Whole Genome Sequencing
Infantile fibrosarcoma	90.90%	Whole Genome Sequencing
*The frequency NTRK fusions in appendix tumours in the FMI data set was reported to be 0%, however it has been reported to be higher than 0% in the literature[2] MASC, mammary analogue secretory carcinoma; NSCLC non-small cell lung cancer		

According to the expert advice received by the ERG, the only RNA-based NGS fusion panel available NHS is for a specific subgroup of NSCLC patients, targeting a range of genes including *EGFRALK*, and *ROS1*. Whilst this panel does not currently target *NTRK1-3* rearrangements, genomic advisers informed the ERG that the costs of adding additional gene targets to an RNA-based NGS panel are nominal.

The European Society for Medical Oncology (ESMO) propose that the standard testing pathway should differ depending on the frequency of *NTRK* fusions in each tumour type, and whether sequencing is currently provided by the NHS. It is recommended that FISH, RT-PCR, or targeted NGS assays are used first line in tumour types known to have a high prevalence of *NTRK* fusions and where other NGS is already available, and IHC where it is not.[17] In those tumour types thought to have lower frequencies of *NTRK* fusions and where current genomic testing is available, ESMO recommends the use of front-line NGS, followed by confirmatory IHC. In the tumour types where there is thought to be a lower frequency of *NTRK* fusions and where there is no genomic testing available, it is suggested that IHC is used for initial screening; *NTRK* gene rearrangements are then confirmed using NGS. A similar approach, suggested by Penault-Llorca *et al.*[16] is presented in Table 5.

Table 5 Alternative screening pathways according to prevalence

Prevalence of *NTRK*	**Testing strategy**
High prevalence of *NTRK* gene fusions	FISH or IHC
5-25% prevalence of *NTRK* gene fusions	NGS panel
< 5% prevalence of *NTRK* gene fusions	NGS panel
< 5% prevalence of *NTRK* / gene fusions not common	IHC then confirmatory NGS

2.2.2.3 Feasibility of *NTRK* fusion Screening

As discussed in Section 2.1.1, there are 28 tumour types observed to harbour *NTRK* gene fusions, but the ERG's clinical advisers suggested that *NTRK* fusions could potentially occur in any tumour type. Therefore, the feasibility of screening for *NTRK* fusions in all tumour types should be considered.

Using the company's proposed diagnostic testing strategy (IHC followed by confirmatory NGS), and their top-down estimates of the annual population eligible for entrectinib (CS, Budget Impact Model), [redacted] individuals would require testing using IHC every year. If the assumptions of the diagnostic accuracy reported of IHC (89% *NTRK* fusion-negatives identified) are used, [redacted] confirmatory NGS tests would be required per year. This is likely to be an overestimate, as the population defined by the company includes patients with haematological cancers. If these patients are excluded, the estimated number of individuals that would require NGS is reduced to [redacted], with [redacted] individuals requiring confirmatory NGS per year. The company do not consider the positioning of entrectinib, or when testing would be offered, so the size of this population is still likely to be higher than what would be expected in practice. For further details on the company's assumptions and calculations, see Appendix B.

The ERG used a conservative, bottom-up approach to calculate the number requiring testing, based on the tumour types in which there is a known *NTRK* fusion (see Table 2). Using the company's proposed diagnostic algorithm (IHC followed by confirmatory NGS), the number of additional IHC tests required to identify patients meeting the anticipated marketing authorisation of entrectinib would total approximately 51,958 a year in England. Based on the diagnostic accuracy figures supplied by the company, this would mean 5,806 patients would require confirmatory NGS tests annually. For further details on the ERG's assumptions and calculations, see Appendix B.

In order to provide testing for *NTRK* gene fusions, sufficient capacity in genomic testing services is required. With the increasing number of targeted medicines available, the number of individuals requiring genetic testing is increasing. Cancer Research UK estimated that in 2014, 16,000 patients with colorectal cancer or NSCLC did not receive molecular testing, with 3,500 of those individuals expected to be eligible for some form of targeted medicine.[32] To ensure that individuals are able to access the appropriate testing, and consequently, correct targeted medicine, substantial investment in the NHS genomics services are needed to increase capacity and to ensure that staff have appropriate skills and training for specific genetic analysis. There is also an additional need for education and training to ensure that clinicians are aware of where targeted medicines could fit within each cancer type's treatment pathway. Clinical advisers to the ERG report that the provision of testing for patients

can be dependent on their clinician's knowledge of genomic medicine and available targeted therapies.

In addition to the requirement for a larger workforce, investment in laboratory infrastructure is needed to ensure sufficient equipment is available to deal with increasing demands.[33] In laboratories where RNA-based NGS or RNA/DNA hybrid-based NGS is not available, substantial investment would be required to provide infrastructure to enable NGS testing.

While the company acknowledge that screening for *NTRK* is likely to impact significantly upon the total cost of identifying and treating patients with entrectinib, the scale of practical and infrastructural considerations associated with the introduction of such a vast number of tests to NHS pathology services is not addressed. As new tests for molecular markers are introduced, increasing numbers of patients being referred for increasingly complex diagnostic investigations continues to outstrip the ability of the service to increase testing capacity. Cancer Research UK predicts a "severe crisis" in pathology capacity in the next 5-10 years,[34] and the ERG's clinical advisers agreed that existing infrastructure could not accommodate the proposed increases in IHC testing without significant NHS investment. Capacity constraints have been identified as a key barrier to the introduction of precision medicines onto the NHS, but investment in increasing capacity is rarely considered in cost-effectiveness evidence.[33] Therefore economic evaluations that fail to integrate these considerations may not provide meaningful evidence on how to implement precision medicines in a cost-effective way.

3 Critique of company's definition of decision problem

3.1 Population

The clinical efficacy evidence submitted by the company included Trk inhibitor-naïve patients with *NTRK* fusion-positive solid tumours that is limited to 10 tumour types included in the entrectinib clinical trials. This includes sarcoma, NSCLC, MASC, breast cancer, thyroid cancer, CRC, neuroendocrine tumour, pancreatic cancer, gynaecological and cholangiocarcinoma. In descriptions of the integrated efficacy analysis, the company do not differentiate between patients with different breast cancers, thyroid cancers or gynaecological cancers. Throughout this report the ERG therefore refers to 13 tumour types to reflect these subtypes. Following request for clarification, the company provided further ORR data for three additional tumour types, including three in a paediatric population (primary CNS, infantile fibrosarcoma and skin cancer) and one in adults (primary CNS).

Table 6 presents an overview of the 13 tumour types for which the CS presented efficacy data, for a total of 66 patients. The number of patients representing each of the tumours included in the entrectinib efficacy evidence is small, ranging from one (cholangiocarcinoma, paediatric skin cancer) to 13 (soft tissue sarcoma). The most frequently represented solid tumour types in the trial evidence were sarcomas (19.7%), NSCLC (15.2%), salivary gland tumours (MASC) (10.6%), and breast cancer (9.1%), which together accounted for over half of patients (54.7%). However, there is a mismatch between the distribution of tumour types in the efficacy population and the estimated yearly prevalence in England calculated by the ERG. For instance, the efficacy evidence included four patients with secretory breast carcinoma, over 10 times the estimated prevalence of the eligible population (0.3/year). Other over-represented populations include sarcoma (13 patients included, yearly prevalence 4) and MASC (7 patients, vs. 2). Conversely, the tumour types included in the efficacy evaluable population with the highest estimated eligible population in England were represented by relatively fewer patients: three papillary thyroid tumour cancers (26/year), three pancreatic cancer (15/year) and four CRC patients (14/year).

The ERG estimate that, of the 13 tumour types included in the trial evidence (including CNS primary and the included paediatric population), approximately 159 patients per year will be eligible for entrectinib. This represents 81.0% of the ERG's estimated annual Trk-inhibitor eligible population of 196 patients, which includes CNS primary and paediatric patients (see Appendix A). This indicates that the trial evidence includes a number of tumour types likely to harbour a larger number of patients who would be eligible for entrectinib according to the anticipated marketing authorisation.

The CS noted that in clinical practice, *NTRK* gene fusions may be present in additional tumour types and histologies. Clinical advisers to the ERG noted that theoretically *NTRK* fusions may be present in over 400 solid tumour types. The Foundations Medicine Inc. dataset only identified *NTRK* fusion in

41 tumour types, out of circa 116,000 samples. Therefore a significant number of tumour types and populations known to harbour *NTRK* fusions are not represented in the entrectinib efficacy evidence.

Table 6 Tumour types included in the entrectinib efficacy evidence

Tumour type/population (high level)	**Tumour Type (low level)**	**N included in efficacy evidence**	**ERG estimated prevalence per year (eligible population)[#]**
Salivary gland (MASC)	[redacted]	[redacted]	2
Lung	[redacted]	[redacted]	10
Breast	[redacted]	[redacted]	4
	[redacted]	[redacted]	0.3
Thyroid	[redacted]	[redacted]	26
	[redacted]	[redacted]	NE
	[redacted]	[redacted]	NE
Colon/colorectal	[redacted]	[redacted]	14
Neuroendocrine	[redacted]	[redacted]	4
Cholangiocarcinoma	[redacted]	[redacted]	0.3
Pancreatic	[redacted]	[redacted]	15
Sarcoma/soft tissue sarcoma	[redacted]	[redacted]	NE
	[redacted]	[redacted]	NE
	[redacted]	[redacted]	NE
	[redacted]	[redacted]	NE
	[redacted]	[redacted]	3
	[redacted]	[redacted]	NE
	[redacted]	[redacted]	4
Gynaecological	[redacted]	[redacted]	1
	[redacted]	[redacted]	3
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]		66	97

[#] according to the positioning of entrectinib presented in CS table 6; values were rounded to nearest integer unless <1. [+] Estimated prevalence of thyroid tumour (NOS) was 5.6; [redacted]

Table 7 summarises the characteristics of *NTRK* fusion positive patients included in the efficacy evaluable population (EEA). This includes the combined population of 54 adult patients across 10

tumour types (NSCLC, MASC, sarcoma, breast cancer, thyroid cancer, CRC, neuroendocrine tumour, pancreatic cancer, gynaecological and cholangiocarcinoma) with at least 6 months follow-up enrolled into entrectinib studies up to November 30th 2017, and excludes adults with CNS primary tumours and paediatric patients. Given the limited evidence on the *NTRK* population the extent to which trial population characteristics reflect those of the broader population under the NICE scope is difficult to assess. Clinical advisers to the ERG confirmed the EEA population characteristics were broadly representative of the population defined in the anticipated marketing authorisation beyond standard of care

Table 7 Characteristics of *NTRK* trial population (from CS Table 9)

Characteristic	**Description**	***NTRK* efficacy cohort (n=54)**
Age (years)	Median (range)	57.5 (21-83)
	<65	34
	≥65	20
Gender	Female	59.3%
Race	White	79.6%
	Asian	13.0%
	Not reported	7.4%
Mean BSA, m^2 (SD)		1.85 (0.26)
Mean BMI, kg/m^2 (SD)		25.68 (5.30)
Median time since diagnosis, months (range)		21.4 (2.1–433.1)
Disease stage at initial diagnosis, n (%)	0, I or II (A/B)	15 (28.3)[a]
	III (A/B/C) or IV	33 (62.3)[a]
	Unknown	5 (9.4)[a]
Performance status	ECOG 0	42.6%
	ECOG 1	46.3%
	ECOG ≥2	11.1%
Smoking status	Never-smoker	56.6%
Metastatic disease	Any site	96.3%
	Baseline CNS metastases	20.4%
No. of lines of therapy since metastatic disease[b], n (%)	0	37.0%
	1	20.4%
	2	25.9%
	3	7.4%
	≥4	9.3%
Previous therapy[c]	Any systemic therapy	88.9%

	Surgery	79.6%
	Radiotherapy	66.7%

CNS, central nervous system; ECOG, Eastern Cooperative Oncology Group
[a] Percentages calculated based on denominator of 53 patients as one patient in the ALKA study for whom the initial diagnosis field on the Case Report Form was blank was excluded.
[b] Patients may have received other therapies in the adjuvant or neo-adjuvant setting that are not included as a line of therapy from the time of metastatic disease diagnosis.
[c] Includes chemotherapy, immunotherapy, targeted therapy or hormonal therapy.

The EEA population includes only one *NTRK2* fusion positive patient. The company did not provide data on *NTRK2* prevalence in the adult primary CNS and paediatric population. Epidemiological evidence of *NTRK* suggests that the prevalence of subtypes 1/2/3 vary across tumour types,[2] although the ERG agrees with the company that estimates are uncertain given the rarity of *NTRK* fusions and variation in testing methods. Clinical advisers to the ERG noted that there is theoretically no reason to suggest that only one type of *NTRK* fusion should be present between patients within any given tumour type. In response to a clarification request from the ERG, the company stated that the low *NTRK2* prevalence in the trial population is reflective of that observed in the wider *NTRK* population ([redacted]) according to the Foundation Medicine Inc. dataset. This is much higher than the 1.9% prevalence reported in the EEA population, therefore the ERG believe that the *NTRK2* population is significantly underrepresented in the efficacy evaluable population.

The company's proposed positioning of entrectinib is as first line therapy for five tumour types (MASC, soft-tissue sarcoma, pancreatic cancer, cholangiocarcinoma and gynaecological cancers) and second line or beyond for five tumour types (NSCLC, breast, thyroid cancer, CRC, neuroendocrine tumours) although the CS noted that some of these patients may receive entrectinib as first line systemic therapy if not eligible for targeted treatments or immunotherapies. In response to clarification, the company positioned entrectinib as second line for the following tumour types: CNS primary (adults), CNS primary (paediatric), sarcoma (paediatric); and as second line or beyond for paediatric skin cancer (paediatric). The company stated they anticipated use of entrectinib in [redacted]

Following a request from the ERG, the company reported individual participant data (IPD) including the number of lines of prior systemic therapy since diagnosis for the efficacy evaluable population, as well as five *NTRK* patients with primary CNS and seven paediatric patients who were excluded from the efficacy evaluable population, from the [redacted] clinical cut-off data. Table 8 presents the distribution of this population by line of therapy. This shows that [redacted] of patients received entrectinib as first line systematic therapy, [redacted] as second line, and [redacted] as third line or beyond. The company provided no breakdown of line of therapy by tumour type received between 3rd line and subsequent lines. Table 8 shows that entrectinib was administered as either first line or as a subsequent line of therapy in all tumour types except cholangiocarcinoma, gynaecological cancers and paediatric skin cancer. The

absence of an alternative "acceptable therapy" was not an eligibility criterion in STARTK-2, which formed the large majority of the total clinical efficacy population. For this reason, some trial participants may not match the population as defined in the NICE scope.

Table 8 indicates there is a mismatch between the proposed positioning of entrectinib and the trial population evidence submitted. For all five cancer types proposed as 1st line therapy, only 35% of the efficacy evaluable population received entrectinib as 1st line therapy. For the remaining cancer types that were positioned as 2nd line or beyond, 40% of patients received entrectinib as first line therapy; however, it is not clear what proportion of these patients received entrectinib as first line because they were not eligible for targeted treatment or immunotherapies. There was no overlap between the proposed positioning of entrectinib and the trial population submitted for two cancer types: although positioned as first line therapy, patients with cholangiocarcinoma and gynaecological cancers received entrectinib as 3rd line therapy and/or beyond. Overall, this mismatch limits the extent to which the trial evidence supports the company's proposed positioning of entrectinib. Although a higher response rate and better survival outcomes may be expected from patients receiving entrectinib as first line therapy, there is insufficient survival outcomes evidence to determine whether this mismatch may have favoured entrectinib. This matter is further discussed in section 4.2.6.1.

Table 8 Distribution of *NTRK* participants by line of systemic therapy and tumour type (efficacy evaluable population +5 CNS primary adults and 7 paediatric patients)

Line of therapy	**1st line**	**2nd line**	**3rd line & beyond**	**Total**
*Company proposed positioning: 1st line**				
Cholangiocarcinoma	[redacted]	[redacted]	[redacted]	1
Gynaecological (endometroid, ovarian)	[redacted]	[redacted]	[redacted]	2
Pancreatic	[redacted]	[redacted]	[redacted]	3
Salivary glands (MASC)	[redacted]	[redacted]	[redacted]	7
Sarcoma	[redacted]	[redacted]	[redacted]	13
Total (of tumour types proposed as 1st line)	[redacted]	[redacted]	[redacted]	**26**
Company proposed positioning: ≥2nd line*				
Breast	[redacted]	[redacted]	[redacted]	6
CRC	[redacted]	[redacted]	[redacted]	4
Neuroendocrine	[redacted]	[redacted]	[redacted]	3
NSCLC	[redacted]	[redacted]	[redacted]	10
Thyroid	[redacted]	[redacted]	[redacted]	5
CNS primary (adults)	[redacted]	[redacted]	[redacted]	5
CNS primary (paediatric)	[redacted]	[redacted]	[redacted]	4
Sarcoma (paediatric)	[redacted]	[redacted]	[redacted]	2

Skin cancer (paediatric)	[redacted]	[redacted]	[redacted]	1
Total (of tumour types proposed as ≥2nd line)	[redacted]	[redacted]	[redacted]	**40**
Total (all tumour types)	[redacted]	[redacted]	[redacted]	**66 (100%)**

*Patients ineligible for curative surgery or radiotherapy with no immunotherapy or targeted therapy options

Some patients may receive first-line entrectinib treatment if not eligible for targeted or immunotherapies

In response to clarification questions, the company provided data on the subsequent therapies received by the trial participants. [redacted] patients in the efficacy evaluable population received a subsequent chemotherapy after progression. A number of these were [redacted]. This suggests that for these patients there are 'acceptable' alternative standard therapies available besides chemotherapy, hormone therapy or best supportive care. No information was provided on the total number of patients receiving subsequent targeted therapies, or on which patients received which subsequent targeted therapy. This further limits the extent to which the integrated efficacy analysis matches the proposed population as defined in the NICE scope.

In summary, the trial population includes only a subset of the total population that is potentially eligible for entrectinib. Although the trial evidence includes a number of tumour types expected to harbour a larger number of patients who may be eligible for entrectinib according to the anticipated marketing population, the large majority of tumour types potentially harbouring *NTRK* fusions are not represented in the evidence submitted. There is also a mismatch between the trial population and the proposed positioning of entrectinib. For an unknown number of patients, there appeared to have been 'acceptable' alternative standard therapies available besides chemotherapy, hormone therapy or best supportive care. Due to concerns about the large number of missing tumour types, the under-representation of *NTRK2* patients, the small sample size of the *NTRK* efficacy trial population, and concerns about the positioning of entrectinib in the trial evidence, the ERG is concerned that the population presented in the evidence submitted does not match the NICE final scope. In particular, the high proportion of patients receiving entrectinib in earlier lines of therapy across tumour types may lead to overestimating its survival benefits.

3.2 Intervention

The intervention is entrectinib at the recommended dose of 600 mg orally once daily for adults, and 300 mg/m^2 orally, once daily for paediatric patients who have the ability to swallow whole capsules. This is in line with the NICE scope.

3.3 Comparators

As the CS evidence only includes trials with no control arms, the company adopted a pragmatic approach to identify comparator data for established management without entrectinib. The company conducted a search of NICE pathways to identify comparators approved by NICE for each of the tumour types represented in the CS efficacy evidence population, using tumour type search terms. Where searches resulted in multiple possible pathways on the NICE Pathways website, the company made a decision on the pathway most relevant to the decision problem, for example the pathway referring to management of advanced/metastatic patients. Although the ERG understand that the company chose not to conduct separate systematic reviews to identify comparator data due to the large number of comparators and tumour types, the risk that relevant evidence may be been omitted cannot be excluded.

Therapies including chemotherapy, hormone therapy and best supportive care that had received a positive NICE recommendation were included. Excluded therapies were: surgery with curative intent, radiotherapy (non-palliative), immunotherapy, targeted agents and biological therapy. The clinical advisers to the ERG confirmed that these criteria are likely to be generally applicable for the majority of clinical scenarios.

Choices of lines of therapy by tumour histology were made by the company, and comparators were selected following current NICE recommendations. Median PFS and OS data were extracted from the clinical effectiveness data presented within the Committee slides, or where not available, from the company's submission. Trial participant characteristics, estimates of precision, or the committee-preferred parametric models used to extrapolate OS and PFS data were not extracted. Where chemotherapies were recommended by NICE but clinical evidence was not specifically available, the comparator data was not included. The company did not clarify which comparators and tumour types this criterion was applied to. The ERG believe this approach to be inappropriate and that a targeted systematic search for relevant evidence, and where possible, extracting data directly from survival curves for a better estimate of median PFS/OS and its variance, would have been preferable.

Where multiple PFS and OS values were available for a given comparator, the company extracted median values from primary analyses of individual trials informing the NICE TA analyses. Subgroup values were not used. The company stated that technology appraisals were informed primarily by one randomised controlled trial, and there was only one median value provided for each outcome that was relevant to the decision problem or the scope of the technology appraisal for the given comparator. However, the ERG found that in some instances, two different estimates where used for a single agent within the same line of therapy. For instance, for best supportive care for 2L+ thyroid cancer, figures adjusting and not adjusting for cross-over were both extracted.

In response to a clarification request, the company stated that comparator efficacy data was drawn from multiple technology appraisals for individual comparators in the same line of therapy where available (for example, docetaxel in NSCLC). They noted that this decision was taken to increase the robustness of the comparator data, by taking a mean of multiple values, and to ensure that an outlying or extreme value was not inadvertently used. The ERG found that in a number of instances (such as trifluridine-tipiracil for CRC, or eribulin for breast cancer) where more than one trial informed the TA, or where subgroups where combined (e.g. everolimus and best supportive care for neuroendocrine tumours in different sites) using inputs from robust meta-analytical techniques (for instance, as reported in company submissions or conducted by ERGs) would have been preferable to naively pooling unweighted means of medians, which is statistically inappropriate.

Where no chemotherapies were recommended by NICE, no additional targeted systematic reviews were conducted. Instead, the company used one of two approaches to identify relevant comparator data. In the case of MASC, surrogate trial data for best supportive care was used to derive OS data. However, it is not clear how this trial data was identified. Other tumour types for which only one patient was included in the efficacy population (cholangiocarcinoma, endometroid and ovarian cancers) were grouped into a single "other" category. PFS and OS estimates were derived for this category by calculating an average of PFS and OS median estimates from comparator data selected for the other tumour types (colorectal cancer, NSCLC, breast cancer, soft tissue sarcoma, pancreatic cancer, neuroendocrine tumours and thyroid cancer). The same method was used to derive PFS comparator data for MASC, reportedly due to lack of evidence. The ERG finds this method inappropriate, as prognosis for patients with a given tumour type such as cholangiocarcinoma, gynaecological cancers or MASC may differ significantly from patients with other unrelated tumour types.

The final choice of comparators was validated by clinical advice for seven of the ten cancer types included in the entrectinib efficacy evaluable population. These include: colorectal cancer, NSCLC, breast cancer, soft tissue sarcoma, pancreatic cancer, neuroendocrine tumours and thyroid cancer. The company did not state whether PFS and OS values were validated by clinical advice for salivary gland cancer (including MASC), cholangiocarcinoma, and gynaecological cancers. It is not clear whether comparator data provided in response to clarification was clinically validated.

Extracted comparator data is presented in table 30 of the CS appendix, and in the company model. The ERG identified some discrepancies between the two sources, which the company addressed in response to clarification. Table 9 summarises PFS and OS values for comparator data selected by the company that informed the economic model. These also include ORR values extracted by the company. This shows that all tumour types except one (MASC) were assigned at least two comparators across multiple lines of therapy. For each cancer type, individual median PFS and OS

estimates from each comparator treatment were pooled to calculate an unweighted mean PFS and OS. This method implies there is an even distribution of patients receiving each of the therapeutic options within a given line of therapy, which does not reflect clinical practice. For instance, due to its toxicity, irinotecan is less likely to be administered than FOLFIRI as second line therapy for CRC, therefore these therapies should not be given the same weight when pooling estimates across comparators. Similarly, this approach assumes there is an equal distribution of patients receiving different lines of therapy within a tumour type, which is not reflective of clinical practice. For instance, doxorubicin and trabectedin, respectively ≥1L and ≥2L treatments for soft tissue sarcoma, were given equal weight to generate an average value of all comparator efficacy data for this tumour type. This approach also did not take into account the distribution of the underlying data including heterogeneity in prognosis factors across different comparators within each tumour type, or estimates of variance and precision in survival estimates, and is therefore invalid.

Median and mean survival estimates were then applied at an individual patient level to calculate an overall mean PFS and OS across all tumour types, weighted by the number of individual patients with each tumour type in the integrated efficacy population. These estimates were used to inform naïve, unadjusted comparisons with entrectinib efficacy data. These comparisons do not account for any potentially important patient characteristics, such as age, performance status, *NTRK* fusion status or the prevalence of CNS metastases.

Following request for clarification, the company provided comparator data for the following tumour types: primary CNS (adults and paediatric), infantile fibrosarcoma and malignant melanoma (Table 10).

Table 9 Selected comparator data (from CS, entrectinib Roche model, Inputs for SoC *NTRK+*)

Tumour type	Therapy	Entrectinib proposed line of therapy (from CS table 6)	Comparator line of systemic therapy		ORR	Median PFS (months)	Median OS (months)	Reference	Average Estimation	
									PFS	OS
			CS appendix table 30	ERG extracted from trial data						
Non Small-cell Lung Cancer	Docetaxel	≥2L	≥2L	≥2L	NR	3.3	8.7	Average of values from NICE TAs 520, 428, 483, 484, 403, 347, 124		
	Docetaxel + nintedanib		≥2L	2L	4.7	4.2	12.6	NICE TA347	3.75	10.65
Colorectal Carcinoma	FOLFIRI	≥2L	2L	2L	11.1	4.7	12.1	NICE TA307		
	Irinotecan		2L	2L	34.8	4.4	14.3	NICE Guideline CG121 - Kim et al 2009		
	Trifluridine-tipiracil		≥3L	≥3L	0.9	2	9	NICE TA405		
	Trifluridine/Tipiracil		≥3L	≥3L	1.6	2	7.2	NICE TA405		
	Best supportive care		≥3L	≥3L	0.0	1	6.6	NICE TA405		
	Best supportive care		≥3L	≥3L	0.0	1.7	5.2	NICE TA405	2.63	9.07

Breast cancer incl. secretory breast	Capecitabine	≥2L	2L	1 to 3L	11.5	4.1	14.5	NICE TA515		
	Eribulin		≥3L	1 to 6L	12.2	3.6	13.2	NICE TA423		
	Vinorelbine		≥3L	NR	4.7	2.2	10.5	NICE TA423		
	Gemcitabine + paclitaxel		≥3L	NR	4.7	2.2	10.5	NICE TA423	3.03	12.18
Salivary Gland Cancer (incl. MASC)	Best supportive care (Platinum+Gemcitabine data used as surrogate)	1L	≥1L	1-3L	NR	4.3	13.8	Surrogate data for BSC - Laurie et al. 2010	4.35	13.80
Soft Tissue Sarcoma	Doxorubicin	1L	≥1L	≥1L	7.5	4.1	14.7	NICE TA465		
	Trabectedin		≥2L	≥2L	5.1	3.7	13.9	NICE TA185	3.90	14.30
Pancreatic	Gemcitabine + nab-paclitaxel	1L	≥1L	1L	23	5.5	8.7	NICE TA476		
	Gemcitabine		≥1L	1L	7	3.7	6.6	NICE TA476		
	FOLFIRINOX		≥1L	1L	31.6	6.4	11.1	NICE Guideline NG85 - Conroy et al 2011	5.20	8.80

Thyroid (papillary), unsuitable/refractory to radioactive iodine	Best supportive care	≥2L	≥2L	1 to 2	1.5	3.7	19.1 (after cross-over adjustment)	NICE TA535 (Cross-over adjusted value from Guo et al 2015)		
	Best supportive care		≥2L	1 to 2	0.5	5.4	42.8	NICE TA535	4.55	30.95
Neuroendocrine tumours	Everolimus (pancreatic)	≥2L	≥2L	≥1L	4.8	11	44.02	NICE TAs 449 and 539		
	Everolimus (GI & lung)		≥2L	≥1L	2	11	37.16	NICE TAs 449 and 539		
	Best supportive care		≥2L	≥1L	2	4.6	37.68	NICE TAs 449 and 539		
	Best supportive care		≥2L	≥1L	1	5.5	39.56	NICE TAs 449 and 539	8.025	39.605
Others*	[redacted]					4.6#	17.2^			

*Cholangiocarcinoma, uterine and ovarian ; #unweighted average of all PFS estimates except MASC. ^unweighted average of all OS estimates.
ORR: objective response rate; OS: overall survival; PFS: progression free survival; GI: gastrointestinal; TA: technology assessment

Table 10 Selected comparator data (from company clarification, entrectinib Roche model, Inputs for SoC NTRK+)

	Treatment	**Entrectinib proposed line of therapy (from CS table 6)**	**Line of systemic therapy**		**ORR**	**PFS**	**OS**	**Source**	**Average Estimation**	
			CS appendix table 30	ERG extracted from individual trials					PFS	OS
High grade glioma (after surgery/radiotherapy)	Temozolomide	NR	2L	2L	5.4	2.89	7.34	TA23		
	Procarbazine, CCNU (lomustine) and vincristine	NR	2L	1L	NR	3.6	6.7	Brada M et al, 2010 35		
	Single agent CCNU (lomustine)	NR	2L	2L	14.4	3.0	9.8	Batchelor T et al, 2013 36	3.16	7.95
Infantile Fibrosarcoma (after surgery/chemotherapy)	Best supportive care	NR	2L	NA	NA	4.1 (average of known)	15.8 (average of known)	NA	4.11	15.85
Malignant melanoma	Dacarbazine	NR	2L+	1L	12.1	1.5	6.4	NICE guideline NG14 (Middleton MR et al, 2000) 37	1.5	6.4

Limitations in reporting meant that in some cases calculation of comparator data could not be replicated. For instance, reporting was insufficient to replicate calculation of average PFS and OS values for the large number of NSCLC comparators and TAs used.

The company noted that, given that *NTRK* fusion status and high prevalence of CNS metastases (20.4%) in the entrectinib trial population are not accounted for, the comparator OS and PFS values may be overestimated. The ERG found insufficient evidence to support this statement. There is insufficient evidence to suggest that *NTRK* fusions are associated with a worse prognosis for most of the tumour types presented in the trial efficacy population (see section 2.1). The prevalence of CNS metastases in the comparator data is also uncertain. Relevant participant characteristics from trials informing PFS and OS inputs were not extracted. Therefore, the ERG checked key participant characteristics of all trials informing the comparator data reported in the publications from the trials. Most comparator trial publications did not report whether CNS metastases were excluded, except for the following comparators: FOLFIRI and irinotecan for CRC, and best supportive care (BSC) for MASC. Other trials reported inclusion restrictions, and only included patients with treated/stable CNS metastases (capecitabine & eribulin for breast cancer, doxorubicin and trabectedin for sarcoma), similarly to STARTRK-2 (see Section 4.2.2). Most comparator trials did not report baseline prevalence of CNS metastases, except for NSCLC trials (ranging from 6% to 14%). Further details on comparator trial extracted by the ERG on population characteristics, end of life and survival distributions used in TAs, are reported in Appendix C.

The ERG found a mismatch between the lines of therapy used in the comparator data and those reported in the efficacy evaluable population for some tumour types. As reported in Table 8, just over a third of the company's trial participants received entrectinib as first line systemic therapy, and entrectinib was administered in treatment naïve patients in 10 of the 13 tumour types (all except cholangiocarcinoma, gynaecological, and paediatric melanoma) represented in the trial evidence. The company identified comparator data including treatment naïve patients in all of those 10 tumour types, with the exception of NSCLC and CRC.

The ERG also identified a mismatch between the company's proposed positioning of entrectinib (as reported in CS table 6) and the line of therapy in which the comparators were used in the trials identified by the company. Soft tissue sarcoma and MASC comparator trials included patients in second line therapy and beyond, although the company placed entrectinib as first line for these tumour types. Conversely, comparator trials included first line patients where entrectinib was positioned as 2nd line or beyond for the following tumour types: breast cancer, thyroid and neuroendocrine tumours.

In response to clarification, the company stated that their proposed lines of therapy by tumour type are provided in the appendix comparator table, which the ERG presents in Table 9. The ERG found a discrepancy between this source and the proposed positioning presented in CS table 6 for soft tissue sarcoma, the appendix table includes a second line comparator (trabectedin) while CS table 5 positions entrectinib as first line for soft tissue sarcoma.

Adverse events of comparator therapies were not extracted, therefore the safety of entrectinib could not be compared to other relevant therapies. Utilities were extracted from TA documentation and are further discussed in Section 5.2.7.2.

The ERG consider the methods used to identify, select and combine comparator data to have a number of important limitations. Overall, the populations included in the comparator trials do not match the entrectinib efficacy population, notably due to likely limited prevalence of *NTRK* fusion in comparator evidence, and mismatch in lines of therapy within the treatment pathway. Comparisons were naïve and do not account for any potentially important prognostic factors, such as age, performance status, *NTRK* fusion status, prevalence of CNS metastases, or specific tumour mutations within each tumour type. In the base case analysis ,no attempts were made to adjust for differences in population characteristics between the entrectinib and comparator trial populations. Overall, the ERG conclude that the comparator data used to inform the company model is highly unreliable. Due to a high risk of confounding bias, comparisons with entrectinib are unlikely to be reliable. Alternative methods for addressing the uncertainty associated with the comparator evidence are presented in Section 5.2.6.1.

3.4 Outcomes

The outcomes presented in the CS include overall survival, progression free survival, overall response rate, duration of response, adverse effects of treatment and health-related quality of life. These match the outcomes specified in the NICE scope.

3.5 Other relevant factors

The CS provided analyses of ORR for the subgroups specified in the NICE scope (previous therapy and tumour type) alongside other subgroups the company considered relevant. Following a request from the ERG, the company provided IPD level data for ORR outcomes by tumour type and line of therapy, but not for PFS and OS.

The company proposed a data collection plan via the cancer drug fund (CDF). A Patient Access Scheme discount of [redacted] off the entrectinib list price has been agreed with NHS England. The CS did not identify any equality issues.

4 Clinical Effectiveness

This section contains a critique of the methods of the review of clinical effectiveness data, followed by a description and critique of the trials included in the review, including a summary of their quality and results, and the results of evidence syntheses performed by the company.

4.1 Critique of the company review methods

4.1.1 Searches

The company performed a systematic search for randomised controlled trial (RCT), non-randomised, and observational studies that investigated the efficacy and safety of entrectinib for the treatment of patients with *NTRK*-positive solid tumours.

The literature searches for each of the systematic literature reviews (SLRs) – clinical effectiveness, economic evaluation, health state utility value – were carried out "in parallel". Consequently one search was conducted of each resource/database that included each of these aspects. For the cost/resource use systematic literature review, separate searches were conducted. Overall, the ERG considers that the searches carried out were well conducted and reported and appropriate sources were used so the likelihood of relevant studies not being identified is low.

The databases used for the effectiveness review are reported as being MEDLINE (segments used were 1946 to present, Daily and In Process & Epub Ahead of Print), Embase, EconLit and EBM Reviews. The latter resource includes a range of other databases: Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, Database of Abstracts of Reviews of Effects, Health Technology Assessment database, NHS Economic Evaluation Database, ACP Journal Club, Cochrane Clinical Answers and Cochrane Methodology Register. The search strategies used in each of the databases are fully reproduced in section D.1.1.3 of the CS and the date that they were conducted is given. The numbers of records retrieved matches the number given in the PRISMA diagram (CS page 25).

Additional searches of conference websites (ASCO, ECCO, ESMO,AACR) were conducted to identify potentially relevant posters and abstracts and the reference lists of identified studies were reviewed. Searches of the trials registers ClinicalTrials.gov and the WHO ICTRP were also conducted to find ongoing studies. HTA websites were scanned by the company to identify previous regulatory submissions, including NICE, SMC, AWMSG, PBAC, CADTH including pCODR.

Other searches were conducted using the Cost Effectiveness Analysis Registry, RePEc (EconPapers within Research Papers in Economics), www.euroqol.org, www.inahata.org, www.hta.org.ac.uk, ScHARRHUD utility database, University of York Centre for Reviews and Dissemination, and Google Scholar. No details are reported about the terms in the searches of these additional resources.

The strategy used in the Embase, MEDLINE and EBM Reviews databases consists of three sections i.e. 1) *NTRK* fusion 2) entrectinib OR comparators and 3) quality of life. Sets were combined to retrieve studies about: a) *NTRK* fusion and entrectinib or comparators or b) *NTRK* fusion and quality of life and c) comparators and quality of life. The strategy for the systematic review of resource use consists of terms for the condition (*NTRK* Fusion) combined with search terms for resource use/health care costs.

The overall structure of the strategy is appropriate and there are no errors in how the sets are combined. Neither are there any typographical errors within the search terms used. A validated search filter to identify HSUV was incorporated into the search strategy

The search strategy used for Embase, MEDLINE and EBM Reviews consists mainly of free text terms rather than a combination of thesaurus and free text terms. This broad approach can be successful when seeking to identify studies that are available in Embase as conference abstracts as they have less detailed indexing applied to the database record. The Embase search strategy did not include the EMTREE heading “protein tyrosine kinase inhibitor” although after testing it was clear that this did not change the overall numbers retrieved.

The search of EconLIT appropriately was much broader than that conducted in Embase and MEDLINE, consisting solely of terms for *NTRK*, TRK fusion combined with terms for the intervention and comparators. No information is given about how the website searches were conducted

4.1.2 Inclusion criteria

The company provided full details of the inclusion criteria used in the systematic literature review in Table 1 of CS Appendix D. Studies were single-screened for inclusion and independently checked by a second reviewer, with any discrepancies resolved through consensus. Studies were eligible for inclusion if they recruited patients with *NTRK*-positive solid tumours to prospective RCT (stage 2-4), non-randomised, or retrospective/prospective observational cohort studies. Studies evaluating the efficacy, safety, and/or HRQoL associated with entrectinib and a number of comparators (e.g. belizatinib, cabozantinib, larotrectinib, repotrectinib) were eligible for inclusion. While these interventions were included in the company’s original SLR, as per the NICE scope only those studies assessing entrectinib were included in the main CS. Efficacy outcomes included overall survival, progression-free survival, time-to-progression, duration of response, time-to-response, and objective response rate. Safety outcomes of interest were any treatment-related adverse event and tolerability issues, i.e. dose reductions and interruptions, treatment discontinuation. Details of HRQoL and patient reported outcome measures administered as part of clinical trials were also included.

There were no restrictions by location and date; the primary focus of the review was on English language publications, or non-English language publications with an abstract in English.

4.1.3 Critique of data extraction

Data extraction was performed and reported adequately. Appendix D of the CS stated that data were extracted from the included studies by one reviewer, with all extracted data checked against the source document by a second reviewer.

The company's main submission presents detailed information about the included studies (ALKA, STARTRK-1, STARTRK-2, and STARTRK-NG), with a summary of the methods, participant characteristics, and results presented in Table 7 and Table 8 of the CS (pages 31-38). The company provided further patient characteristic and efficacy data in their clarification responses at the ERG's request.

4.1.4 Quality assessment

The company did not present a formal quality assessment of any of the studies included in the SLR, reasoning that as they were not primary full publications there was insufficient available evidence to adequately assess the quality of the study.

The ERG did not consider the company's lack of quality assessment appropriate, particularly given the availability of evidence to the company about their own trials. Therefore the ERG conducted its own quality assessment on STARTRK-2, the primary source of clinical data used in the company's analysis, based on the Downs and Black checklist[38] (see Section 4.2.2), which assessed quality of reporting, external validity, internal validity, confounding, and study power.

4.1.5 Evidence synthesis

The four included studies enrolled a total of 357 patients; however, as the majority of patients recruited to these trials were not *NTRK* fusion-positive, the company pooled across the trials the patients who met the following criteria:

- Had at least 6 months follow-up
- Had *NTRK* gene fusion positive tumours
- Received at least 1 dose of entrectinib
- Had not been previously treated with a Trk inhibitor

This population is referred to by the company as the *NTRK* Efficacy Population (n=62), from which further *post hoc* exclusions were made; six patients with primary CNS tumours, one paediatric patient, and one patient with non-measurable disease. The resulting patient group is the '*NTRK* Efficacy Evaluable Analysis Set' (EEA) (n=54), which forms the basis of the company's efficacy analyses and

is the population upon which the economic model is focused. Following the ERG's request, the company provided updated analyses which included the 5 primary CNS and 7 paediatric patients for whom outcomes were available.

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

Four non-randomised single-arm Phase I/II basket studies were included in the company's 'integrated efficacy analysis set', which was a *post hoc* pooling of participants designed to maximise the number of patients included in the analysis. These studies investigated the efficacy and safety of entrectinib in adult patients (ALKA, STARTRK-1/-2), and paediatric/adolescent patients (STARTRK-NG) with tumours positive for *NTRK1/2/3, ROS1,* or *ALK* molecular alterations. The STARTRK-2 trial contributed 51 of the 54 patients included in the company's efficacy analyses, therefore this study, and the pooled integrated efficacy dataset forms the focus of the following section.

4.2.1 Design and analysis of basket trials

A common approach to evidence generation in histology independent cancer therapeutics is the basket trial. In the context of the present appraisal, a basket trial designed to address the decision problem would evaluate a single drug targeting a single mutation (i.e. *NTRK* gene fusion) in multiple disease cohorts defined by histology or tumour type. Typically, two-stage studies are designed to recruit a certain number of patients to each 'basket', and if a pre-specified proportion of patients in a particular basket respond, then recruitment is expanded within this disease area. If too few responses are observed within a basket then recruitment is stopped due to low promise of efficacy.

While the company cites FDA, EMA, and EUnetHTA opinion stating that basket trials are acceptable for HTA of tumour agnostic therapies, the basket study (STARTRK-2) comprising the majority of the company's evidence submission was not designed as a basket trial in the sense intended by these bodies. The baskets in this study were based on molecular targets (*ALK*, *ROS1*, *NTRK*) rather than tumour type for each molecular target. Therefore assumptions underpinning the analysis of a basket trial may not hold for the analysis of *post hoc* subgroups within the *NTRK* fusion positive basket.

Heterogeneity of response across baskets is an important issue in the design and analysis of conventional basket trials, and in this case extra care must be taken to accommodate the potentially large variation and imprecision in response rate estimates introduced by very small sample sizes. There are a number of possible analytical approaches; one method is to analyse each basket separately as though it were an independent study. However, this approach does not allow for the possibility that some populations may respond in a homogeneous way, which is plausible given the common molecular target. The approach taken in the company's analysis was to assume equal efficacy across all baskets and to generate a pooled response estimate, but in doing so reject the potential for heterogeneity of response across baskets. A third approach assumes similar efficacy across baskets,

with the different histologies not determining a particular ordering of effectiveness *a priori*, i.e. the baskets are exchangeable and a Bayesian hierarchical model (BHM) can be used.[39] This type of design acknowledges the heterogeneity of response across baskets by assuming that response rates are exchangeable, rather than equal across baskets. This allows borrowing of information on the probabilities of response across baskets and increases precision of estimates, whilst reducing the chances of obtaining extreme estimates in specific baskets with few patients. Alternative forms of BHM have also been proposed, which allows borrowing of information across similar baskets while avoiding optimistic borrowing from extreme baskets.[40]

However, this hierarchical approach may increase uncertainty unnecessarily. Therefore, when there is a strong rationale for expecting a uniform level of response it may be preferable to use a simple pooling of information across subgroups as in the CS.[41] However, the company did not state any reasons to expect homogeneity of response across tumour types *a priori*, and indeed previous basket trials have shown heterogeneity in effectiveness of chemotherapeutic agents across tumour types. A recent trial of vemurafenib in 122 patients with *BRAF* V600–mutated cancers across multiple tumour types (including CRC, NSCLC, Erdheim–Chester disease and Langerhans'-cell histiocytosis, primary brain tumours, cholangiocarcinoma, anaplastic thyroid cancer) found evidence of response in some tumour types including NSCLC and Erdheim–Chester disease and Langerhans'-cell histiocytosis, but not in colorectal cancer.[42] A trial of imatinib, a tyrosine kinase inhibitor, that included 196 patients across 40 different subtypes, found evidence of activity of imatinib in only five malignancies.[43] Another basket trial of imatinib in 10 histologic subtypes of advanced sarcoma concluded that although rare dramatic responses were seen, imatinib was not an active agent in these subtypes, although it had previously shown effectiveness in another subtype of soft tissue sarcoma, gastrointestinal stromal tumour.[44]

Thus, it does not seem reasonable to make an assumption of homogeneity across tumour types, given the variability of the entrectinib trial results in the absence of a plausible clinical argument. Furthermore, as the included trials were not designed or sufficiently powered to test the assumption of heterogeneity of response across subgroups, the ERG consider it inappropriate and overly optimistic to assume equal response independent of tumour histology.

The ERG explores the effect of heterogeneity of response across tumour types using BHM in section 4.3.1.

4.2.2 STARTRK-2

STARTRK-2 is an ongoing multicentre, single arm, open-label, phase II basket study of entrectinib in patients aged ≥18 years with solid tumours harbouring *NTRK1/2/3*, *ROS1*, or *ALK* molecular alterations. As discussed in Section 4.2.1, the baskets in this study were based upon the three types of

genetic alteration, and not on tumour type. A total of 206 patients were enrolled and treated with entrectinib, with 63 enrolled to the *NTRK* basket. Patients were recruited across 84 sites in 15 countries, including 3 centres in the UK. Fifty-one patients met the criteria for inclusion (i.e. >6 months follow up and measurable disease at baseline) and thus formed the main part of the *NTRK* efficacy evaluable analysis set. The study design of STARTRK-2 is summarised below in Table 11, and eligibility criteria are summarised in Table 12. STARTRK-2 is a non-randomised, uncontrolled, open-label trial. Only a small subgroup (*NTRK*-fusion positive patients) of the trial informed the submission. Therefore, the evidence from this trial is considered at high risk of bias, and is not appropriately designed to assess the relative efficacy and safety of entrectinib against current established management. The trial eligibility criteria did not specify that inclusion into the trial was dependent on the lack of alternative effective and suitable standard therapy. As discussed in section 3.1, this may limit the extent to which the trial population matches the anticipated licence.

Table 11 Study design of STARTRK-2

Study details	
Location	84 sites in Australia, Belgium, France, Germany, Hong Kong, Italy, Japan, South Korea, The Netherlands, Poland, Singapore, Spain, Taiwan, United Kingdom, USA
Design	Non-randomised, one-arm, open-label
Duration of core study	4 years
Method of randomisation	None
Method of blinding	None
Intervention(s)	Entrectinib (RXDX-101)
Comparator(s)	None
Primary outcome	Objective Response Rate (ORR)
Data cut-off	35 months (Nov 2015 – [redacted])
Secondary outcomes	Duration of response (DOR), best overall response (BOR) time to response (TTR), progression free survival (PFS), safety outcomes.

Table 12 Eligibility criteria for STARTRK-2 (adapted from CS Table 8, Pages 34-37)

Inclusion criteria	**Exclusion criteria**

Age ≥18 Histologically- or cytologically-confirmed diagnosis of locally advanced or metastatic solid tumour that harbours an NTRK1/2/3, ROS1, or ALK gene rearrangement that is predicted to translate into a fusion protein with a functional TrkA/B/C, Ros1, or Alk kinase domain, respectively Measurable disease as assessed locally using RECIST v1.1 ECOG performance status ≤2 and minimum life expectancy of ≥4 weeks Prior anticancer therapy is allowed but 2 weeks must have elapsed following prior chemotherapy, and 4 weeks since completion of antibody-directed therapy Patients with CNS involvement, which is either asymptomatic or previously-treated and controlled.	Concomitant secondary oncodrivers (e.g., epidermal growth factor receptor, KRAS) Prior treatment with an approved or investigational TRK, ROS1, or ALK inhibitor in patients with tumours testing positive for the respective gene rearrangements Active gastrointestinal disease or other malabsorption syndromes

The ERG used the Downs and Black checklist to quality assess STARTRK-2 using the Interim Clinical Study Report provided by the company. This checklist scores the quality of reporting, external validity, internal validity, internal validity-confounding, and power of non-randomised trials. Results of the ERG's quality assessment using the Downs and Black checklist are presented in Appendix D. Overall, the ERG considers this trial to be at high risk of bias given that it is uncontrolled and only a fairly small subgroup of patients from this trial are included in the analysis.

4.2.3 ALKA

ALKA is an ongoing multicentre, single arm, open-label, phase I ascending dose and dose escalation study of entrectinib in patients aged ≥18 years with advanced/metastatic solid tumours with *NTRK1/2/3, ROS1,* or *ALK* molecular alterations. The primary objective of this study was to determine first cycle dose-limiting toxicity and the maximum tolerated dose of entrectinib. While only one patient from this study was included in the efficacy evaluable analysis dataset, a total of 57 patients were evaluable for safety outcomes.

4.2.4 STARTRK-1

STARTRK-1 is an ongoing multicentre, single arm, open-label, phase I ascending dose and dose escalation study of entrectinib in patients aged ≥18 years with solid tumours harbouring *NTRK1/2/3, ROS1,* or *ALK* molecular alterations. This study contributed two patients to the efficacy evaluable analysis dataset, and 76 patients were evaluable for safety outcomes.

4.2.5 STARTRK-NG

STARTRK-NG is an ongoing multicentre, single arm, open-label, phase I/Ib dose escalation and expansion study of entrectinib in patients aged 2-22 years with solid tumours harbouring *NTRK1/2/3, ROS1*, or *ALK* molecular alterations. At the time of the data cut used in the CS, only one patient from the STARTRK-NG trial had at least 6 months of follow-up, thus, no patients were included from this trial in the EEA dataset as originally presented by the company. Sixteen patients from this trial were analysed in the pooled safety population.

In response to a request by the ERG, the company provided a scenario analysis including 7 paediatric patients from the STARTRK-NG who had reached 6 months of follow-up by the latest data cut ([redacted]). This group included four patients with primary CNS tumours, two with sarcoma and one with malignant melanoma.

4.2.6 *NTRK* Efficacy Evaluable Analysis Set

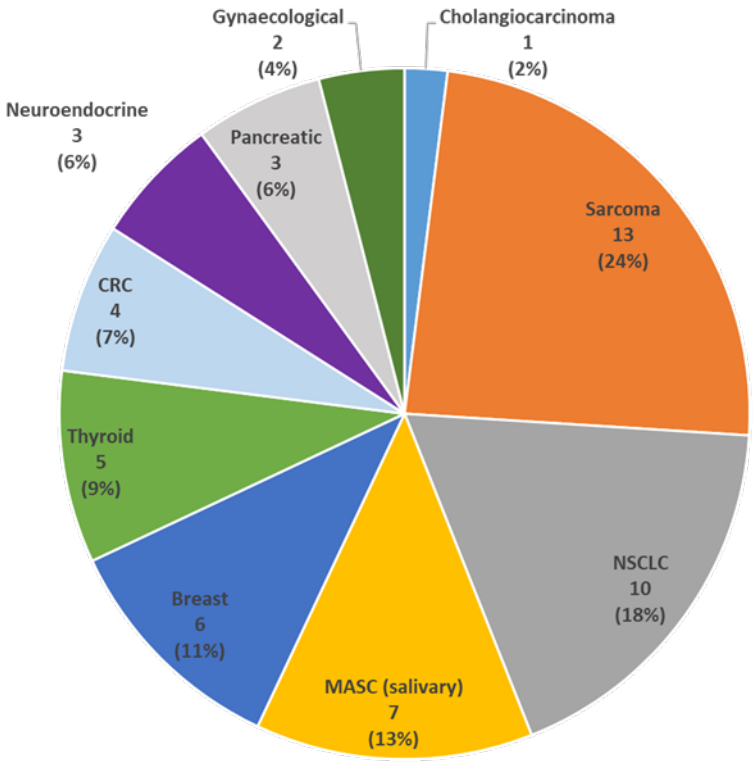
The primary source of efficacy data used in the company's efficacy analyses was based on the *NTRK* Efficacy Evaluable Analysis Set, which included patients derived from the four entrectinib trials who met the criteria described in Section 4.1.5, and excludes paediatric patients and those with primary CNS tumours. The company provided further analyses including 5 patients with primary CNS tumours and 7 paediatric patients in response to a request by the ERG (see section 4.2.6.1).

The demographics and baseline characteristics of the EEA population are summarised in Section 3.1 (Table 7).

As discussed in Section 3.1, there was a high degree of heterogeneity in the numbers of previous therapies received by patients within tumour types, making it likely that patients within subgroups were at different stages of the treatment pathway with varying disease history and prognosis. The ERG considered it unfeasible to conduct formal analyses on such data due to very small numbers of patients with such a high degree of heterogeneity in characteristics and response. The effect of heterogeneity across different tumour types is further explored in Section 4.3.

Figure 1 illustrates the distribution of cancer types included in the EEA dataset. The most represented solid tumour types were 'sarcomas' (24.1%), NSCLC (18.5%), salivary gland tumours (13.0%), and breast cancer (11.1%). The majority of patients had gene fusions involving *NTRK1* (40.7%), and *NTRK3* (57.4%), while only one patient was included with an *NTRK2* gene fusion. The company suggested in their clarification response that this was simply due to a lower absolute prevalence of *NTRK2* gene fusions, which comprise only [redacted] of *NTRK* fusions. However, the ERG were concerned that the distribution of tumour types and gene fusion types in this patient population did not closely match or represent that expected in the NHS population, an issue discussed further in Section 3.1.

Figure 1 Tumour types included in the EEA dataset (n=54) (CS Fig. 7, Page 46)



CRC, colorectal cancer; MASC, mammary analogue secretory carcinoma; NSCLC, non-small cell lung cancer

4.2.6.1 Summary of clinical efficacy results

This section presents a critical summary of efficacy results presented by the company for the *NTRK* fusion trial population, including the EEA dataset, CNS primary and paediatric population.

Overall, the trial evidence showed a clinically meaningful objective response rate (ORR) ([redacted] in EEA dataset, CNS primary and paediatric population) including in patients with CNS metastases at baseline. However, there is considerable uncertainty regarding the extent to which the high response rates observed translate into clinically meaningful survival benefits. The ERG identified a number of important issues, particularly due to the significant immaturity of the PFS and OS data.

Despite substantial censoring and the small number of patients at risk in the tails of the Kaplan-Meier curve, the crossing of OS curves between entrectinib responders and non-responders is of some concern. Due to limited data there is considerable uncertainty about the precision of response and survival benefit estimates and heterogeneity by tumour type and line of therapy. Due to the lack of control group in the entrectinib trial evidence, the relative clinical benefits of entrectinib compared with relevant alternative cancer therapies are highly uncertain.

Response rate

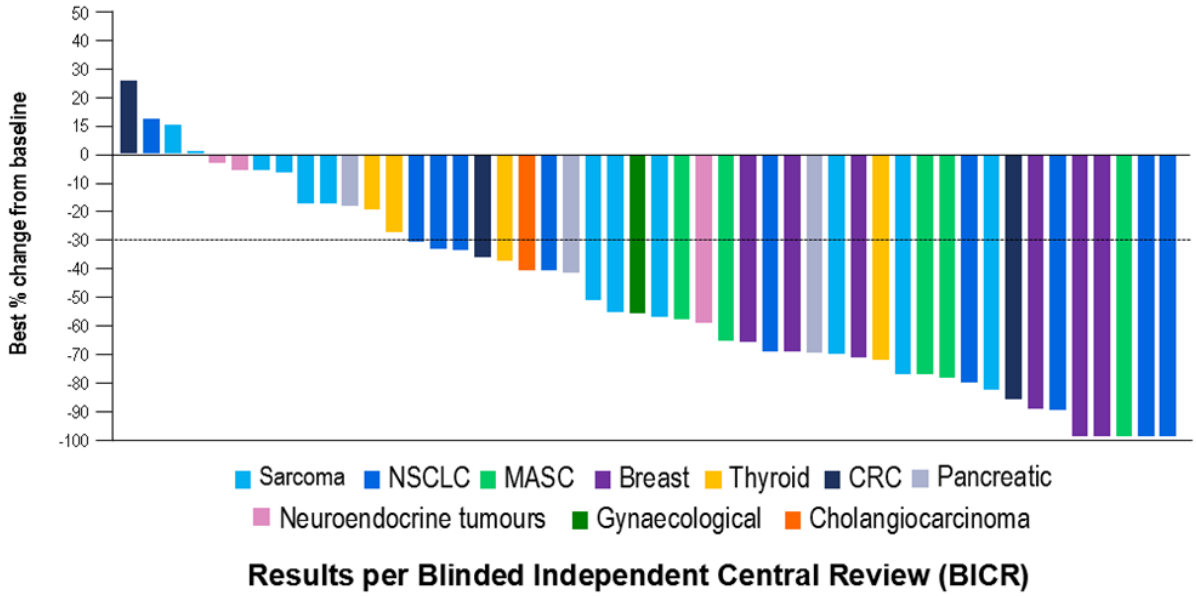
Objective response was assessed according to RECIST 1.1, assessed independently by blinded independent central review (BICR, primary analyses) and by the investigator. Response was defined as partial response (PR) or complete response (CR) confirmed by repeat imaging at least 28 days following first documentation of response. Table 13 presents the overall response rate (ORR) and best overall response for the efficacy evaluable population (data cut-off [redacted] BICR). Objective response was achieved in a high proportion of patients (57.4%, 95% CI: 43.2% to 70.8%). [redacted] achieved CR and [redacted] had a PR. Disease progression was found in four patients (7.4%). Investigator-assessed response rate estimates were consistent with the BICR (53.7%, 95% CI: 39.6% to 67.4%).

Table 13 Objective response rate and best overall response (efficacy evaluable population, data cut-off [redacted], from CS table 13)

	N (% of 54)
Responders	[redacted]
95% CI for response rates	[redacted]
Non-responders	[redacted]
Complete response (CR)	[redacted]
Partial response (PR)	[redacted]
Stable disease (SD)	[redacted]
Progressive disease (PD)	[redacted]
Non-CR/PD	[redacted]
Missing or unevaluable	[redacted]

Figure 2 presents individual patient responses measured as best percentage change from baseline in sum of longest tumour diameter (SLD) for the efficacy evaluable population. The 30% line of best percentage change corresponds to the RECIST 1.1 definition of partial response. Six patients from the efficacy evaluable population had missing SLD % change and were excluded from this plot. The company stated that response was observed across tumour types. Although the ERG agrees with this statement, no clear trend in response by tumour type can be inferred from visual inspection of this plot due to the small sample size and large number of subgroups.

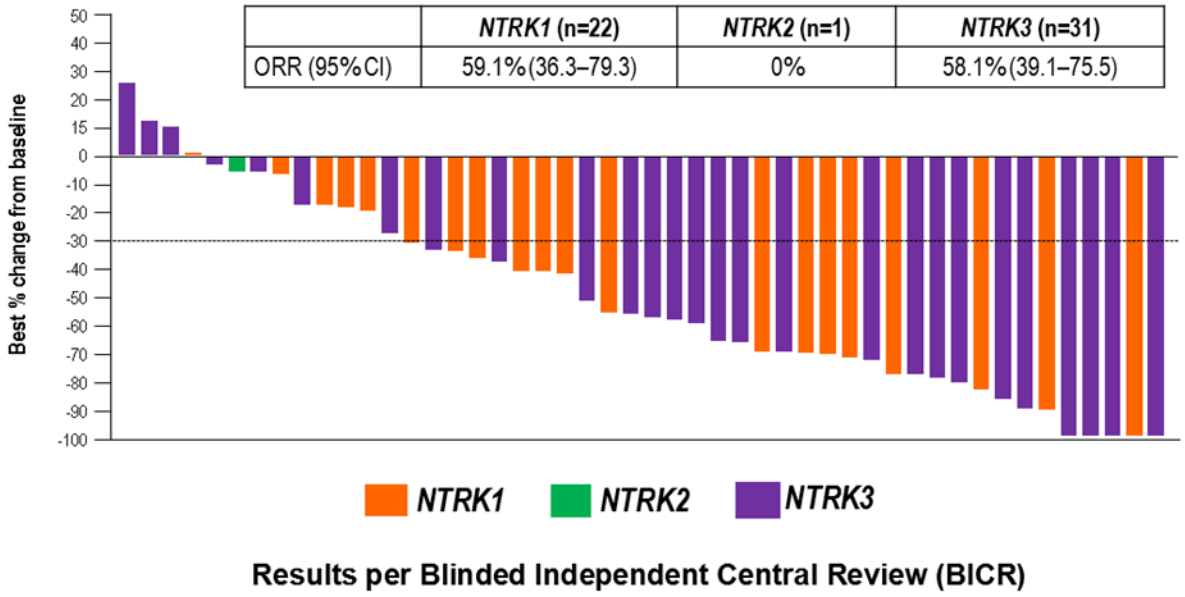
Figure 2 Entrectinib activity in *NTRK* fusion-positive solid tumours: individual patient responses by tumour type, BICR assessment - data cut-off 31 May 2018 (efficacy evaluable analysis, N=48*, from CS Figure 8)



*6 had missing SLD % change

Figure 3 presents individual patients' response for the efficacy evaluable population. This shows similar ORR for patients with *NTRK1* fusion (59.1%; 95% CI 36.3-79.3) and *NTRK3* (58.1%, 95% CI 39.1-75.5). The only patient with *NTRK2* fusion did not respond. The company stated that responses were independent of the *NTRK* fusion gene. The ERG believes this interpretation to be highly uncertain due to the lack of evidence for *NTRK2* fusions and small size of the *NTRK1* and *3* subgroups.

Figure 3 Entrectinib activity in *NTRK* fusion-positive solid tumours: individual patient responses by *NTRK* gene, BICR assessment - data cut-off 31 May 2018 (efficacy evaluable analysis N=48*, from CS Figure 9)



*6 had missing SLD % change

Figure 2 and Figure 3 exclude CNS primary and paediatric patients. In response to an ERG request, the company provided individual patient response data by line of therapy and tumour type for 66 patients, including the EEA population, as well as five adult primary CNS tumours patients, and seven paediatric patients (four primary CNS tumours, two sarcoma and one malignant melanoma) from the [redacted] clinical cut-off date. For primary CNS tumours, response was measured according to different criteria (Response Assessment in Neuro-Oncology Criteria, RANO) than other included tumours (RECIST v1.1), therefore the ERG agree with the company these results should be interpreted with some caution.

Detailed results assessed by BICR and investigator are presented in tables 1 and 2 and figures 1 and 2, in the company's clarification response (entrectinib clarification response 04072019K). Table 14 summarises response rates by line of therapy for 66 patients for EEA population, adult primary CNS tumours patients, and paediatric patients. Response data were extracted from the Company's clarification response (entrectinib clarification response 04072019K). BIRC-assessed response data by tumour type based on the more recent [redacted] clinical cut-off date (Company's clarification response Table 1) were used.

In the adult primary CNS tumour population, investigator-assessed response data are available for the five patients. However, the BIRC data only include one primary CNS tumour patient, as BIRC data from the four STARTRK-2 adult primary CNS patients are not available. In the paediatric population,

only investigator-assessed response data are available. Investigator-assessed response rates are provided in the Company's clarification response Table 2 and were used to impute response status where BIRC-assessed response is missing. However, patients were not listed in the same order in Tables 1 and 2 of the Company's clarification response, and no patient identification numbers were provided. Therefore, imputation was done by carefully matching the missing patients by tumour type and line of therapy, checking that a value for the correct patient was imputed. The ERG is confident that imputation was adequate. Table 14 shows high response rates across first, second and third line therapy and beyond. ORR and CR rate were higher in patients receiving entrectinib as 1st line therapy [redacted] than as 2nd line [redacted] and third or subsequent line [redacted], although these findings are based on small subgroups.

Table 14 Response rates by and line of therapy (data cut-off [redacted]

	[redacted]	[redacted]	[redacted]	[redacted]
Response	[redacted]	[redacted]	[redacted]	[redacted]
CR	[redacted]	[redacted]	[redacted]	[redacted]
PR	[redacted]	[redacted]	[redacted]	[redacted]
No response	[redacted]	[redacted]	[redacted]	[redacted]

[#] if BICR was missing, data imputed from investigator assessment where possible. * Missing patients treated as non-responders

ORRs were [redacted] and [redacted] in patients with and without baseline CNS metastases, respectively (cut off [redacted]). As above, these results may not be reliable due to the small number of patients in each subgroup as reflected in the wide confidence intervals (11 patients had CNS metastases at baseline) and the exclusion of CNS primary patients from this subgroup analysis.

Duration of response

Table 15 shows that responses in the efficacy evaluable population were durable with a median DOR of [redacted] among the 32 responders, although this was subject to significant censoring, as [redacted] of the 32 responders had an event at the [redacted] cut-off.

Table 15 Duration of response, BICR assessment (efficacy evaluable analysis), data cut-off [redacted]

Pts included in analysis (Responders)	[redacted]
Pts with event (%)	[redacted]
Progressive Disease	[redacted]
Death	[redacted]
Median	[redacted]
95% CI for Median	[redacted]
25% and 75%-ile	[redacted]

Range	[redacted]

* Subject to censoring. NE: not estimable

KM data were also reported for earlier cut-off of 31st May 2018. A swimmer plot for the 31 responses in the *NTRK* efficacy evaluable analysis set is shown in CS Figure 11, although once again this was subject to significant censoring (15/31 censored).

Progression free survival

Table 16 and Figure 4 present Kaplan-Meier analyses results for PFS based on the BICR assessment in the EEA population (data cut-off [redacted]). The estimated median PFS was [redacted]. The ERG note that these results are subject to significant censoring, as only [redacted] patients had an event. In addition, these results only apply to the EEA population, and do not account for heterogeneity across tumour types. In response to a clarification request, the company stated they were not able to provide PFS data stratified by line of therapy or tumour type.

Table 16 Progression-free survival BICR assessment (efficacy evaluable analysis) – data cut-off [redacted]

	BICR-assessed PFS (n = 54)
Patients with event (%)	[redacted]
Progressive Disease	[redacted]
Death	[redacted]
Median PFS (95% CI)	[redacted]
25% and 75%-ile	[redacted]
Range	[redacted]

* Subject to censoring. NE: not estimable

Figure 4 Kaplan-Meier curve for BICR-assessed PFS (efficacy evaluable analysis) - data cut-[redacted] *Figure redacted*

The ERG requested further survival data stratified by response status from the company. PFS KM data was provided for the EEA with and without CNS primary and paediatric populations from the [redacted] cut-off. Table 16 and Figure 5 present KM results for the combined EAA, adult primary CNS and paediatric populations. This shows that [redacted] of these patients were responders, and [redacted] were non-responders. As expected, median PFS was higher in responders ([redacted] months) compared with non-responders ([redacted] months). The reliability of these results may be limited, notably due to the immaturity of the PFS data. From visual inspection of KM curves, approximately [redacted] of the trial population [redacted] were censored. In addition, these analyses are limited by the lack of adjustment for potential confounding factors between responders and non-responders, including differing baseline risk and use of subsequent therapy. However, in the

absence of direct comparator data, these provide a proxy for the potential magnitude of PFS benefits observed in entrectinib responders. Similar results were reported for the EAA population only.

Table 17 Progression-free survival BICR assessment (efficacy evaluable analysis + CNS primary adults and paediatric population) – data cut-off [redacted]

	[redacted]	[redacted]
Progressive Disease	[redacted]	[redacted]
Median PFS	[redacted]	[redacted]
95% CI	[redacted]	[redacted]

Figure 5: PFS Kaplan-Meier curves for responders vs non-responders - integrated analysis population plus primary CNS and paediatric patients data cut-off [redacted]

Figure redacted

Overall survival

Table 18 presents OS results for the EEA population with and without the CNS adult and paediatric population. At the cut-off date of [redacted], [redacted] had died. The KM estimated median OS for the total efficacy population was [redacted]. Although this is potentially clinically significant, these estimates are highly uncertain due to significant data immaturity. The extent to which OS is driven by the efficacy of subsequent therapies is also unclear. As discussed previously, [redacted] of the trial population received entrectinib as first line therapy, and [redacted] received subsequent cancer treatments. There is insufficient evidence to explore whether survival outcomes may have been greater in the first line population compared to patients further down the treatment pathway. The extent to which OS may vary by tumour type is also uncertain. In response to a clarification request, the company stated they were not able to provide OS data stratified by line of therapy or tumour type, but provided further OS data stratified by response status.

Table 18 Overall survival, BICR assessment (efficacy evaluable analysis set with and without CNS primary adults and paediatric population), [redacted] (from CS Table 22)

	Total n=54	Total n=66
Pts with event (%)	[redacted]	[redacted]
Median	[redacted]	[redacted]
95% CI for Median	[redacted]	[redacted]
25% and 75%-ile	[redacted]	[redacted]
Range	[redacted]	[redacted]

* subject to censoring

OS KM data was provided for the EEA with and without CNS primary and paediatric populations from the [redacted] cut-off. Table 19 and Figure 6 present KM results for the combined EAA, adults primary CNS and paediatric populations. Table 19 shows that at the [redacted] cut-off, the numbers of deaths recorded in responders and non-responders was [redacted] and [redacted] respectively. Median OS was [redacted] for responders, but was not reached in non-responders (survival was 50.1% at the last time point). The KM data and the survival curves presented in Figure 6 indicate that the OS benefit observed in responders ceases approximately at [redacted], at the point where the two survival curves cross. The ERG found this potentially concerning as it suggests there may be no long-term OS benefit for those who respond to entrectinib compared with those who do not. However, these OS data are immature. From visual inspection of KM curves, approximately [redacted] of the analysed population, including [redacted] of responders and [redacted] non-responders were censored; the survival curve for non-responders reaches a plateau at [redacted] at which point the estimated survival probability was still [redacted] The crossing of survival curves and substantial immaturity of the data mean that the longer-term OS benefit of entrectinib in this population is highly uncertain. KM data and survival curves were also reported for responder analyses excluding the adult CNS primary and paediatric population, and are presented in Figure 7 below. These data include [redacted] responders and [redacted] non-responders. These responder analyses [redacted] Again, this raises concerns about the true longer-term OS benefits of entrectinib in treatment responders, and emphasises the need for more mature survival data.

As discussed above, the lack of adjustment for potential baseline imbalances between responders and non-responders (as noted by the company) and other confounding factors including subsequent therapies means that these results may not be reliable. As reported in section 3.1, the company clarified that [redacted] of the EEA population received a wide range of subsequent cancer therapies, although these data are not broken down by response status. In addition, the responder analyses do not take into account the heterogeneity in survival by age, tumour type, line of therapy or presence of CNS metastases. The ERG agree with the company that uncertainties associated with these analyses are further compounded by the small number of patients, limited follow-up, exclusion of non-responders who died prior to outcome assessment, and the use of different definitions of response in CNS primary patients. However, a responder analysis approach avoids some of the significant limitations of the company's naïve comparison with external comparator data as discussed in section 3.3. In particular, all patients included in these analyses had an *NTRK* fusion, and the distribution of *NTRK1* and *NTRK3* fusions were similar between responders and non-responders in the EEA responder analysis (*NTRK* fusion subtypes were not reported for the adult primary CNS and paediatric populations).

The company noted that it cannot be assumed that entrectinib had no activity in patients classed as non-responders, and that in these patients tumour progression may have been temporarily halted or slowed down by treatment, thereby improving survival outcomes of the non-responder group. The ERG agrees that SLD reductions between 10% and 30% were observed for several patients across lines of therapy and tumour types (see response to clarification response 04072019K figure 1). Although it is theoretically possible that entrectinib may have improved survival results in patients not classed as responders according to the RECIST definition, the company did not provide evidence to support this, and the ERG believe this is unlikely to fully explain the positive survival outcomes observed in non-responders and the crossing of curves in the KM responder analyses. Therefore, the ERG believes there to be significant uncertainty about the longer-term survival benefits of entrectinib, regardless of response status or depth of response.

Table 19 OS BICR assessment (efficacy evaluable analysis + CNS primary adults and paediatric population) – data cut-off [redacted]

	[redacted]	[redacted]
Death	[redacted]	[redacted]
Median OS	[redacted]	[redacted]
95% CI	[redacted]	[redacted]
25% and 75%-ile	[redacted]	[redacted]

Figure 6 OS Kaplan-Meier curves for responders vs non-responders - integrated analysis population plus primary CNS and paediatric patients data cut-off [redacted]

Figure redacted

Figure 7 OS Kaplan-Meier curves for responders vs non-responders - EEA population (without primary CNS and paediatric patients) data cut-off [redacted] (from company follow-up clarification response 11072019KM, figure 2)

[redacted]

Figure redacted

To further explore uncertainty and heterogeneity in the survival estimates presented by the company, the ERG requested from the company individual patient data on PFS and OS by line of therapy and tumour type for the integrated analysis population, adult primary CNS tumours and paediatric populations. The company replied that this was not possible due to legal and governance reasons, although they noted they may be able to conduct further prospective analyses as required.

Health-related quality of life

HRQoL was evaluated for 51 of the 54 *NTRK* fusion patients included in the EEA population enrolled in the STARTRK-2 trial, and was assessed prior to the first dose of each cycle and at the end of treatment. The following questionnaires were administered: European Organization for Research and Treatment of Cancer Core Quality of Life Questionnaire (QLQ-C30) and EuroQol-5 Dimension (EQ-5D). Nine NSCLC patients also completed the lung cancer module (QLQ-LC13) and three patients with metastatic colorectal cancer (mCRC) completed the colorectal cancer module (QLQ-CR29). Completion rate for QLQ questionnaire was reported to be ≥80% at most study visits. Results were reported in the interim STARTRK-2 CSR (31st May 2018 cut-off) in tables and narratively.

The QLQ-C30 assesses five functional domains (physical, role, cognitive, emotional, social) and an overall global health status score. Baseline functioning scores were moderate-to-high for QLQ-C30 for global health status (69.79), physical functioning (74.17), role functioning (67.01), and cognitive functioning (84.72) on a score ranging from 0 to 100 with 100 reflecting better functioning. While receiving entrectinib, mean GHS scores were generally maintained or improved (-4.17 to 9.72). Physical functioning and role functioning scales results were moderate to high, with a trend towards clinical improvement. Cognitive functioning showed a negative trend (worst mean change score of -11.11 at Cycle 20 Day 1). Further results for QLQ-LC13 and QLQ-CR29 were reported in the STARTRK-2 interim CSR. Results for emotional and social functioning were not reported. Results of the EQ-5D questionnaire are discussed in section 5.2.7.

The ERG generally agree with the company's interpretation of the quality of-life questionnaire results, which suggest that overall general health functioning is not signigicantly affected duing entrectinib treatment. However, a reduction in cognitive functioning was observed, and there was no evidence on emotional and social functionning specifically. As above the reliability of the reported results is limited due to the small sample size.

Adverse effects of treatment

The CS provided adverse events data for 355 patients from three ongoing adult studies: ALKA (n = 57), STARTRK1 (n = 76), STARTRK2 (n = 206)) and one paediatric trial STARTRK-NG (n = 16). Patient safety data from these four trials have been pooled and analysed as the 'integrated safety population', with a data cut-off of 31st May 2018. Patients included in the integrated safety population were followed up for at least 6 months. The results for patients with < 6 months at the 31st May 2018 cut-off are reported separately. The integrated safety population includes adult and paediatric patients with *NTRK*, *ROS1* and *ALK* as well as paediatric patients with other/no known gene fusions. Table 20 presents the demographic characteristics of the integrated safety analysis.

Table 20: Baseline characteristics of the integrated safety population (adapted from CS Table 36)

	Adult Patients (N = 339)	Paediatric Patients (N = 16)	*NTRK* Adult Patients (N = 68)	Integrated safety population (N=355)
Sex, n (%)				
Male	151 (44.5)	10 (62.5)	31 (45.6)	161 (45.4)
Female	188 (55.5)	6 (37.5)	37 (54.4)	194 (54.6)
Median age, years (range)	55.0 (15, 86)	9.5 (4, 20)	57.5 (21, 83)	55.0 (4–86)
Age group, years, n (%)				
<65	249 (73.5)	16 (100.0)	43 (63.2%)	265 (74.6)
≥65	90 (26.5)	0	25 (36.8%)	90 (25.4)
Race, n (%)				
Asian	82 (24.3)	3 (18.8%)	9 (13.2)	82 (23.2)
White	222 (65.7)	13 (81.3%)	52 (76.5)	235 (66.4)
Black of African American	13 (3.8)	0	1 (1.5)	16 (4.5)
Other	5 (3.6)	0	0	5 (1.4)
Not reported	6 (4.4)	0	6 (8.8)	16 (4.5)
Mean BSA, m2 (SD)	1.79 (0.26)	1.07 (1.07)	1.83 (0.28)	1.76 (0.30)
Mean BMI, kg/m2 (SD)	24.79 (5.17)	17.33 (4.45)	25.12 (5.63)	24.45 (5.36)
ECOG PS, n (%)				
0	140 (41.3%)	0	26 (38.2)	140 (41.3)
1	170 (50.1%)	0	33 (48.5)	170 (50.1)
2	25 (7.4%)	0	7 (10.3)	25 (7.4)
3	3 (0.9%)	0	2 (2.9)	3 (0.9)
4	1 (0.3%)	0	0	1 (0.3)
Metastatic disease at baseline, n (%)				
Any site	NR	12 (75)	NR	311 (87.6)
CNS lesions	NR	0	NR	138 (38.8)

A summary of adverse events reported in the integrated safety population are reported in Table 21. [redacted]

Grade 3/4 AEs were reported for [redacted] of patients in the overall safety population. Treatment-related grade 3/4 AEs were found in [redacted] of the patients. Treatment-related serious adverse events were reported in [redacted] of the overall safety population. Deaths associated with adverse events were seen in [redacted] AEs leading to discontinuation of entrectinib were reported in [redacted] of the integrated safety population.

Table 21: All-causality and treatment Related Adverse Events (Integrated Safety Population, data cut-off [redacted], from CS Table 38, Page 80)

Adverse Event, No. Patients (%)	Adult Patients[a] (n=339)		Paediatric patients (n=16[c])		*NTRK* Fusion patients (n = 68)		All Patients (N=355)	
	All causality	Treatment-related[b]	All causality	Treatment-related[b]	All causaslity	Treatment-related[b]	All causality	Treatment-related[b]
Number of Patients:								
with AE	[redacted]							
with SAE	[redacted]							
with Grade ≥3 AE	[redacted]							
Adverse events associated with:								
Discontinuation	[redacted]							
Dose reduction	[redacted]							
Drug interruption	[redacted]							
Death	[redacted]							

AE, adverse event; SAE, serious AE

[a] Includes 68 NTRK, 134 ROS1 NSCLC and 137 other adult patients

[b] Treatment Related Adverse Events refer to adverse events that was considered by the investigator to be related to entrectinib treatment.

[c] Paediatric patients include 1 NTRK patient

[d] The ERG noted a small discrepancy in reporting of the proportion of all-cause Grade 3/4 adverse events (61.1% in the CS Table 38 and CS confidential docs 6.2.4.7 Table 9. 60.3% reported in CS page 81; CS confidential docs 6.2.4.7 and CS Appendix F. Table 19).

Table 22 presents adverse drug reactions (ADR) by organ class in the integrated safety population as reported in the Summary of Product Characteristics provided in the CS. [redacted]

Table 22: Adverse drug reactions, integrated safety population. (from CS Appendix C, Summary of Product Characteristics, Version 1, 10/2018. Table 5).

System Organ Class **Adverse Reaction**	**All Grades (%)**	**Grade 3 – 4 (%)**	**Frequency Category (All Grades)**
[redacted]			
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]

[redacted]			
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]			
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]			
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]			
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]			
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]			
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]			
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]			
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]			
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]

[redacted]			
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]			
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]			
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]			
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
ALT: Alanine aminotransferase. AST: Aspartate aminotransferase. Very common: ≥1/10, common: ≥1/100 to <1/10			

The ERG agrees with the manufacturer that the rates of AEs were broadly similar between adult patients across *NTRK, ROS1* NSCLC and other *ROS/ALK* patients.

The most frequently reported all-causality adverse events in the *NTRK* adult population were similar to those seen in the total integrated safety population.

[redacted]

The Summary of Clinical Safety reports the frequency of patients who experienced weight gain (Table 23); [redacted]

Table 23: Patients experiencing treatment-related weight gain (Integrated Safety Population, data cut-off 31st May 2018, from Summary of Clinical Safety (Roche Confidential Docs, 6.2.7.4) Page 61)

Adverse Event	No. Patients (%) [redacted]
≥ 5% weight increase	[redacted]
10 to ≥ 20% weight increase	[redacted]
≥ 20% weight increase	[redacted]

Overall, the ERG found that adverse events were generally well reported in the CS. The company did not extract safety data for comparator included in the NICE scope (see section 3.3). Due to the lack of comparator data, the relative safety of entrectinib compared with established management is highly uncertain. Rarer adverse events may not have been identified due the relatively small size of the safety population, particularly in paediatric patients.

4.3 Additional work on clinical effectiveness undertaken by the ERG

4.3.1 Exploring heterogeneity in response rates across tumour types

The ERG considers the company's assumption that such a variety of tumour types will have identical response rates when treated with entrectinib to be very strong and, as yet, untested. Therefore. an analysis of the potential heterogeneity in response rates across tumour types represented in the EEA, adult CNS primary and paediatric populations, and the additional uncertainty around this potential variability, was conducted.

Regardless of how the STARTRK trials were originally designed and analysed (Section 4.2), we can consider each of the tumour types as a "basket" or group and analyse the response data using a Bayesian Hierarchical Modelling (BHM) framework[39] to explore the potential heterogeneity in effects across tumours. Although originally developed as an adaptive trial design with stopping rules for unpromising treatments, we can ignore the adaptive phase and use the method to estimate posterior probabilities of response for each tumour type, as well as a pooled posterior probability of response across all tumour types, accounting for the potential lack of uniformity of effect across tumours. An additional advantage of this type of model is the ability to predict the response probability that would be expected in a "new" tumour type (i.e. a tumour that is not represented in the trial data), which will give a measure of the uncertainty in the response rates in tumour types in the target population but for which no data are available (see Appendix A).

4.3.1.1 Methods

For the response outcome, data available for each of the tumour types in the integrated analysis population, plus primary CNS tumours and paediatric tumours, are the number of responders, x_j, out of the total number of patients, n_j for tumour type j, which are assumed to follow a binomial likelihood

$$x_j \sim \text{Binomial}(n_j, p_j)$$

where p_j is the probability of response for tumour type j, with $j = 1,...,G$, and G is the total number of tumour types. We model the log-odds of response in tumour type j, θ_j, on the log-odds scale: $\text{logit}(\theta_j) = p_j$. The BHM assumes that for each of the G tumour types, the log-odds of response, θ_j, are exchangeable and follow a Normal distribution

$$\theta_j \sim \text{Normal}\left(\mu, \sigma^2\right)$$

where σ is the standard deviation quantifying the between-tumour heterogeneity and μ is the pooled mean effect across all tumours. Prior distributions must be selected for μ and σ and are likely to have some influence on the posterior estimates,[39, 45] particularly when a small number of groups and patients per group are included.

A relatively conservative normal prior distribution for μ is used, centred around a probability of response of 0.3 (a log-odds of -0.8473) which is often considered as a promising response rate, with a variance of 10 across all tumour types. Sensitivity of results to a more favourable prior distribution where the prior probability of response across all tumour types is centred around a mean of 0.5 (i.e. a log-odds of 0) with the same variance.

The prior for the between-tumour heterogeneity standard deviation is specified as Uniform(0,5) which was found to be robust in a simulation study.[45] An Inverse Gamma(2, 20) prior distribution for the between-tumour variance had previously been proposed[39], which means the between-tumour precision has prior mean 0.10 and variance 0.005. Inverse-gamma prior distributions were found to lead to posterior distributions which are highly sensitive to the chosen parameters and are therefore not recommended in most cases.[45] For completeness we present the results obtained using this prior distribution for the base-case dataset in Appendix E.

We also calculate the probabilities that the response rate for each tumour type is at least 30% or at least 10%.

Because the tumour types included in the integrated analysis population, plus primary CNS tumours and paediatric tumours are not reflective of the full licensed indication (i.e. some tumour types are missing, see section 3.1), the predictive distribution for the response rate in a new tumour type is calculated to reflect the full degree of uncertainty both due to the sample size and the observed heterogeneity in effects across the observed tumours. The resulting distribution is the probability of response in a "new", i.e. unrepresented tumour type.

The model was adapted from Thall et al[39] and estimated using Markov chain Monte Carlo in OpenBUGS,[46] implemented in R[47] (version 3.6.0) using R2OpenBUGS[48] (version 3.2.3.2). Code and implementation details are presented in Appendix E.

Model fit was assessed by plotting individual tumour contributions to the residual deviance (in a well-fitting model these are expected to be close to 1) and by comparing the total residual deviance to the number of tumour types, G.

4.3.1.2 Description of included data

Response data were extracted from the Company's clarification response (entrectinib clarification response 04072019K). BIRC-assessed response data by tumour type based on the more recent [redacted] clinical cut-off date (Company's clarification response Table 1) were used. Where BICR data was missing, data were imputed where possible using investigator-assessed response data. Further details are reported in section 4.2.6.1. The number of patients and responses by tumour type are given in Table 24.

Table 24 Number of responders by tumour type ([redacted] clinical cut-off date, with imputed response status where BIRC-assessed response is missing)

Tumour ID	Tumour type	Number of patients (n)	Number of responders (x)
1	Sarcoma	[redacted]	[redacted]
2	NSCLC	[redacted]	[redacted]
3	CRC	[redacted]	[redacted]
4	Neuroendocrine tumours	[redacted]	[redacted]
5	Pancreatic	[redacted]	[redacted]
6	Gynaecological	[redacted]	[redacted]
7	Cholangiocarcinoma	[redacted]	[redacted]
8	MASC	[redacted]	[redacted]
9	Breast	[redacted]	[redacted]
10	Thyroid	[redacted]	[redacted]
11	CNS Primary	[redacted]	[redacted]
12	Paediatric CNS Primary	[redacted]	[redacted]
13	Paediatric (non-CNS)	[redacted]	[redacted]
	Total	[redacted]	[redacted]

The company advised that caution should be exercised in the interpretation of response for CNS tumours as it is measured according to different criteria than for systemic solid tumours (Section 4.2.6.1). Whilst the ERG agrees with this advice, it is still valid to assess the heterogeneity in response across all included tumours regardless of how response is defined, as the overall response rate is an important clinical result.

4.3.1.3 Results

Results for the base-case analysis, which includes all adult and paediatric tumours (Table 24), are presented in this section. The prior distributions used for the base-case analysis are

$$\mu \sim \text{Normal}(-0.8473, 10)$$
$$\sigma \sim \text{Uniform}(0, 5) \quad (1)$$

The BHM estimates moderate between-group heterogeneity (posterior median [redacted], on the log-odds scale) although there is considerable uncertainty 95% credible interval (CrI) ([redacted]) (Figure 8). This suggests that there could be considerable variability across tumour types, although the possibility of very little variability is also not ruled out.

Figure 8 [redacted]

[redacted]*Figure redacted*

The estimated mean response rate across all tumour types is [redacted] with 95%CrI [redacted]. This is similar to the response rate that would be obtained if the tumour types were all assumed to have identical response probabilities [redacted] 95%CrI [redacted], which is consistent with the company's submission. The 95% CrI for the response probability predicted for an unrepresented tumour type is wide (Table 25, Figure 9), meaning that this probability could be as low as [redacted], or as high as [redacted].

Table 25 Probabilities of response according to the BHM.

	Overall posterior probability of response		
	mean	median	95% CrI
Posterior probability of response	[redacted]	[redacted]	[redacted]
Predictive probability of response	[redacted]	[redacted]	[redacted]

Prior distribution for log-odds of response centred on a probability of 0.3;
Uniform prior distribution for the between-tumour standard deviation

Figure 9 [redacted]

[redacted] *Figure redacted*

The estimated probabilities of response for each tumour type are shown in Table 26. The effect of allowing borrowing of information across the tumour types is to shrink the observed response probabilities towards the pooled mean response probability in Table 25. Tumour types with few patients borrow more information than tumour types with more patients.

Table 26 Probabilities of response for all tumour types

	Tumour type	Observed response (%)	Estimated mean response based on BHM (%)	Prob of response rate at least 30%	Prob of response rate at least 10%
1	Sarcoma	[redacted]	[redacted]	[redacted]	[redacted]
2	NSCLC	[redacted]	[redacted]	[redacted]	[redacted]
3	CRC	[redacted]	[redacted]	[redacted]	[redacted]
4	Neuroendocrine tumours	[redacted]	[redacted]	[redacted]	[redacted]
5	Pancreatic	[redacted]	[redacted]	[redacted]	[redacted]
6	Gynaecological	[redacted]	[redacted]	[redacted]	[redacted]
7	Cholangiocarcinoma	[redacted]	[redacted]	[redacted]	[redacted]
8	MASC	[redacted]	[redacted]	[redacted]	[redacted]
9	Breast	[redacted]	[redacted]	[redacted]	[redacted]
10	Thyroid	[redacted]	[redacted]	[redacted]	[redacted]
11	CNS Primary	[redacted]	[redacted]	[redacted]	[redacted]
12	Paediatric CNS Primary	[redacted]	[redacted]	[redacted]	[redacted]
13	Paediatric (non-CNS)	[redacted]	[redacted]	[redacted]	[redacted]

Prior distribution for log-odds of response centred on a probability of 0.3; Uniform prior distribution for the between-tumour standard deviation

Figure 10 shows the posterior distributions of the probabilities of response for each of the 13 tumour types. Whilst all distributions overlap, the distributions of response for [redacted] and their 95% CrI (Appendix E) suggest that response rates [redacted] are plausible. These tumour types also have the lowest probabilities of having a response rate greater than 30% (Table 26).

Figure 10 Posterior distribution for the probabilities of response in each tumours type, including primary CNS and paediatric.

[redacted] *Figure redacted*

Three sensitivity analyses were carried out and presented in Appendix E.

1. To assess sensitivity of results to the inverse-gamma prior distribution for the between-tumour heterogeneity variance, as suggested by Thall *et al* [39]
2. To assess sensitivity of results to the use of a more favourable prior for the log-odds of response;
3. To assess sensitivity to excluding primary CNS and paediatric patients from the data.

4.3.2 Exploring heterogeneity in time-to-event outcomes across tumour types

Heterogeneity in time to event outcomes (PFS, OS) can be explored using the BHM in a similar way.[39] The model assumes a common parametric distribution for each tumour type, but with a different location parameter. Information on this parameter can be borrowed across the different tumours, according to an estimated heterogeneity parameter. The results from this type of model would be different distributions of PFS or OS for each tumour type which could be incorporated in the economic model in order to further explore how heterogeneity in outcomes by tumour type influences the expected ICERs.

Although the BHM can borrow information across tumour types, and is designed to allow inferences with few events per tumour type, it is unclear whether this type of model would provide useful results in this appraisal, given the immaturity of the survival data and the small number of patients in most tumour types. PFS and OS data were not available to the ERG by tumour type so the feasibility of this type of analysis could not be assessed. Nevertheless, as more data become available, this could be a useful way to determine the extent of heterogeneity in PFS and OS across the different tumour types, and would allow predictive distributions of PFS and OS to be used to inform the survival of patients with unrepresented tumour types.

4.4 Conclusions on clinical effectiveness

The CS efficacy evidence was supported by four uncontrolled basket trials that included a total of 66 patients with metastatic or locally advanced *NTRK* fusion positive solid tumours, including seven paediatric patients. Thirteen tumour types were included: sarcoma, NSCLC, MASC, breast, thyroid, CRC, neuroendocrine tumours, pancreatic cancer, gynaecological cancers, cholangiocarcinoma, CNS primary, infantile fibrosarcoma and paediatric melanoma. Each tumour type was represented by between one and 13 patients.

The ERG found that the intervention and outcomes presented in the CS evidence match the NICE scope. However, due to concerns about the large proportion of unrepresented tumour types, the under-representation of *NTRK2* patients, and the small sample size of the *NTRK* efficacy trial population, the ERG is concerned that the population presented in the evidence submitted is not representative of the population defined in the NICE final scope. A significant proportion ([redacted]) of trials patients received entrectinib as first line systemic therapy, and for some there appeared to have been ‘acceptable’ alternative standard therapies available.

The company adopted a pragmatic approach to identify PFS and OS comparator data for established management without entrectinib, by searching NICE pathways to identify NICE approved comparators for each of the tumour types represented in the CS efficacy evidence. Median PFS and OS from each tumour types were averaged and then pooled to calculate mean overall PFS and OS

across all tumour types, weighted by the prevalence of each tumour type within the trial population. The ERG found that the population included in the comparator trials is unlikely to match the entrectinib efficacy population, notably due to the unknown prevalence of *NTRK* fusion in most of the comparator evidence, and mismatch in lines of therapy within the treatment pathway. In the base case analysis, no attempts were made to adjust for differences in population characteristics between the entrectinib and comparator trial populations; comparisons were naïve and do not account for any potentially important prognostic factors. The ERG conclude that the methods used to identify, select and combine comparator data are inappropriate, and that the comparator data used to inform the company model is highly unreliable.

Overall, the trial evidence showed a clinically relevant overall response rate across tumour types. However, there is considerable uncertainty regarding the extent to which the response observed translate into clinically meaningful survival benefits. The ERG found a number of important issues, particularly due to the significant immaturity of the PFS and OS data. Despite substantial censoring and the small number of patients at risk in the tails of the Kaplan-Meier curve, the crossing of OS curves between entrectinib responders and non-responders is of some concern.

The ERG explored heterogeneity in response rates across tumour types using a Bayesian hierarchical model, which assumes the response probabilities are similar (i.e. exchangeable) across tumour types, rather than identical (the company's preferred assumption). The ERG's analyses found that response rates obtained were similar to those observed when equal response probabilities are assumed, although there was considerable uncertainty in the level of heterogeneity of response rates across tumour types. Therefore, the possibility that some tumour types could have response rates that differ significantly from the pooled [redacted] response rate cannot be excluded. Due to limited data there is considerable uncertainty about the precision of response and survival benefit estimates and heterogeneity by tumour type and line of therapy. The lack of control group in the entrectinib trial evidence means that the relative effectiveness and safety of entrectinib compared with relevant alternative cancer therapies are highly uncertain. Due to lack of appropriate data and the uncertainty in response rates, the efficacy of entrectinib in tumour types not represented in the company's trials is unknown.

5 Cost Effectiveness

This section focuses on the economic evidence submitted by the company and the additional information provided in response to the points for clarification. The submission was subject to a critical review on the basis of the company's report and by direct examination of the electronic model. The critical appraisal was conducted with the aid of a checklist to assess the quality of the economic evaluation and a narrative review to highlight key assumptions and uncertainties.

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

The CS describes the search strategies used to identify relevant cost-effectiveness studies for the treatment of patients with *NTRK* fusion–positive solid tumours. Full details of the search strategy used are provided in Appendix D of the CS.

5.1.1 Searches

The ERG considers the searches undertaken by the company to be appropriate. For details of the searches undertaken by the company, see Section 4.1.1.

5.1.2 Inclusion/exclusion criteria used for study selection

The inclusion/exclusion criteria are summarised in Table 27 (Appendix G) of the CS and follow the usual PICOS framework. In brief, the review included any economic analyses and systematic reviews of pharmacological treatments for patients with *NTRK* fusion–positive solid tumours. Articles were assessed by a single reviewer against each eligibility criteria and independently checked by a second reviewer. Any discrepancies between reviewers regarding the inclusion of studies were resolved by discussion.

The ERG considers that the inclusion/exclusion criteria applied were appropriate and likely to identify any relevant studies.

5.1.3 Studies included and excluded in the cost effectiveness review

A total of 2,645 studies were identified in the searches following de-duplication. Of these, 80 full text articles were screened for inclusion in the review. No studies were, however, found to meet the review inclusion criteria and as such no published evidence was identified on the cost-effectiveness of entrectinib. Supplemental searches conducted by the ERG also did not identify any studies on the cost-effectiveness of pharmacological treatments for patients with *NTRK* fusion–positive solid tumours.

5.1.4 Conclusions of the cost effectiveness review

In the absence of any previously published cost effectiveness studies in patients with *NTRK* fusion–positive solid tumours, the *de novo* analysis in the CS represents the most relevant evidence for the stated decision problem.

5.2 ERG's summary and critique of company's submitted economic evaluation

The company presented a *de novo* economic analysis comparing entrectinib with established management in 13 tumour types. The model estimates a single composite ICER considering cost-effectiveness across all 13 tumour types and does not attempt to estimate individual ICERs for each individual tumour types. A summary of the company's economic evaluation is presented in Table 27, with justifications for key aspects and signposts to the relevant sections of the CS. The ERG has considered the methods applied in the company's economic evaluation in the context of a detailed checklist, reported in Appendix F.

Table 27 Overview of the company's economic evaluation

	Approach	Source / Justification	Location in CS
Model	Cost-effectiveness (cost-utility) analysis: Entrectinib: A partitioned survival analysis (PartSA) approach. Established management: Based on a PartSA approach, but rather than deriving time in state from parametrically extrapolated curves fitted to KM data, derives mean time from extracted median PFS and OS estimates for comparator therapies.	PartSA approach allows modelling relevant outcomes (PFS, TTOT and OS). The model structure is stated to be in line with the NICE decision support unit guidance [49] Modelling of comparator data attempts to simulate the PFS and OS benefit of the comparator therapies.	Section B.3.2.2; p91-92
States and events	The PartSA model contains 3 states: progression free (PF), progressed disease (PD) and death.	The PartSA model health states partitions OS into states of interest pre and post progression, and on and off treatment.	Section B.3.2.2; p91
Comparators	Entrectinib was compared to established practice which was a pooled comparator consisting of chemotherapy regimens and BSC.	Established practice was modelled as a "simulated "chemotherapy comparator generated by averaging clinical outcomes derived from previous NICE appraisals. These were weighted according to the proportion of patients in the integrated analysis with each tumour type.	Section B.3.2.2; p94
Natural History	Based on partitioned survival model. Transitions for patients receiving entrectinib were based on the integrated analysis of the ALKA, STARTRK-1 and STARTRK-2 single arm trials. Transition for patients receiving established management were based on median PFS and OS data extracted from relevant NICE technology appraisals (TAs) and extrapolated assuming an exponential function.	PFS and OS estimates were modelled independently, with the proportion of progressed patients at each cycle, calculated as the difference between the OS and PFS curves. Comparator outcomes were based on mean PFS and OS.	Section B.3.2.2; p91
Treatment effectiveness	Clinical outcomes included PFS and OS. Entrectinib PFS and OS were extrapolated from pooled analysis of relevant patients from the ALKA, STARTRK-1 and STARTRK-2 trials using single standard parametric models.	In the base-case analysis, an uncontrolled and unadjusted comparison was established between the pooled patient level data from the three entrectinib single arm trials and pooled data from previous NICE appraisals used to estimate comparator outcomes.	Section B.3.2.2; p94 Section B.3.3.1; p94-97 Section B.3.3.2; p97-100 Section B.3.3.3; p100-103

	Approach	**Source / Justification**	**Location in CS**
	Comparator OS and PFS was extrapolated based on median PFS and OS extracted from relevant NICE appraisals assuming an exponential parametric survival function		
HRQoL	Pre-progression health state utilities for patients receiving entrectinib were estimated from EQ-5D-3L collected in the STARTRK-2 trial. Health state utilities in the comparator arm and post-progression health states for both treatment arms were based on a weighted average of utility values used in previous NICE TAs Utility decrements for adverse events were applied in scenario analysis only.	The health state utilities (PF and PD) were assumed to differ across treatment arms in the pre-progression health sate and assumed to be equal in the post progression health state. Scenario analysis was also undertaking assuming no utility difference between entrectinib and comparator arms of the model. EORTC and EQ-5D-3L data were collected at baseline; day 1 of each subsequent treatment cycle, and after treatment discontinuation. Questionnaires were also completed in the period after treatment discontinuation. Observed EQ-5D-3L responses were classified into three categories according to the patient treatment or progression status: 1) Base-line assessment (assessment prior to treatment start date) ([redacted]), 2) patients in PFS (after treatment star data but prior to disease progression ([redacted]), and 3) patients post-PFS (after IRC assessed progression) ([redacted]). Utility values were derived from the collected EQ-5D-3L values and assigned to the entrectinib pre-progression health states. Post progression data was not used as the reported mean was considered implausible. Pre- progression utilities for established practice and post progression values for both treatment arms were based on a weighted average of utility values reported in previous NICE TA's. These were weighted according to the proportion of patients in the integrated analysis with each individual tumour type. The specific source of individual utility values used in the model other than the source TA were not reported in CS. Utility decrements associated with adverse events relating to entrectinib and chemotherapy treatment were not included in the model as it was assumed that these were already captured in the trial-based utility values used. Scenario analysis was also undertaken incorporating additional disutilities associated with adverse events specifically associated with entrectinib treatment (weight gain).	Section B.3.4.1. p109-110 Section B.3.4.3. p111-112 Section B.3.4.4. p111-113

	Approach	**Source / Justification**	**Location in CS**
		All AE disutilities were applied as a one-off decrement applied to the first cycle of the model.	
Adverse events	Adverse events were included for grade 3/4 events occurring in ≥5% of subjects. Event rates were assumed to be identical for intervention and comparator arms with the exception of increase in weight which was assumed to only occur in patients receiving entrectinib.	Event rates were drawn from the integrated analysis of the ALKA, STARTRK-1 and STARTRK-2 trials. In the base-case analysis, the AE rates were applied in the model to estimate associated costs only. Scenario analysis was also presented in which rates were used to estimate treatment related disutilities.	Section B.3.5.3 p123
Resource use and costs	Cost categories were: • Treatment and administration costs • Subsequent therapy • Health state resource use and costs • Testing of *NTRK* status • AE costs	Drug and administration unit costs were sourced from BNF, and NHS reference costs. Resource use was informed by UK hospital chemotherapy protocols. A Patient Access Scheme (PAS) discount of [redacted] off the entrectinib list price has been agreed with NHS England. PAS are also available for a number of the comparator therapies, but are not included in the company's base-case analysis. To estimate health state resource use comparator therapies were classified into three categories (oral chemotherapy, single agent chemotherapy and combination therapy). Resource use for each category was drawn from recent TA's (TA515[50], TA520[51], TA476[52]). Unit costs were sourced from the Personal Social Services Research Unit (PSSRU) and NHS reference costs. Testing costs were based on current testing algorithms already used in practice and a proposed testing algorithm for *NTRK* fusions based on IHC followed by confirmatory NGS. Costs of individual tests were based on values used in TA406, costs cited in a Scottish science advisory council report and values elicited from five national genomic laboratory hubs. The costs of adverse events grade 3-4 with incidence ≥ 5% were included in the base-case. The cost of end of life care was included for the last cycle that patients were alive in the model and for both intervention and comparator.	Section B.3.5 p115-126

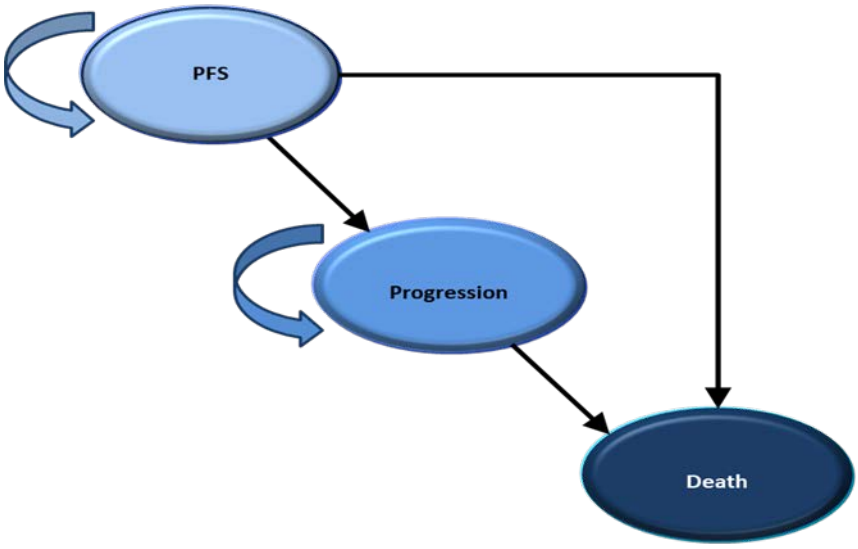
	Approach	**Source / Justification**	**Location in CS**
Discount rates	Costs and benefits were discounted at 3.5% per annum	In accordance with the NICE reference case.	Section B.3.2.2; p93
Sensitivity analysis	Probabilistic sensitivity analysis was performed. Deterministic univariate probabilistic analysis was performed on a series of model parameters. A series of scenario analyses was also performed.	In accordance with the NICE reference case.	B.3.8; p133-138

5.2.1 Model structure

The *de novo* analysis presented by the company compares entrectinib with established management. Established management consisted of a composite comparator represented through a weighted average of comparators from the tumour types represented in the integrated analysis for entrectinib, see Section 5.2.4

The cost-effectiveness analysis presented by the company is based on a partitioned survival model (PSM) or "area under the curve" analysis, depicted in Figure 11. It comprises three mutually exclusive health states: (i) PFS (progression free), (ii) progressive disease (PD; progression), and (iii) death. Within the PFS and PD health states, the model distinguished between patients who are receiving treatment and those who are not. The model predicted the total costs and QALYs separately for the entrectinib arm and the pooled comparator arm in order to estimate a single ICER.

Figure 11 Model structure (Figure 16 in CS)



Transitions between states are not explicitly incorporated into the analysis using probabilities. Instead the distribution of patients in each health state is determined by using estimates of PFS and OS.

For entrectinib, transitions were based on extrapolated KM data from the *NTRK* efficacy evaluable analysis set (Section 4.2.6). Time-to-event data for OS was used to determine the proportion of patients alive, while the proportion of patients in the PD state was calculated as the difference between OS and PFS. In scenario analysis time to off treatment (TTOT) was also used to determine time to discontinuation of treatment; in the base-case this was assumed to coincide with progression.

Transitions between states in the comparator arm were modelled using a different approach to that used to model the entrectinib arm. Time-to-event data were not used, instead estimates of mean OS

and PFS for each tumour type were modelled to estimate time in each health state. For each tumour type, time alive was estimated using mean OS, while time spent in the PD health state was estimated as the difference between the mean OS and mean PFS. These estimates of time in state were then used to estimate total costs and QALYs for each tumour type. Total costs and QALYs for the comparator arm were then estimated as weighted averages using the distribution of tumours in the integrated analysis of entrectinib (See Section 4.2.6) to determine the appropriate weight. As for the entrectinib arm, time on treatment for the comparator arm was assumed to align with PFS.

The cycle length used in the model was one week. Transitions between health states were assumed to occur at any time within the cycle. To account for the over- or under-estimation of transitions occurring at the beginning or end of the cycle, half-cycle corrections were applied to each time interval.

ERG comment

While the model structure is consistent with previous technology appraisals in advanced cancers, the ERG notes a number of issues, regarding the selection of endpoints from the clinical trials to define transitions between health states and the estimation of a single ICER for the full population covered by the marketing authorisation.

Choice of clinical endpoints to model health state transitions

As described above, the company's approach to estimating the cost-effectiveness of entrectinib follows the typical approach adopted in cancer appraisals of directly using extrapolated PFS and OS to populate a partitioned survival model. This approach, however, may not be a suitable model structure to leverage the available data for the present decision problem as it requires the availability of reliable, mature PFS and OS data for both the intervention assessed and the comparator. As outlined in Section 5.2.6, the available PFS and OS data for entrectinib and the reliability of the PFS and OS for the constructed comparator data set are severely limited, with concerns raised regarding the representativeness of the recruited population, uncertainties around positioning of entrectinib and potential confounding due to secondary therapy received.

The OS data from the efficacy evaluable analysis set is also immature with median OS not yet reached and as such, a significant proportion of the predicated benefit of entrectinib is based on the survival extrapolation. This issue is further exacerbated by the difficulty in validating predictions from the survival extrapolation analysis of entrectinib. It is good practice to assess the plausibility of the extrapolated portions of parametric survival models through the use of external data and clinical validity informed by clinical expert opinion and biological plausibility.[53] This is of particular importance in the present appraisal given the limitations of the OS data and lack of any external

datasets characterising the long-term prognosis of *NTRK* fusion positive patients. The context of this appraisal, however, makes this particularly challenging because the survival analysis is based on a population of patients with many different tumours types, which makes elicitation of clinical opinion particularly complicated.

Furthermore, while entrectinib has not yet been assessed by any regulatory bodies at the time of the ERG report, it is anticipated that the evaluation by these bodies will be based on a similar profile of evidence to that of larotrectinib,[54] which uses response outcomes (ORR and DOR) as the main regulatory endpoints. As the main regulatory endpoints in the larotrectinib evidence, it is these that formed the basis of the FDA decision to approve larotrectinib and is from these outcomes that they inferred the likelihood of clinical benefit, rather than using PFS or OS. The company's approach, however, ignores these outcomes and asks us to infer clinical benefit on the basis of PFS and OS, which appear to have been considered unsuitable for such a purpose in a regulatory setting, at least for larotrectinib. Alternative model structures built around response may therefore have been more suitable to address this decision problem and could represent a more robust approach upon which to predict long-term outcomes. Such an approach may also better lend itself to characterise uncertainty resulting from any heterogeneity in the treatment effect and therefore increase the opportunity to identify cost-effective subgroups as well as help focusing future data collection activities. In section 6 the ERG explores a response based model as alternative to the company's PFS and OS based approach.

The estimation of a single "full population" ICER

The model was designed to provide an estimate of a single "full population" ICER. This approach does not capture the heterogeneity in the patient population, and diverges from the Committee preference to date for a tumour type-specific treatment recommendation. The preference for making tumour type-specific decisions has been demonstrated in two previous NICE appraisals of interventions with a broad marketing authorisation, in which the Committee preferred to make tumour type-specific recommendations. In two appraisals of neuroendocrine tumours [55, 56] and of bone metastases from solid tumours and multiple myeloma [57], the NICE scope specified the consideration of the location of tumour or type of primary cancer, and the Committee deemed it appropriate to perform separate clinical and cost-effectiveness analyses given differences in prognosis and HRQoL associated with each of the tumour types. It is notable that in both of these appraisals, either separate studies were available for different tumour types or it was more feasible to undertake subgroup analyses than in the present appraisal.

Having a single "full population" ICER is not appropriate as the ICER is likely to differ across tumour types, and will be driven by a range of factors such as differences in treatment effect and comparator

effectiveness and comparator drug acquisition costs. Further, given the amount of heterogeneity associated with a histology-independent indication, estimating the average cost-effectiveness for the full patient population covered in the scope may not provide enough information to decision-makers about whether the drug is cost-effective across all subgroups.

While it is generally the view of the ERG that an optimised decision is preferable where possible because it increases allocative efficiency, it is acknowledged that this is more challenging in the present decision problem and an analysis of outcomes within each individual tumour type would not be sufficiently robust for decision making, since they would be based on very small patient numbers In this respect, a response based model, referenced above, may also confer advantages in accommodating any heterogeneity across the population because far fewer observations are required on response outcomes to draw meaningful conclusions about differences between tumour types, than would be required by time-to-event outcome such as PFS and OS.

5.2.2 The company's economic evaluation compared with the NICE reference case checklist

Table 28 summarises the ERG's assessment of whether the company's economic evaluation meets NICE's reference case and other methodological recommendations.

Table 28 Comparison of company's economic evaluation with NICE reference case

Attribute	Reference Case	Included in CS	Comment on whether *de novo* evaluation meets requirements of NICE reference case
Comparator(s)	Alternative therapies in the NHS, including those currently regarded as current best practice	Partly	There is significant uncertainty regarding the position of entrectinib within the patient pathway. It is therefore not clear whether the included comparators represent those patients would receive in clinical practice. Furthermore the modelled population does not cover all tumour types covered by the NICE scope and therefore comparator therapies relevant to these tumour types were not modelled.
Type of economic evaluation	Cost-effectiveness analysis	Yes	
Perspective - costs	NHS and PSS	Yes	NHS and PSS costs have been taken into account.
Perspective - benefits	All health effects on individuals	Yes	QALY benefits to treated individuals were considered.

Time horizon	Sufficient to capture differences in costs and outcomes	Yes	The economic model uses a lifetime horizon (30 years). Less than 0.001 % of patients are expected to survive beyond this period.
Synthesis of evidence on outcomes	Systematic review	Yes	
Outcome measure	QALYs	Yes	EQ-5D-3L was collected in the STARTRK-2 trial and used to populate utilities for patients receiving entrectinib in the pre-progression health state. Quality of life for patients receiving comparator therapies and either therapy in the post progression health state were based on a weighted average of utilities reported in previous technology appraisals
Health states for QALY measurement	Described using a standardised and validated instrument	Yes	Derived from EQ-5D
Benefit valuation	Time Trade Off or Standard Gamble	Yes	Time Trade Off
Source of preference data	Representative sample of the public	Yes	Societal tariffs from EQ-5D.
Discount rate	3.5% on costs and health benefits	Yes	Costs and benefits have been discounted at 3.5% per annum.
Equity weighting	No special weighting	Yes	No special weighting undertaken.
Sensitivity analysis	Probabilistic sensitivity analysis	Yes	Probabilistic sensitivity analysis was undertaken.

5.2.3 Population

The modelled population considered in the company's base-case was assumed to represent the population in the integrated efficacy analysis, with clinical evidence on the effectiveness of entrectinib drawn from this analysis, see Section 3.1 for details. The modelled population therefore includes the 13 tumour types represented in this analysis. The distribution of the tumour types in the integrated efficacy analysis is also used in the comparator arm of the model, as comparator effectiveness, utilities and costs for each tumour type are all weighted according to their distribution in the integrated analysis.

Modelled patient characteristics were mean body weight and mean height which were drawn from the integrated analysis. These were used to estimate mean body surface, used in the dose calculation for some chemotherapy regimens.

ERG comment

The ERG's concerns regarding the model population centre around a key, but implicit assumption of the company's economic analysis that the modelled population represents the wide, histology-independent, anticipated market authorisation. As outlined in Section 3.1, the ERG has several substantive concerns regarding the population recruited to the integrated efficacy analysis and the degree to which it represents the population potentially eligible to receive entrectinib therapy. These limitations include the distribution of tumour types, the exclusion of available evidence on CNS and paediatric patients, unrepresented tumour types and the underrepresentation of the *NTRK2* gene fusion population. Each of these issues is discussed in turn in the sections below, with specific focus on the implications for the economic analysis presented.

Distribution of tumour types

The company's approach to constructing an established management comparator was to produce an average of clinical outcomes derived from NICE appraisals, weighted by the proportion of tumour types represented in the integrated analysis population. See Section 3.3 for a discussion of the comparator. The proportions used in the company's base case can be seen in Table 29.

The ERG is concerned that the estimated cost-effectiveness of entrectinib is being driven by the proportion of tumour types seen in the integrated efficacy analysis. In applying the distribution of tumour types observed in the trial, it is assumed that the trial population is reflective of practice, but this is unlikely to be the case. An alternative distribution is provided in Table 29. This is important because the prognosis and costs vary substantially across tumour types, in particular, the costs associated with screening for *NTRK* fusions can vary significantly.

The impact of alternative distributions of tumour types is also illustrated in scenario analyses whereby 100% weighting is given to a single tumour type. These show that the ICER varies significantly by tumour type. The result of this scenario show the ICER can range from £114,524 if 100% weighting is given to pancreatic cancer to £31,064 if 100% weighting is given to MASC. For further details, see Section 5.2.9.3.

Reflecting these concerns, the ERG queried the company regarding the representativeness of the distribution of patients across tumour types in the integrated efficacy analysis. The CS stated in their response that given [redacted] patients were screened for *NTRK* gene fusions, it is reasonable to expect that the

proportions of tumour types used in the base-case, may reflect the population seen in clinical practice, with the exception of MASC, which the company agreed is over represented.

Consideration of alternative data sources regarding the likely distribution of tumour types, however, undermines the company's stated position. A comprehensive data set of over 166,000 tumour samples along with the observed frequency of *NTRK*-gene fusions in specific tumour types was provided to the ERG in response to clarification questions. This information was used to estimate the number of patients in the population eligible for entrectinib, which in turn represents an alternative distribution of tumour types. The method used to estimate this can be found in Appendix B. This alternative distribution can be seen in Table 29.

Table 29. Distribution of tumour types in the entrectinib integrated efficacy evaluable analysis set and an alternative ERG distribution

Tumour Type	Proportion in CS	ERG
Sarcoma	24%	[redacted]
NSCLC	18%	[redacted]
MASC	13%	[redacted]
Breast	11%	[redacted]
Thyroid	9%	[redacted]
CRC	7%	[redacted]
Neuroendocrine	6%	[redacted]
Pancreatic	6%	[redacted]
Gynaecological	4%	[redacted]
Cholangiocarcinoma	2%	[redacted]
CS, company submission; FMI, Foundation Medicine Inc.; NSCLC, non-small cell lung cancer; MASC mammary analogue secretory carcinoma; CRC colorectal cancer		

To explore the impact of this alternative distribution on the results of the economic model, a scenario analysis is implemented in Section 6.

Exclusions from available evidence

The patient population considered in the base-case failed to encompass the entire population as defined in the NICE scope and was limited to the 10 categories of tumour types present in the entrectinib efficacy evaluable analysis set. The base-case population also excluded patients with primary CNS tumours and paediatric patients from the efficacy evaluable analysis set despite their eligibility for inclusion. The patient population considered in the base-case failed to encompass the entire population as defined in the NICE scope and was limited to the 10 categories of tumour types present in the entrectinib efficacy evaluable analysis set. The base-case population also excluded

patients with primary CNS tumours and paediatric patients from the efficacy evaluable analysis set despite their eligibility for inclusion.

The ERG believes patients with primary CNS tumours and paediatric patients should be included in the analysis as they fall within the population in which the company is seeking a recommendation and as such, requested a scenario analysis including these patients. The company response to this request was to highlight that the inclusion of paediatric patients in the economic analysis is challenging due to the absence of a counterfactual or at least any robust comparator data. The company also noted that patients with primary CNS tumours were excluded for a number of reasons including the lack of follow-up and that response was measured using Response Assessment in Neuro-Oncology Criteria (RANO). This is discussed in Section 4.2.6.1. However, the company reiterated that a NICE recommendation in accordance with the proposed license is anticipated, which includes all paediatric and adult patients harbouring *NTRK* fusion including primary CNS tumours.

The ERG is concerned that the company is seeking a recommendation in patients with primary CNS tumours and paediatric patients, yet despite data being available at the original CCOD for these patient groups, the company has decided to omit information from the base case provided in the CS due to differing response measurements when in fact response outcomes are not used in the economic model.

The company provided an updated economic model in the updated response to clarifications with the inclusion of the five efficacy-evaluable adult primary CNS tumour patients and the seven paediatric patients added to the model. The ERG welcomes the inclusion of these two additional patient populations into the economic analyses. See Section 5.2.9.4 for a discussion of the impact of the inclusion of these populations on the company's base case ICER.

Unrepresented tumour types

As highlighted in Section 3.1 an important limitation of the integrated efficacy analysis is that it does not include all tumour types covered by the anticipated marketing authorisation. For a list of those tumour types identified, see Table 2. This issue of unrepresented tumour types is potentially significant as based on current knowledge of which tumour types exhibit an *NTRK* fusion, a [redacted] will be covered by the anticipated marketing authorisation based on those tumour types in which an *NTRK* fusion has been identified. Furthermore, clinical advice received by the ERG has suggested that it is plausible that *NTRK* gene fusions could potentially be present in 400+ possible tumour types.

The impact of this omission in the modelled population is potentially very significant particularly given the weak support for the assumption of homogeneous response rates for the different tumours (Section 4.3.1), which suggests that different response rates may be observed in the missing tumour

types. Furthermore, this issue persists beyond the treatment effect and also impacts upon costs and utilities of both entrectinib and the comparator as the 10 tumour types present in the trial are being assumed to represent all of the potential tumour types in the population, which as outlined above is very unlikely to be the case. See Section 5.2.6 for a discussion of how the tumour types may impact on the treatment effect and costs respectively.

Underrepresentation of NTRK2

The efficacy evaluable analysis set includes 54 patients, with only one of these patients harbouring an *NTRK2* gene fusion. The ERG is concerned about the low representation of patients with *NTRK2* fusions in the trial and that the population used in the underrepresents this specific fusion type. At the points for clarification stage, the ERG queried the underrepresentation of these patients, with the company responding that the low number of recruited patients reflects the low prevalence of this fusion within the wider population. However, based on The Foundation Medicine Inc. data provided by the company in their clarification response, *NTRK2* patients may make up [redacted] of the *NTRK*-fusion positive tumours, much higher than the 2% included in the entrectinib integrated efficacy analysis.

There is insufficient evidence to establish whether patients with an *NTRK2* gene fusion have different prognoses to patients with *NTRK1* and -*3* gene fusions or whether there is potential for different responses to entrectinib based on *NTRK* fusion type. Clinical advice given to the ERG suggests that both of the scenarios are plausible and it was explained that it will depend upon the role *NTRK2* fusions play in the tumour growth within the tumour types they occur in. Further, data presented to the FDA as part of an NDA Multidisciplinary Review and Evaluation of larotrectinib in patients with *NTRK*-gene fusions, suggests that patients with *NTRK2* gene fusions had a lower overall response rate than those with *NTRK1* and -*3* gene fusions, which may suggest differential response to TRK inhibitors in this population.[54]

The under representation of this fusion type therefore presents additional uncertainty and a further problem with the representativeness of the population that makes up the integrated efficacy analysis. The size and direction of the consequences of this in the economic model are not fully apparent, but if, as suggested by the FDA, *NTRK2* positive patients are less likely to respond, it may lead to an overestimation of the treatment effect and consequently an underestimation of the ICER.

5.2.4 Interventions and comparators

The economic model presented in the CS compares entrectinib with established management which was assumed to consist of a blended comparator of chemotherapy regimens and BSC.

The modelled dose of entrectinib was assumed to align with the anticipated recommended dose of entrectinib, which is detailed as 600 mg once daily, with each 600 mg dose administered as 3x 200mg capsules. Duration of treatment for entrectinib was assumed to be aligned with the anticipated marketing authorisation, i.e. [redacted] and in the base-case analysis was set equal to progression free survival. Scenario analysis was also presented where time on entrectinib treatment was based on observed time on treatment in the integrated efficacy analysis.

The modelled blended comparator consisted of chemotherapy regimens and BSC. This blended comparator was based on previous NICE TAs identified as providing relevant effectiveness data. As outlined in Section 3.3 comparator effectiveness data for each tumour type was generated from multiple TAs therefore the modelled comparator was blended both at the individual tumour type level as well as at the across tumour types. The active comparators consisted of combination of alkylating agents (oxaliplatin, trabectedin), antimetabolites (capecitabine, fluorouracil, gemcitabine), anti-tumour antibiotics, multi-kinase inhibitors (Nintedanib), topoisomerase inhibitors (irinotecan), mitotic inhibitors (docetaxel, paclitaxel) and therefore cover a wide range of agents. Dosing of comparator therapies was based on their Summary of Product Characteristics (SmPC) guidance with duration of therapy based on PFS, i.e. treatment until either progression or death.

A list of comparators for each tumour type is presented in Table 30. Note that the comparators listed reflect those used to generate comparator effectiveness as comparator costs were based solely on active comparators. The modelled effectiveness data was therefore inconsistent with the modelled comparator costs, see Section 5.2.8 for further discussion of comparator drug acquisition costs.

Table 30 Summary of comparators modelled and data sources

	Therapy	**Reference**
Non Small-cell Lung Cancer	Docetaxel	Average of values from NICE TAs 520, 428, 483, 484, 403, 347, 124
	Docetaxel + nintedanib	NICE TA347
Colorectal Carcinoma	FOLFIRI	NICE TA307
	Irinotecan	NICE Guideline CG121 - Kim et al 2009
	Trifluridine-tipiracil	NICE TA405
	Trifluridine/tipiracil	NICE TA405
	Best supportive care	NICE TA405
	Best supportive care	NICE TA405
Breast Cancer incl. secretory breast	Capecitabine	NICE TA515
	Eribulin	NICE TA423
	Vinorelbine	NICE TA423
	Gemcitabine + paclitaxel	NICE TA423
Salivary Gland Cancer (incl. MASC)	Best supportive care (Platinum Gemcitabine data used as surrogate)	Surrogate data for BSC - Laurie et al. 2010
Soft Tissue Sarcoma	Doxorubicin	NICE TA465
	Trabectedin	NICE TA185
Pancreatic	Gemcitabine & nab-paclitaxel	NICE TA476
	Gemcitabine	NICE TA476

	FOLFIRINOX	NICE Guideline NG85 - Conroy et al 2011
Thyroid (papillary), unsuitable/refractory to radioactive iodine	Best supportive care	NICE TA535 (Cross-over adjusted value from Guo et al 2015)
	Best supportive care	NICE TA535
Neuroendocrine tumours	Everolimus	NICE TAs 449 and 539
	Everolimus	NICE TAs 449 and 539
	Best supportive care	NICE TAs 449 and 539
	Best supportive care	NICE TAs 449 and 539

In addition to the above, the economic model also allowed for subsequent therapy following discontinuation of entrectinib treatment; no subsequent therapy was assumed for patients receiving established management. Subsequent active therapy was assumed to be received by [redacted] of patients based on the proportion of patients with progressed disease who received subsequent therapy in the integrated efficacy analysis. Subsequent therapies were assumed to consist of established management as defined above. Patients receiving subsequent therapy were assumed to do so from the time of progression until death. Scenario analysis was also presented in which different rates of subsequent therapy used were assumed in the entrectinib model arm (50% and 80%) as well as a scenario in which 50% of chemotherapy patients were assumed to continue therapy post-progression. See Section 5.2.9.3 for the result of the scenario analyses.

ERG comment

As discussed in Section 3.3 significant uncertainties existing regarding the positioning of entrectinib in the patient pathway because the anticipated marketing authorisation allows entrectinib to be used at several points in the treatment pathway. Table 6, reported on page 30 of the CS provides some indication of where the company anticipate entrectinib will be positioned in UK practice. However, this table does not cover all of the tumour types represented in the integrated analysis or the additional tumour types know to harbour *NTRK* fusions, see Table 2 Tumour types included in ERG population size calculations. As outlined in Section 3.1 there are also some concerns about whether the indicated positions proposed in Table 6 of the CS, reflect the likely positioning of entrectinib. This uncertainty in the positioning of entrectinib means that it is not possible to validate the selected comparators as it is not clear at what position entrectinib will be positioned in the respective pathways.

The company's approach to identifying comparators also does not help to provide clarity regarding which are the appropriate comparator because for each tumour type, multiple TAs spanning multiple lines of therapy have been selected. The ERG considers this a logical inconsistency and that at least for the purposes of generating an externally valid comparator it would have been preferential to instead select a single line therapy to represent the anticipated positioning of entrectinib. Furthermore, while the ERG has not been able to fully validate the selected comparators for every tumour type, there are clear examples of where comparators have been selected that are rarely used in UK practice.

For example, with respect to the third line treatment of breast cancer, gemcitabine plus paclitaxel is rarely used while in colorectal cancer, irinotecan is less likely to be administered than FOLFIRI as second line therapy.

A further issue relates to the representativeness of the modelled comparators which, in principle should not only represent the trial population, but the wider population covered by the marketing authorisation i.e. all patients with *NTRK* fusions. As discussed in Section 3.10, the distribution of tumour types is not fully reflective of the eligible population with some types over/under represented in the trial population as well as there being a significant number of missing tumour types that may represent 20% of the population potentially eligible for treatment with entrectinib, based on those tumour types in which an *NTRK* gene fusion has been identified. See Table 6 for a list of these tumour types.

The validity of the modelled comparators is therefore subject to significant uncertainties, and at least for a number of tumour types, the selected comparators do not represent current UK practice, see Section 3.3. Furthermore, even ignoring these issues, the modelled comparators are only appropriate to the degree that they are representative of tumour types missing from the integrated efficacy analysis, and there is no reason to believe that this is the case.

In addition to the above issues, the ERG also has concerns regarding the modelling of subsequent therapies both with respect to the duration of subsequent therapy assumed in the company's base-case model and the mix of therapies assumed.

With respect to duration of therapy, the base-case analysis assumes subsequent therapies are received from progression until death. The ERG, however, considers that this assumption is likely to be overly pessimistic and that many patients will move to BSC before death. This may reflect either exhaustion of treatment options or lack of fitness to continue to receive therapy. The impact of this assumption is to decrease costs in the entrectinib arm of the model and therefore to substantially decrease the ICER.

At the points for clarification stage the ERG requested that the company provide further analyses in which patients were assumed to discontinue subsequent therapy prior to death. In response to the ERG's request the company provided two scenarios: one in which patients continued on subsequent therapy for three months and a second in which they continued on subsequent therapy for six months. The resulting ICERs from this analysis were £39,849 and £40,093 per QALY respectively, both of which are substantially lower than the company base-case of £52,609 per QALY. The ERG's preference is for a 6 month period of treatment following progression given that this represent roughly half the PPS, though the ERG acknowledges that this is a rather arbitrary assumption.

With respect to the mix of subsequent therapies used, the ERG considers the company's assumptions reasonable to the extent that entrectinib is likely to displace currently used therapy, but notes that this mix of therapies was very different to that received by patients in the integrated analysis which includes a wide range of therapies including several targeted therapies and immunotherapies, see points for clarification response question B10.

5.2.5 Perspective, time horizon and discounting

Consistent with the NICE methods guide [49] the company's analysis used NHS and Personal Social Services (NHS & PSS) perspective and discounted costs and benefits at a rate of 3.5%.

A lifetime horizon of 30 years was chosen as it was considered sufficient to capture all relevant differences in costs and benefits between comparators. The impact of shorter 5 year, 10 year and 15 year time horizons were also explored in a scenario analysis. The ERG considers this an appropriate time horizon, as it is very unlikely that any patients would remain alive beyond this time period.

5.2.6 Treatment effectiveness and extrapolation

As stated in Section 5.2.1, the company used a partitioned survival approach to provide a direct comparison of the timing and rates of progression and death. The main effectiveness inputs included in the company's economic model are therefore PFS and OS. For the model base case, OS and PFS survival estimates for entrectinib were drawn from the integrated analysis which pooled data from the ALKA, STARTRK-1, and STARTRK-2 trials. The integrated analysis set included 54 patients across 13 different tumour types, but excluded 6 patients with primary CNS and a paediatric patient. The CS states that the patients with primary CNS tumour were excluded from the integrated analysis because progression was measured by a different criteria, RANO rather than RECIST 1.1. The paediatric patient was excluded as the integrated analysis only includes patients from the adult studies, thereby excluding patients from the paediatric study STARTRK-NG. See Section 3.1 for further discussion of the population and critique of these exclusions.

The data-cut off used in the economic model was the 31st of May 2018; note this differs from the clinical section of the CS submission which presents data from the later [redacted] cut off. At the points for clarification stage the ERG requested that this latest data cut be integrated into the economic model which was provided by the company in their response. Results from this updated cut are presented in Section 0.

For the comparator therapies PFS and OS outcomes were sourced from previous TAs identified by the company to represent established management in the NHS for each of the tumour types modelled. A list of the source TAs used by the company to model comparator PFS and OS is presented in Table 30 (Section 5.2.4). For each tumour type median PFS and OS was extracted for all relevant TAs and then

a simple average of median PFS and OS estimated for each tumour type. Mean PFS and OS used to estimate time in pre-progression and post-progression health states was then estimated for each tumour type by assuming PFS and OS followed an exponential survival function.

Figure 12 illustrates the KM curve and extrapolated exponential OS curve for entrectinib (using the latest data cut, [redacted]) along with the average survival curves generated for patients receiving established management. Figure 13 illustrates the KM curve and various extrapolated OS curves presented in the CS using the latest data cut. Figure 14 and Figure 15 present similar data for PFS.

Figure 12 Kaplan-Meier, parametric extrapolated exponential OS for entrectinib and average survival for established management ([redacted] data cut)

Figure redacted

Figure 13 Kaplan-Meier and parametric extrapolations of OS for entrectinib (▌ data cut)

Figure redacted

Figure 14 Kaplan-Meier, parametric extrapolated exponential PFS for entrectinib and average survival for established management (▌ data cut)

Figure redacted

Figure 15 Kaplan-Meier and parametric extrapolations of PFS for entrectinib (▌ data cut)

Figure redacted

As can be seen above, available PFS and in particular OS data is immature, with ▌ and ▌ of patients having experienced a PFS and OS event respectively.

5.2.6.1 Uncontrolled comparison of treatment effectiveness

Generating an appropriate comparator dataset poses a significant challenge due to the histology-independent nature of this appraisal. As described above, the company's approach focuses on generating a comparator dataset using data sourced from previous TAs, which are then weighted by the distribution of tumour types in the integrated efficacy analysis. The principal concern regarding the company's approach is the fact that it relies on an unadjusted naïve comparison between the weighted comparator and the integrated efficacy analysis which has significant scope for confounding biases resulting from differences in population characteristics.

One potentially significant source of confounding results from differences in the number of lines of therapy patients have received. As highlighted in Section 4.2.6 a significant proportion of the patients in the integrated efficacy population (37.0%) received entrectinib as a first-line systemic therapy with a further 20.4% receiving entrectinib as a second line therapy. The comparator data set however, draws predominantly from patients in later lines of therapy, with 7 of the 10 tumour types, representing 57% of patients, including no evidence from patients receiving first line therapy. Further, a feature of the integrated analysis is that patients received entrectinib at different points in the clinical pathway including within single tumour types. For example, the integrated analysis includes NSCLC patients who received entrectinib as a first, second and third or later lines of therapy. The comparator data set generated, however, makes no account for this and does not attempt to match the comparator data used to the position patients are in the integrated analysis. This difference between the comparator data set and the integrated analysis is a potentially significant source of bias as line of therapy is an important determinant of prognosis. As a result, it is very likely that estimates of PFS and OS are confounded in favour of entrectinib.

A further and potentially important source of confounding bias is the fact that only a small proportion of patients in the comparator data set are likely to be *NTRK* fusion positive. This is problematic because there is evidence to suggest that *NTRK* fusions are prognostic, though available evidence is limited. To account for the potential prognostic value of *NTRK* fusions the company presents a scenario analysis which draws on evidence from patients with colorectal cancer. This analysis suggests *NTRK* is an indicator of poor prognosis and results in much lower ICERs than the base-case analysis. In the ERG's view, however, this scenario should be interpreted with caution as literature identified by the ERG suggests that the prognostic value of *NTRK* may be different across tumour types (Table 2). This was also confirmed by the clinical adviser to the ERG, who suggested that such variability in prognosis was possible, and likely dependent on the role *NTRK* fusions play in that specific tumour type.

In addition, there is also the possibility of differences in a wide range of other patient characteristics, such as age and ECOG status which are commonly prognostic. The CS does not report any baseline characteristics for the comparator arm and interpretation of these would have been complicated by the large number of tumour types and data sources. However, one approach the company could have taken to better understand the potential for differences is to have extracted commonly reported prognostic baseline characteristics such age, ECOG status and presence of brain metastasis from the source TAs. These could then have been used to generate a weighted set of baseline characteristics in a similar way to how the effectiveness data was generated. This could then have been compared to the integrated efficacy population. Such an approach could also potentially have been extended by implementing a Matching-adjusted Indirect Comparison (MAIC) or Simulated Treatment Comparison (STC) [58] to adjust the effectiveness data for entrectinib. Implementing such an approach would, however, have been very challenging not least because of the large number of source data sets and would make strong assumptions about the prognostic value of characteristics across tumour types. Furthermore, even if a suitable adjusted comparison could be generated it would only be able to account for a small number of observed characteristics due to the small sample size in the integrated analysis and therefore there would likely be significant residual confounding bias.

In summary, while the ERG considers that the broad approach adopted of using a weighted comparator data set to be reasonable, there are significant challenges associated with implementing this successfully, as well as further issues resulting from the company's execution of this approach. As such, while rectifying the specific issues highlighted above could potentially improve upon the validity of the comparator data set, it is likely that substantive concerns regarding the suitability of the comparator data would remain. Because of this, the ERG considers that company should have also considered alternative methods of generating comparator effectiveness estimates. The company could have for example considered two approaches discussed in a recent publication by Hatswell *et al.*[1] and

described below, which, while also subject to limitations, could have provided some degree of reassurance regarding predicted comparator effect estimates.

The first approach proposed by Hatswell *et al.*[1] uses effectiveness data on non-responders as a proxy for patients not receiving an active treatment. Comparator effectiveness estimates of PFS and OS under this approach would therefore be based on observed PFS and OS amongst non-responders in the integrated efficacy analysis. The rationale behind this approach is that patients in which no response is observed represent those with a lack of treatment effect (as they have no response to treatment) and therefore are representative of a counterfactual where no effective therapy exists. The advantages of this approach are that the patients are likely to be better matched with the intervention arm because they are drawn from the same population. However, this approach has disadvantages and makes a number of strong assumptions. It assumes that response is not systematically correlated with tumour type, which is unlikely to be true as is observed in analysis presented in Section 4.3. It also assumes that lack of response is indicative of comparator treatment effects which is likely to depend on the treatment considered as a comparator, but may be reasonable given the anticipated marketing authorisation which permits use of entrectinib [redacted].

The second approach outlined by Hatswell *et al.* [1] compares the outcomes for patients on entrectinib with their outcomes on the previous line of therapy. Under this approach the average time to progression (TTP) on the previous line of therapy is compared with average patients TTP when treated with entrectinib and a ratio estimated. The inverse of this ratio would then be applied to the observed PFS data from the entrectinib integrated efficacy analysis, to estimate comparator PFS, with PPS survival for both the entrectinib arm and comparator arm assumed to be same and also sourced from the integrated efficacy analysis. As with the first method, the advantage is that effect estimates are drawn from the same population as the intervention arm and therefore better matched, however there are also disadvantages. Firstly, this can only be implemented for patients who have received a previous line of therapy. Secondly, it also assumes that the ratio of TTP across lines of therapy is indicative of the treatment effect and it is uncertain to what degree this is likely to hold true. Finally, because this method can only estimate PFS it assumes that PPS survival is the same across therapies which similarly may not hold true.

To explore the uncertainties in the estimated treatment effect, the first method is implemented using non-responders as controls using data provided by the company at the points for clarification stage. See Section 6.5.1 for the results of this responder-based approach. The ERG also considered the second approach potentially valid, but did not feel that that this could be implemented within the time and resource available as the data requirements are significantly higher, but this could be considered as a reasonable scenario to be implemented at a later date.

5.2.6.2 Heterogeneity in treatment effect

A significant issue in the context of the present appraisal is the possibility for heterogeneity in the treatment effect across tumour types, as well as across other clinical characteristics such as age (paediatric vs adults), fusion type and position in the treatment pathway. This issue is however, largely neglected in the CS, with minimal analysis devoted to exploring the potential for heterogeneity in PFS or OS or indeed other measures of effectiveness. This is important as an implicit assumption in the company's base-case analysis is that the modelled treatment effect is constant not only across the modelled tumour types, but also across all tumour types covered by the marketing authorisation, see Table 6 for a list of the tumour types in which an *NTRK* fusion has been identified in the literature.

As demonstrated in the ERG exploratory analyses on response data (see Section 4.3.1), there is evidence to suggest that the treatment effect is heterogeneous across tumour types. These analyses showed that response outcomes for entrectinib vary considerably across tumour types ranging from [redacted] to [redacted], see Table 26, Section 4.3.1.3. Further, the predictive distribution, which provides an estimate of the likely response rate in an unrepresented tumour has a credible interval of [redacted] to [redacted] (Table 25, Section 4.3.1) implying that mean response across all eligible tumour types could be very different to that estimated in the integrated analysis. It is unclear how heterogeneity in response outcomes impacts on survival outcomes, and consequently cost effectiveness estimates, but these analyses do illustrate the potential for heterogeneity. The ERG presents further analyses in Section 6 exploring this potential heterogeneity in the treatment effect using a response based model to integrate the results of this analysis into the model.

5.2.6.3 Overall survival

Entrectinib

To extrapolate OS, standard parametric models (exponential, Weibull, log-logistic, lognormal, Gompertz, and generalised gamma) were fitted to the available KM data. To determine the most appropriate model, the CS states that reference was made to fit statistics (AIC/BIC; see Table 46 of the CS), visual fit to the observed KM curves, and clinical plausibility of survival estimates. Figure 16 provides a graphical summary of each curve and their fit to the observed KM data.

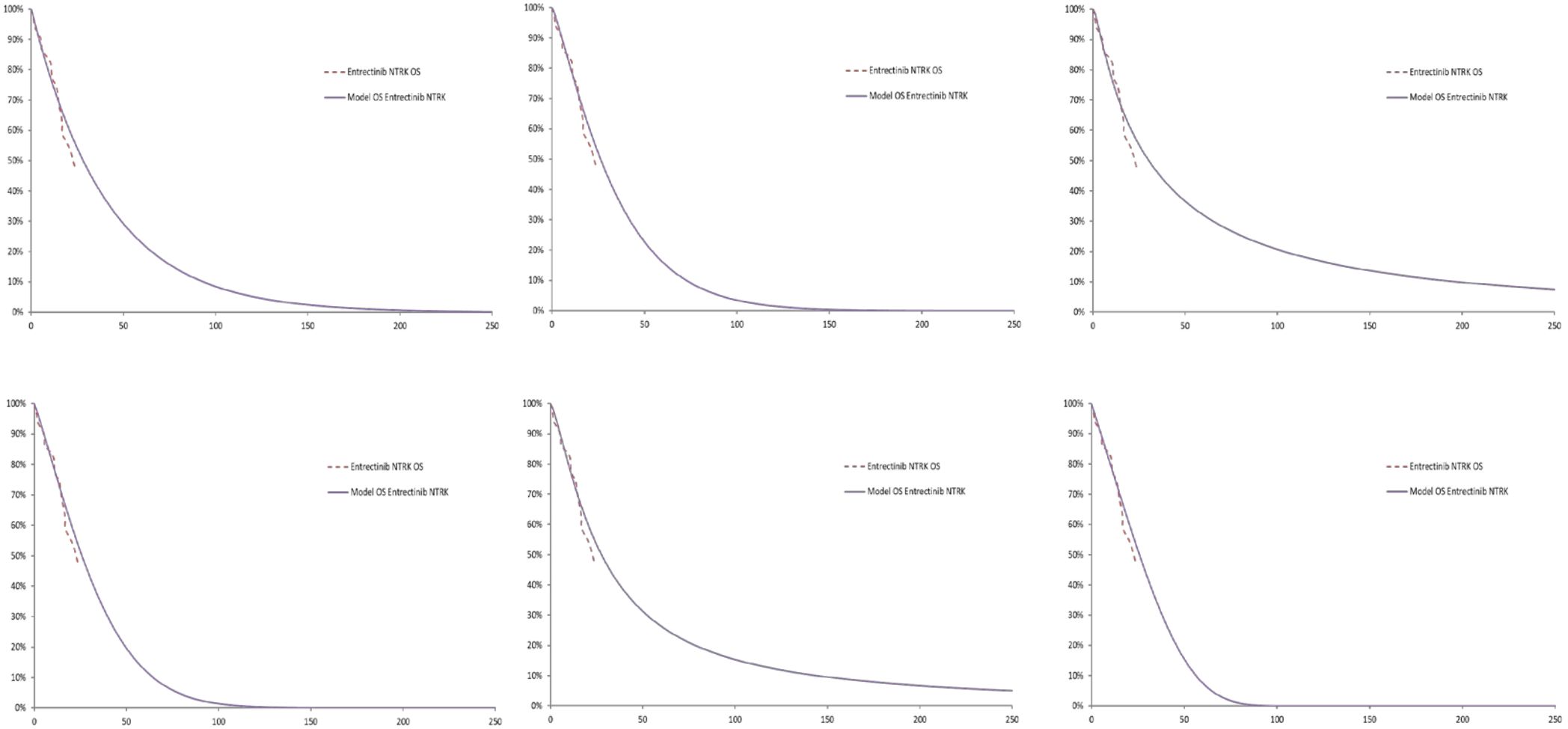
Consideration of clinical plausibility made reference to a landmark analysis of OS, which considered predicted proportion of patients alive at 2, 5, 10, 15, 20 and 30 years. This analysis is summarised in Table 47 of the CS. On the basis of the landmark analysis, only two of the curves were considered by the company to produce clinically plausible survival predictions. These were the exponential and Weibull curves. The base-case survival model presented in the CS selected an exponential curve and was justified on the basis that this had the best statistical fit, with all six other parametric curves considered in scenario analyses. The updated base-case based on the latest data cut ([redacted]) and presented

as part of the company's points for clarification response also retained the exponential function as company's preferred extrapolation. In their response the company cited statistical fit as the justification for selecting the exponential functions stating that it consistently had the best statistical fit across all scenarios.

ERG Comment

In the context of the present appraisal which combines so many tumour types, consideration of clinical plausibility of alternative extrapolations of OS is challenging. However, the ERG considers the exponential function selected to be a plausible extrapolation of OS, with good statistical fit to the observed OS data. The exponential function further makes reasonable predictions regarding long-term survival with most patients predicted to have died after five years and nearly all at 10 years. The ERG, however, notes that other survival functions similarly have good statistical fit while also making reasonable predictions about long term survival extrapolations, including the Weibull, Gompertz and Gamma functions. In considering the appropriateness of these individual functions the ERG notes that the exponential function is the only one to predict that post progression survival is longer than pre-progression survival. The ERG questions the clinical plausibility of this given that entrectinib therapy is discontinued on progression and that only [redacted] of patients received any subsequent therapy. Further, the positioning of entrectinib as a therapy of last resort suggests that few effective treatment options would remain to regain tumour control after progression, with consequences for post-progression mortality. The ERG therefore considers the Weibull, Gompertz and Gamma to represent a more reasonable extrapolations of OS with the ERG favouring the Weibull function due to its marginally better statistical fit over the other two functions.

Figure 16 Alternative entrectinib OS parametric curves (1st row, left to right: Exponential, Weibull, Log-normal; 2nd row, Gamma, Log-logistic, Gompertz, Exponential)



Comparator therapies (established management)

As described in Section 5.2.4, OS for comparator therapies was drawn from previous TAs and extrapolated assuming an exponential survival function. Because the company extracted only median OS values and not KM data, no other survival functions were explored. Survival predictions were, however, validated by clinical experts who "endorsed" the model predictions.

ERG Comment

The ERG has two main concerns about the modelling and extrapolation of the comparator data.

The ERG's first concern relates to the method used to pool median OS values from the NICE TAs selected to represent established management. Specifically, the ERG considers the approach of averaging median OS estimates at the individual tumour level to be inappropriate and mathematically incorrect. The company should instead have estimated mean OS for each TA and then pooled these. The impact of this calculation error is small, but is corrected in Section 6.

The ERGs second concern relates to the use of the exponential function to extrapolate OS. The ERG notes that this is a consequence of the approach taken by the company to identifying relevant effectiveness evidence, but considers it less than ideal. Examination of the source TA reveals that the exponential curve was rarely favoured by the committee in the considered appraisals, with the consequence that comparator OS is likely overestimated for some tumour types and underestimated for others. Furthermore, the estimates of post progression survival appear excessively long, with mean survival time post progression twice that of survival time prior to progression. In the context of the comparator data it is, however, unclear whether this is driven by underestimation of PFS or overestimation of comparator OS, though both of these will result in the ICER being overestimated.

The ERG further considers that other methods could have been adopted by the company to develop a comparator dataset, which would have greater face validity and flexibility. For example, the company could have extracted estimated life years gained from the committee's preferred scenario which would have accounted for the committee's preferred extrapolation. Alternatively, the company could also have extracted reported KM data from each TA, which would have given the company the flexibility to fit the best parametric extrapolation.

5.2.6.4 Progression free survival

Entrectinib

In common with the approach used for OS, PFS was extrapolated by fitting standard parametric functions to the available KM data with selection of an appropriate parametric function similarly

based on references to the statistical fit, reference to the hazard trend and to landmark analysis considering the clinical plausibility of predicted PFS at 2, 5, 10, 15 and 20 years.

On the basis of the landmark analysis depicted in Table 45 of the CS four of the parametric functions were considered clinically plausible. These were the exponential, Weibull, Gamma and Gompertz functions. The company base-case presented in the CS selected the exponential curve on the stated grounds that this represented a "conservative, but statistically and clinically plausible estimate of progression-free survival for entrectinib patients". As with OS, the updated base-case using the latest data cut ([redacted]) retained the exponential function as the company's preferred extrapolation, with statistical fit cited as the main reason for selecting this curve.

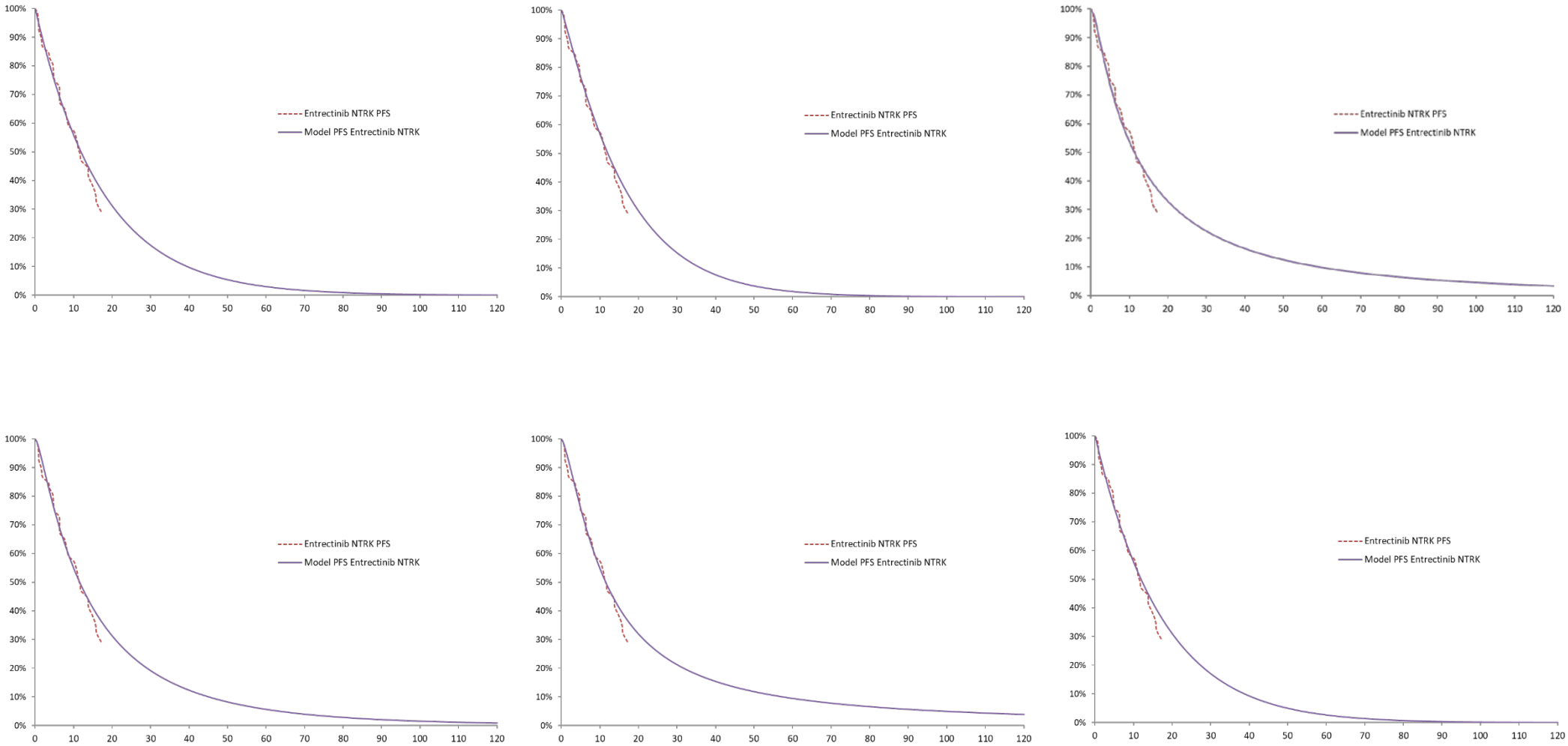
A graphical comparison of the extrapolations of PFS using the base case and alternative parametric function is presented in Figure 17 below.

The ERG considers that the exponential, Weibull, Gamma, and Gompertz functions all represent reasonable extrapolations of PFS and produce predictions that are consistent with the OS evidence when an exponential function is used; the ERG notes that after a certain time point, the Log-normal and Log-logistic functions yield estimates of progression that were higher than any of the plausible OS survival curves. The ERG, further notes that all four of these curves produce very similar estimates of mean PFS ranging from 15.85 months using the Weibull function to 17.65 months using the Gamma function. This small variation in predicted mean PFS and the relative insensitivity of the model to this input means that the ICER is relatively robust to the function adopted, changing by less than £1,000 per QALY across all four plausible extrapolations. The ERG considers the Weibull function as the preferred extrapolation function, as it is likely the most appropriate given its good statistical fit and for consistency with the ERG's preferences regarding the extrapolation OS; combining the Weibull function for OS with an exponential function for PFS implies a decreasing hazard for progression events which is clinically unlikely.

Comparator therapies (established management)

The company's approach to modelling PFS for patients receiving comparator therapies was the same as for OS and used median PFS values drawn from relevant previous TAs. These were then extrapolated assuming an exponential function with no other functions considered.

Figure 17 Alternative entrectinib PFS parametric curves (1st row, left to right: Exponential, Weibull, Log-normal; 2nd row, Gamma, Log-logistic, Gompertz, Exponential)



ERG Comment

The ERG's concerns regarding the extrapolation of PFS are largely similar to that of OS, with issues relating to both the method of calculation and limitations of assuming that PFS follows an exponential function. The impact of this uncertainty in comparator PFS is, however, not as significant as the uncertainties relating to OS as it is a less significant driver of the model. Importantly, however, because PFS determines time on treatment, increasing PFS actually decreases the ICER rather than increasing it. This is because the relatively small increases in QALY results from extending PFS are outweighed by increased drug acquisition costs.

5.2.6.5 Adverse events

Adverse events from treatment with entrectinib and established management were considered in the economic model to capture associated costs. Scenario analysis also considered the impact of AEs on quality of life by including disutilities. Only Grade 3-4 events were modelled and only if they occurred in >5% of patients. Four AEs were modelled: anaemia, fatigue, neutropenia and weight increase. Event rates for both the entrectinib and established management arms of the model were drawn from the integrated efficacy analysis, see Table 65 of the CS Page 123 for event rates. Because of the lack of data on event rates for AEs occurring in patients receiving established management, the model assumed identical event rates for both entrectinib and established management arms with the exception of weight increase which was assumed to only occur in patients who received entrectinib. The AE weight increase, was, however, associated with zero cost and therefore the assumption of differential event rate had no impact on the results of the base-case analysis. In effect, therefore the base-case analysis assumes no difference in AE rates between entrectinib treatment and established management.

The company considered this assumption to be conservative with respect to entrectinib as the company noted that many of the chemotherapy regimens patients would receive as part of established management have a poor toxicity profile.

ERG Comment

The ERG considers the use of the integrated analysis to model AE rates in patients receiving entrectinib to be reasonable and recognises the difficulty of modelling AE rates for patients receiving established management. However, the ERG considers the approach taken by the company to be less than ideal. Alternative approaches could have been considered for generating adverse event rates, such as using the source TAs to identify AE rates for the comparator therapies. Further, the company's assertion that the assumption of largely equal AE rates is a conservative one is not a certainty and will depend significantly upon the comparator being considered. For example, a number of the comparators listed by the company, see Table 30 above, consist of BSC and therefore patients

are receiving no active therapy. In such cases we may expect adverse event rates for patients receiving entrectinib to exceed those of established management. The impact of these simplifying assumptions is, however, likely to be minimal as even in an extreme scenario where 0% event rates are assumed for patients receiving established management the ICER falls only very slightly.

5.2.7 Health related quality of life

The company estimated health state utility values for entrectinib based on EQ-5D-3L data collected in the STARTRK-2 study. For the comparator arm, these were identified via a review of published literature. A summary of utility values is presented in Table 31. Utilities were not adjusted for age, and disutilities relating to adverse events were not applied in the base case analysis.

Table 31 Utility values used in the cost-effectiveness analysis (adapted from Table 53 of CS)

State	Utility value	95% confidence interval	Justification
Progression-free survival			
Entrectinib	[redacted]	[redacted]	Utility derived from clinical trial and valued according to UK societal preferences
Established management	0.73	Applied at individual tumour level	Weighted average of tumour-specific utilities
Progressed disease			
Entrectinib	0.59	Applied at individual tumour level	Assumption of equivalent progressed utility to comparator
Established management	0.59	Applied at individual tumour level	Weighted average of tumour-specific utilities

5.2.7.1 Utility values for entrectinib

EQ-5D-3L data were collected in the STARTRK-2 trial, from which 51 of the 54 integrated analysis population patients came. EQ-5D assessments were collected from 44 of the 51 STARTRK-2 patients across nine tumour types (the specific tumour types included in the analysis not reported in the CS). Patients completed the EQ-5D-3L questionnaires at baseline, on Day 1 of each subsequent treatment cycle of 28 days thereafter, at the end of treatment visit, and in the period after treatment.

The company estimated the mean utility value for the progression-free and progressive disease health states, based on 409 and 44 observations respectively (Table 32). Completion rates of the questionnaire reduced over the course of the trial: completion rates fell below 50% after cycle 10 of the trial, and 22.9% of patients completed the End of Treatment questionnaire. A model was not fit to the EQ-5D data available post progression given the limited number of observations, and the data was

not felt to be plausible as it provided a higher health state utility than that of the pre-progression health state.

Table 32 Mean utility estimates for entrectinib (Table 50 in CS)

State	Number of Observations	Mean	Minimum	Maximum	Median
Baseline	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Pre-progression	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Post-progression	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

A linear mixed model was fitted to the pre-progression EQ-5D data, adjusting for sex, tumour type and age, and accounting for the repeated observations per subject. The final model resulted in an estimate for utility of 0.8119 (0.76, 0.86). Results from a model with a nested random effect by patient within tumour type were used in the model base-case, and resulted in a utility of [redacted]. This makes the assumption that tumours were randomly sampled from a population of possible tumours and that patients were then sampled randomly from within this tumour pool.

Due to the small sample size and associated uncertainty, the post-progression utility from the integrated efficacy analysis was not used in the economic analysis. The predicted utility was also estimated as being higher than for the progression free health state, which was not considered to be plausible. The company therefore assumed that utility in the PD health state was equal to that of established management.

5.2.7.2 Utility values for established management

A systematic review of utility values undertaken by the company did not identify any studies of utility values specific to an *NTRK* population. The company therefore undertook a search of relevant NICE TAs for appropriate utility values, similar to the approach taken to identify clinical outcomes (see Appendix H in CS), In contrast with the approach taken for the comparator efficacy, where a range of estimates for each tumour type were pooled, the utility values extracted for each tumour type were obtained from a single selected TA. Table 33 reports the selected utility value for each tumour type along with associated uncertainty parameters. Note the company reported that the standard error for each utility estimate was often not available in the source document. Where no standard error was reported, a common arbitrary standard error of 0.14 was used for these estimates (i.e. for colorectal cancer). For PPS, where a pooled utility was estimated, a common standard error was estimated by averaging the standard error for each tumour type.

Table 33 Utility sources for comparator tumour types (Table 51 in CS)

Tumour type	N	Utility estimate – PFS	Measure of uncertainty (SE)	Utility estimate – PPS	Measure of uncertainty (SE)	Source
Colorectal cancer	4	0.73	0.14	0.64	0.14	TA40[59]5
MASC	7	0.725	0.14	0.60	0.14	Assumption: average of known
Thyroid cancer (papillary and anaplastic)	5	0.72	0.14	0.64	0.14	TA535[60]
Non-small-cell lung cancer (squamous and non-squamous)	10	0.74	0.18	0.59	0.06	TA428 [61]
Pancreatic cancer	3	0.70	0.14	0.65	0.14	TA476 [52]
Sarcoma	13	0.72	0.14	0.56	0.14	TA465 [62]
Neuroendocrine tumours	3	0.767	0.14	0.725	0.14	TA539[55]
Breast cancer (including secretory)	6	0.705	0.14	0.496	0.14	TA515[50]
Other (average of known)	3	0.725	0.14	0.65	0.14	Assumption: average of known
Weighted average		0.73		0.59		Calculation

5.2.7.3 Adverse event disutilities

The assumption that any disutility has already been incorporated into the base case health state utilities through trial derived EQ-5D utilities, and incorporating an additional disutility may be considered double counting.

The impact of including disutilities of selected adverse events was explored in a scenario analysis conducted by the company, and found that these had a minimal impact on model results.

ERG Comment

Mapping of EQ-5D data

The ERG considers that the use of a linear mixed model is appropriate for analysing the EQ-5D data and it makes assumptions regarding the random sampling from a population of possible tumours that are consistent with those used in the Bayesian hierarchical modelling presented in Section 4.3.1. However, it was not clear how the variables in the linear mixed model were selected and it was considered that the company could have adjusted for additional characteristics that are likely to

impact on HRQoL, such as CNS metastases or line of therapy, as well as the baseline utility to control for differences between patients.

HRQoL benefit for entrectinib patients

The company base-case analysis applies a higher health state utility value for patients receiving entrectinib in the PFS health state compared with those receiving established management. The company justify this difference stating that "entrectinib is an oral TKI therapy with a more convenient administration and relatively tolerable safety profile when compared with traditional cytotoxic chemotherapies, which form the majority of comparator products". However, as discussed in Section 5.2.7, many of the comparators consist of BSC (no active therapy); in such cases, adverse event rates for patients receiving entrectinib may exceed those of established management and no HRQoL impacts of drug administration would apply.

On balance the ERG considers the assumed quality of life benefit in the pre-progression state to be plausible and reasonable given the safety profile of entrectinib, but is concerned about the lack of evidence to justify the assumption of differential quality of life pre- progression and considers there to considerable uncertainty regarding the magnitude of any difference. The ERG further notes that this assumption is not a significant driver of cost-effectiveness as demonstrated in a scenario presented by the company where the utility value was lowered to that of the comparator arm.

HRQoL of patients on established management

With respect to comparator utilities, the ERG is satisfied that the estimates appear reasonable and comparable with other advanced cancers, but is unable to individually verify each individual utility estimate used given the limited time and resource available to the ERG. The ERG, however, does consider there to be a degree of uncertainty in the provided estimates and notes limitations with the company's approach to selecting utility values. Specifically, the ERG notes the inconsistency in approach between data used to populate the effectiveness of comparator therapies and that used to identify utilities. The impact of this inconsistency is not fully clear, but may impact on estimated utilities as the selected utilities will reflect a specific line of therapy, which may not be directly comparable to the equivalent entrectinib patient. In this respect, the ERG notes that data provided by the company in response to the ERG's clarification request highlighted that entrectinib was often given at earlier lines of therapy compared with the sources of data selected to represent established management (see Section 5.2.6). Given that patients in earlier lines of therapy may have better HRQoL, this may lead to the difference in HRQoL between the two arms being overestimated, biasing the cost-effectiveness analysis in favour of entrectinib.

With regards to the decision not to model AE-related disutilities, the ERG considers that it is a reasonable assumption for entrectinib, since the HRQoL data collected in the STARTRK-2 trial is likely to capture any effects relating to events. It is not possible to determine whether this is the case in the comparator arm without reviewing each individual utility estimate in detail, but the impact of inappropriately excluding a disutility is unlikely to be a driver of the economic analysis.

5.2.8 Resources and costs

The CS (Appendix D, Page 6) describes the search strategies used to identify studies of resource use and treatment costs. The costs included in the model comprised drug acquisition, administration and monitoring for entrectinib and the estimated comparator. Unit costs were sourced from the British National Formulary, NHS reference costs 2017-2018 and the Personal Social Services Research Unit (PSSRU).[63] Costs also included *NTRK* fusion screening costs obtained from previous NICE technology appraisals[64], The Scottish Science Advisory Council [65]and inputs from NHS genomic laboratories.

5.2.8.1 Treatment acquisition cost – entrectinib

Table 34 presents the treatment acquisition costs and drug dosing schedules included in the company's base case analysis. The acquisition cost for entrectinib includes the agreed simple PAS. The dosing intensity applied in the model was 100%, which is higher than the observed dosing intensity taken as an average across all of the trials (96.6%).

Table 34 Entrectinib drug acquisition costs

Drug	Pack concentration	Pack volume	Dose per pack	Cost (£)/pack	Source
Entrectinib	100 mg	30	3,000 mg	[redacted](860.00)	[redacted] (and list) price
Entrectinib	200 mg	90	18,000 mg	[redacted](5,160.00)	[redacted] (and list) price

ERG Comment

The ERG accepts the dosing intensity assumed by the company. Although this is an overestimate of the dosing intensity as observed in the trial, the ERG considers that the applied 100% dose intensity is a reasonable though potentially conservative assumption. The ERG, however, notes that the base-case model presented in the CS fails to account for drug wastage due to discontinuation of therapy. The ERG considers this to be unrealistic because once a pack of tablets has been started these would not be reused should the patient discontinue therapy part-way through a pack. The impact of adding drug wastage for entrectinib is to increase drug acquisition for entrectinib and hence to increase the ICER.

5.2.8.2 Treatment acquisition cost – established management

Drug acquisition costs, dosing frequency, and route of administration information for the interventions forming the established management comparator were obtained from the British National Formulary (BNF), see Table 36 below. The CS does not include the confidential PAS schemes, which have been approved for eribulin, everolimus, nab-paclitaxel, nintedanib, trabectedin and trifluridine-tipiracil. Details of these confidential PAS schemes were made available to the ERG and have been incorporated into the analysis presented in a confidential appendix.

Due to the weighted average approach used to construct the established management arm (as discussed in Section 5.2.4), the CS provides an average of the monthly acquisition costs for each identified comparator for the given tumour type. This can be seen in Table 35.

Table 35 Tumour-specific monthly drug acquisition – average by tumour type

Tumour type	Cost per month
Colorectal cancer	£1,878.09
MASC	£0
Thyroid cancer (papillary and anaplastic)	£0
Non-small-cell lung cancer (squamous and non-squamous)	£1952.05
Pancreatic cancer	£1,507.37
Sarcoma	£3,096.16
Neuroendocrine tumours	£1,354.32
Breast cancer (including secretory)	£1,178.76
Other (average of known)	£1,281.60
MASC, mammary analogue secretory carcinoma	

ERG's comment

Table 36 presents the drug acquisition cost data and drug dosing schedules included in the company's base-case analysis. Table 36 also includes drug acquisition costs obtained from the electronic market information tool (eMIT). [66] This provides information on the average price paid by the NHS for pharmaceuticals, which can differ from the list prices listed in the BNF and is seen as a more accurate and up to date indicator of costs. The ERG considers eMIT to be a more appropriate source of drug acquisition costs and conducts scenario analysis presented in Section 6 using these values.

Table 36 Individual comparator acquisition costs (adapted from Table 55, pg. 115 of CS)

Treatments	Composition (mL or tabs)	Cycle length	Dose per cycle	*BNF Cost*	*eMIT cost*

Capecitabine	150mg/tablet	2 weeks	1250 mg/m^2	*£30.00*	£8.15
Eribulin	0.88mg/2ml	3 weeks	2.26 mg/m^2	£361.00	NR
Vinerolbine	10mg/1ml	Weekly	25-30 mg/m^2	£29.00	£35.83
Gemcitabine	1g/10ml	3 weeks	2500 mg/m^2	£13.09	£8.66
Paclitaxel	100mg/16.7ml	3 weeks	175 mg/m^2	£200.35	£9.49
Docetaxel	20mg/ml	3 weeks	75 mg/m^2	£91.51	£11.61
Irinotecan	40mg/2ml	2 weeks	180 mg/m^2	£39.38	£3.19
Folinic acid	50mg/5ml	2 weeks	500 mg/m^2	£20.00	NR
Fluorouracil	500mg/10ml	2 weeks	12 mg/kg	£6.08	£0.97
5FU	2.5g/50ml	2 weeks	600 mg/m^2	£32.00	NR
Oxaliplatin	50mg/10ml	2 weeks	85 mg/m^2	£155.00	£4.32
Trifluridine-tipiracil	15mg	4 weeks	700 mg/m^2	£500.00	NR
Everolimus	10mg/tablet	Daily	10 mg	£2,673.00	NR
Nab-paclitaxel	100mg	4 weeks	375 mg/m^2	£246.00	NR
Gemcitabine (combination)	1g/10ml	3 weeks	*1000 mg/m2*	£13.00	£8.66
Leucovorin	100mg/10ml	2 weeks	200 mg/m^2	£37.50	NR
Lenvatinib	24	Daily	*24 mg*	£47.90	NR
Sorafenib	200	Daily	*800 mg*	£3,576.56	NR
Doxorubicin	200mg/100ml	3 weeks	60 – 75 mg/m^2	£391.40	£15.59
Ifosfamide	1g	3 weeks	*5-6g/m2*	£115.79	NR
Trabectedin	0.25mg	3 weeks	1.5 mg/m^2	£363.00	NR
Pegylated liposomal doxorubicin	20mg/10ml	*4 weeks*	*50mg/m2*	£360.23	NR
Carboplatin	50mg/vial	*3 weeks*	AUC 5–6 IV	£20.00	£3.59
Nintedanib	100mg/tablet	3 weeks	8000 mg	£2,151.10	NR
BNF, British National Formulary; eMIT, electronic market information tool; 5FU, 5-fluorouracil; NR, not reported					

The ERG is concerned that the monthly drug acquisition costs of the current standard of care for gynaecological cancers and cholangiocarcinoma have not been estimated for the individual tumour types. Rather, the costs associated with these tumour types are an average of colorectal cancer, MASC, thyroid cancer, NSCLC, pancreatic cancer, sarcoma, neuroendocrine tumours and breast cancer. It is unclear whether this reflects costs of relevant therapies for these tumour types, drug

acquisition costs of the comparator therapies were, however, not a major driver of cost effectiveness and therefore the ERG does not consider this issue further.

5.2.8.3 Treatment administration costs

The company state that in order to estimate health state costs across a range of chemotherapy types, it was necessary to apply a simplifying assumption that treatments with similar routes of administration are likely to be associated with similar routine healthcare costs across the different tumour types. The company grouped the interventions into three administration classes: oral, simple IV and complex IV. Each class is associated with an average monthly administration cost, which is applied to each intervention.

The NICE TAs used to inform the costs for each of these classes were TA515 (Eribulin for treating locally advanced or metastatic breast cancer after 1 chemotherapy regimen) [50] for oral chemotherapy, TA520 (Atezolizumab for treating locally advanced or metastatic non-small-cell lung cancer after chemotherapy)[67] for simple IV chemotherapy and TA476 (Paclitaxel as albumin-bound nanoparticles with gemcitabine for untreated metastatic pancreatic cancer) [52] for complex IV chemotherapy.

The oral therapies, including entrectinib, have a monthly cost of £14.59, the simple IV interventions have a monthly cost of £331.69 and the complex IV therapies have a monthly cost of £488.12. The majority of comparators were simple IV chemotherapy or complex IV chemotherapy.

ERG Comment

The ERG is concerned about the appropriateness of the simplifying assumptions made by the company and notes that within categories that infusion time varies significantly. For example, eribulin is classed as a simple IV therapy with infusion time of 2-5 minutes, while trabectidin, also classed as a simple IV therapy, is administered over a period of 24 hours. Similar inconsistencies are seen in the complex category. For example, vinorelbine is infused over a period of 6-10 minutes whereas FOLFIRI is administered as irinotecan infused over 60-90 minutes, folinic acid infused over 2 hours, fluorouracil infused as a bolus and 5FU infused over 46 hours.

The ERG considers that the simplifying assumptions made by the company were not necessary and that individual administration costs for each of the comparator interventions could have been sought. It is unclear whether this approach under or overestimates the costs, however the result is increased uncertainty in the administration costs. Due to the resource required to implement this, the ERG was unable to address this issue in our additional analyses.

5.2.8.4 Health state costs

The three health states in the model are: progression free, progressed disease and death – the model includes those costs associated with patients being in each of these health states.

Progression-free costs

The costs applied in the model in the PFS state are differentiated by the same three classes as the treatment administration costs. These are oral, simple IV and complex IV. The modelled unit costs and resources in each of these classes can be seen in Table 37.

Table 37 Progression free health state costs

Progression free health state costs: Oral treatment					
Item	**Number used**	**% of patients**	**Unit cost**	**Monthly cost**	**Reference***
Medical oncology, outpatient visit	1	100	£162.05	£162.05	Total Outpatient Attendances -row 370 (Outpatient, consultant-led)
GP surgery visit	1	100	£37.40	£37.40	PSSRU 2018 [63]: 10.3b GP unit costs (9.22 minutes patient time)
CT scan	1	33	£132.75	£44.21	RD22Z (CT scan of one area, pre- and post-contrast)
Full blood count	1.55	100	£2.51	£3.89	DAPS05 (haematology)
Liver function tests	1.66	100	£1.11	£1.84	DAPS04 (clinical biochemistry)
Progression free health state costs: Simple IV treatment					
Medical oncology, outpatient visit	1.33	100	£162.05	£215.53	Total Outpatient Attendances -row 370 (Outpatient, consultant-led)
GP surgery visit	1	100	£37.40	£37.40	PSSRU 2018[63]: 10.3b GP unit costs (9.22 minutes patient time)
CT scan	1	33	£132.75	£44.21	RD22Z (CT scan of one area, pre- and post-contrast)
Full blood count	1.55	100	£2.51	£3.89	DAPS05 (haematology)
Liver function tests	1.66	100	£1.11	£1.84	DAPS04 (clinical biochemistry)
Progression free health state costs: Complex IV treatment					
Medical oncology, outpatient visit	1.33	100	£162.05	£215.53	Total Outpatient Attendances -row 370 (Outpatient, consultant-led)
Medical oncology,	1	50	£104.00	£52.00	Total Outpatient Attendances -row 370 (Outpatient,

outpatient, nurse-led					non-consultant-led)
GP surgery visit	1	100	£37.40	£37.40	PSSRU 2018 [63]: 10.3b GP unit costs (9.22 minutes patient time)
Nurse community visit	1	50	£42.00	£21.00	PSSRU 2018 [63]: 10.2 GP practice nurse unit costs (assumes 1 hour patient contact)
CT scan	1	33	£132.75	£44.21	RD22Z (CT scan of one area, pre- and post-contrast)
Full blood count	1.55	100	£2.51	£3.89	DAPS05 (haematology)
Liver function tests	1.66	100	£1.11	£1.84	DAPS04 (clinical biochemistry)

As with administration costs, the resources used in each of the classes were taken from three TAs identified in the company's search for comparator data. These appraisals are TA515 (eribulin for treating locally advanced or metastatic breast cancer after 1 chemotherapy regimen) [50] for oral chemotherapy, TA520 (atezolizumab for treating locally advanced or metastatic non-small-cell lung cancer after chemotherapy)[68] for simple IV chemotherapy and TA476 (Paclitaxel as albumin-bound nanoparticles with gemcitabine for untreated metastatic pancreatic cancer) [52] for complex IV chemotherapy. Monthly HCRU costs for entrectinib were assumed to be entirely associated with those of an oral therapy. The company states that a clinical expert validated the resources used in these appraisals as being generalisable to the tumour types covered in this appraisal. Unit costs were sourced from the 2017-18 NHS Reference Costs and the PSSRU 2018.[63]

ERG Comment

The ERG has concerns about the company's approach to modelling costs in the PFS state based on the type of therapy received by the patients. The company's approach is an oversimplification of the costs associated with the care in different tumour types and as a result, there is significant uncertainty in the modelled cost inputs. The approach taken by the company would suggest that the cost associated with treating a patient with NSCLC and neuroendocrine with oral therapies in the PFS health state are identical. As with the administration costs, the ERG considers that the simplifying assumption of grouping therapies into one of three classes was not necessary and that tumour specific health state costs could have been sought. As with administration costs it is unclear whether this approach under or overestimates the costs, however, the result is increased uncertainty in the administration costs.

However, as the influence of administration costs on the ICER is minimal, the ERG does not consider this issue further.

Progressed disease costs

The costs applied in the model for the PPS state are the same across both entrectinib patients and comparator patients. These costs are presented in Table 38.

Table 38 Progressed disease health state costs (Table 63, page 122 of the CS)

Item	Number used	% of patients	Unit cost	Monthly cost	Reference
Medical oncology, outpatient visit	1	100	£162.05	£162.05	Total Outpatient Attendances - row 370 (Outpatient, consultant-led)
Medical oncology, outpatient, nurse-led	1	100	£104.00	£104.00	Total Outpatient Attendances - row 370 (Outpatient, non consultant-led)
GP home visit	1	100	£37.40	£37.40	PSSRU 2018[63]: 10.3b GP unit costs (9.22 minutes patient time)
Nurse community visit	1	67	£42.00	£30.15	PSSRU 2018 [63]: 10.2 GP practice nurse unit costs (assumes 1 hour patient contact)

ERG Comment

The ERG has no major concerns with the unit costs of the items included in the progressed-disease health state included in the model. The ERG, however, notes that some additional costs could have been included to accurately reflect the care received by patients with cancer in a progressed disease state. These include medication costs (e.g. steroids, NSAIDS, morphine, bisphosphonates and dietary supplementation), and tests and procedure costs (e.g. full blood count, serum chemistry, CT scan, home oxygen and x-ray). The omission of these costs is likely to be small, but will lead to an underestimation of the ICER.

5.2.8.5 End of life costs

The CS model calculated a one-off cost to account for terminal care incurred. The model applied this at the transition from progressed disease to death. The costs were obtained from Georghiou and Bardsley [69], adjusted for inflation to 2017-2018. These costs can be seen in Table 39.

Table 39 Summary of components of end of life costs

Component	Mean cost, last 3 months (2017 – 2018)	Mean cost/month, last 3 months (2017 – 2018)
Emergency inpatient admission	£4049.29	£1349.76
Non-emergency inpatient admission	£1352.75	£450.92
Outpatient attendance	£375.98	£125.33
A&E visits	£79.57	£26.52
Social care	£441.63	£147.21
District nursing care	£584.86	£194.95
GP visits	£363.05	£121.02

ERG Comment

The ERG has no major concerns with the end of life costs as these are common to both groups, and because virtually all participants die within the time horizon, the only differences in these costs between the two treatment groups are as a result of discounting.

5.2.8.6 Adverse event costs

The company only included those adverse events occurring at a rate of ≥5% in the model. All adverse events except increased weight were considered to occur at the same rate for both entrectinib and comparator patients. The adverse events that were included in the model were: anaemia, fatigue, neutropenia and weight increase.

ERG Comment

The ERG did not identify any areas of concern regarding the company's choice of adverse events to include in the model. The ERG, however, notes that event rates were assumed equal for all AE except weight gain which had a zero cost. As such, effectively no costs of AEs were included in the model. As outlined, in Section 5.2.6.5, the ERG considers that the company could have modelled AE for comparator therapies, though it is acknowledged that this is likely to have only a small effect on the estimated ICER.

5.2.8.7 *NTRK*-fusion screening costs

The company included a projected cost for screening eligible patients in the base-case analysis. In the model, the company proposed a hierarchical approach in which IHC is conducted to identify patients with tumours expressing *NTRK* protein, followed by confirmatory testing with an NGS panel to establish whether these patients have specific *NTRK* gene fusions.

To calculate the *NTRK*-fusion screening costs, the company estimated the number-needed-to-screen for each of the tumour types represented in the integrated efficacy analysis (Table 1). The company then identified the screening tests conducted in current practice within the NHS based on the NHS Genomic Testing Directory (Table 40).

Table 40 Frequency of NTRK fusions in enrolled tumour types (Adapted from CS, Table 66, p 125)

Tumour type	NTRK-fusions rate	Number needed to screen	Current Testing
CRC	[redacted]	[redacted]	Other biomarker screening
NSCLC (squamous and non-squamous)	[redacted]	[redacted]	Other biomarker screening
Pancreatic	[redacted]	[redacted]	No molecular testing within directory
Non-secretory breast cancer	[redacted]	[redacted]	Other biomarker screening
Secretory Breast Carcinoma (0.02% HER2-)	[redacted]	[redacted]	Other biomarker screening
Thyroid (papillary/anaplastic)	[redacted]	[redacted]	Other biomarker screening
Neuroendocrine tumours	[redacted]	[redacted]	No molecular testing within directory
Sarcoma (non-paediatric)	[redacted]	[redacted]	WGS
MASC	[redacted]	[redacted]	NTRK-fusion testing
Other	[redacted]	[redacted]	No molecular testing within directory
Paediatric cancers	[redacted]	[redacted]	WGS

CRC, colorectal cancer; NSCLC, non-small cell lung cancer; MASC, mammary analogue secretory carcinoma; WGS, whole genome sequencing; NR, not reported.

In tumour types where a genetic test is already conducted in clinical practice ('other biomarker screening' in Table 40), the company assumed the unit cost of a standard IHC test to be £75.00. In the base-case analysis, the company assumed the cost of screening for the comparators in which genetic testing already occurs, to be the number-needed-to-screen (NNS) (Table 40) multiplied by the cost of the IHC test. In the entrectinib arm, the screening cost was assumed to be the same screening cost calculated for the comparator arm with an additional cost of adding a confirmatory NGS test for 11%

of the patients IHC tested. The unit cost of an NGS test applied in the base case was [redacted] per test. The proportion of patients receiving NGS was based on clinical data provided by an investigator involved in the entrectinib clinical development programme, which suggested that the IHC testing approach will remove 89% of *NTRK*-fusion negative samples. In those tumour types where no genetic testing is currently conducted in clinical practice ('no molecular testing in clinical practice' in Table 40), no screening costs were attributed to the comparator arm in the base case. In these tumour types, the screening costs in the base case for the entrectinib arm was the cost of an IHC test multiplied by the NNS, plus an additional cost a confirmatory NGS test for 11% of IHC tested patients.

In their base case, the company assumed that MASC patients were diagnosed by IHC alone in line with current testing for these patients. As *NTRK* fusions are already included in the Genomic Testing Directory for MASC, this cost was applied in entrectinib and established management arms of the model. For those tumours in which WGS is reimbursed for specific tumour types (paediatric tumours and sarcoma), the company assumed that *NTRK* fusion positive patients would be identified via current testing practice. A unit cost of screening for *NTRK*-fusions is £800.00 per test per patient tested was therefore applied in both entrectinib and established management arms of the model.

The costs of screening each tumour type to identify one entrectinib-eligible patient are shown in Table 41. The model used a weighted average of the costs for each tumour type, weighted by the number of patients with that specific tumour type in the efficacy evaluable population. In the model, the average incremental cost of screening for the entrectinib arm over the comparator arm was an additional £15,828.

Table 41 Costs of screening by tumour type to identify one patient used in the base case (Adapted from CS, Table 69, p 127)

Tumour type	Base case: entrectinib	Base case: comparator
CRC	[redacted]	[redacted]
NSCLC (squamous and non-squamous)	[redacted]	[redacted]
Pancreatic	[redacted]	[redacted]
Non-secretory breast cancer	[redacted]	[redacted]
Secretory Breast cancer	[redacted]	[redacted]
Thyroid (papillary/anaplastic)	[redacted]	[redacted]
Neuroendocrine tumours	[redacted]	[redacted]
Sarcoma (non-paediatric)	[redacted]	[redacted]
MASC	[redacted]	[redacted]
Other	[redacted]	[redacted]

Weighted average within integrated analysis	[redacted]	[redacted]

CRC, colorectal cancer; NSCLC, non-small cell lung cancer; MASC, mammary analogue secretory carcinoma

In addition to the above the, company also presented scenario analyses in which 50% and 25% of the costs of screening are applied to entrectinib to represent scenarios in which 2 or 4 *NTRK* fusion-targeting medicines are available, respectively.

ERG Comment

There are a significant number of uncertainties in estimating appropriate testing costs. These relate to the testing strategy adopted, unit costs applied, feasibility and provision of current services.

Testing Strategy

The ERG considers that the company's proposed, hierarchical approach to testing to be a plausible strategy to identify individuals with an *NTRK* fusion. IHC is high-throughput and is inexpensive, making it a practical screening tool to use in a large population. The diagnostic accuracy of IHC is, however, variable. IHC has a low sensitivity for tumours expressing the *NTRK*3 fusion and high rates of false negatives in smooth muscle and neural tumours. Implementation of this strategy would therefore mean that a proportion of *NTRK* fusion positive patients are likely to be missed. See Section 2.2.2.2 for further discussion of limitations of this approach.

The company's assumption that *NTRK*-fusion positive paediatric and adult sarcoma patients would be identified under established pathways as WGS is disputed by the ERG. The ERGs clinical advisers stated RNA-based NGS would be needed after WGS to confirm an *NTRK*-fusion positive tumour. This will require the entrectinib testing costs to include an additional RNA-based NGS cost for all of the paediatric and sarcoma patients identified through WGS. The effects of this additional cost on the company's base case will increase the costs associated with the entrectinib arm and is also explored in scenario analysis presented in Section 6.3.2.

As noted in Section 2.2.2, RNA-based NGS fusion panels are available on the NHS for a specific subgroup of patients with NSCLC, targeting a range of genes including *EGFR, ALK*, and *ROS1*. Whilst this panel does not currently target *NTRK1-3* rearrangements, genomic advisers informed the ERG that the costs of adding additional gene targets to an RNA-based NGS panel are nominal. Incremental costs associated with the identification of *NTRK* fusion patients with NSCLC may therefore be close to zero. Scenario analysis is implemented in Section 6.3.2 evaluating the impact of removing testing costs for NSCLC patients.

As described in section 2.2.2, there are several other strategies that could be adopted to detect *NTRK* fusions. One potential approach would be to offer NGS as a first line test to identify *NTRK* fusion positive patients, with or without confirmatory IHC. Given the low prevalence of *NTRK* fusions in most tumour types, and the large population of individuals to be tested, a primarily NGS-based testing strategy may be impractical because NGS is more expensive and time-consuming. Costs of NGS and resources required to implement it are falling over time, potentially making more plausible in the future. The advantage of this approach is that NGS has high diagnostic accuracy, consequently, it is less likely that patients with an *NTRK* fusion will be missed.

An alternative strategy outlined is also outlined in recent guidelines published by the ESMO Translational Research and Precision Medicine Working Group.[17] This approach suggests that the testing pathway for detecting *NTRK* fusion positive patients should vary depending on the frequency of *NTRK* fusion and current availability of testing for each tumour type.

- For tumours with a high frequency of *NTRK* positive patients (e.g. MASC & infantile fibrosarcoma), FISH or a targeted NGS panel should be used.
- In the tumour types where genomic testing is currently available (e.g. NSCLC and colorectal cancer), NGS should be used as a first line testing approach.
- For the tumour types where no genomic testing is available (e.g. pancreatic cancer), IHC followed by confirmatory NGS is recommended.

The ESMO recommendations take advantage of current testing availability, and therefore may be an efficient approach with potentially lower incremental testing costs. In the future, the expansion of genomic testing will allow first-line NGS to be used for a wider range of tumour types, with advantages in terms of testing sensitivity.

A limitation of this approach is that current testing may not be widely available for patients within each tumour type, and particular eligibility criteria is likely to limit the number of individuals who would be able to access testing. This may limit the range of tumour sites where first-line NGS will be viable option. A further issue with this approach is that NGS tests currently used on the NHS are primarily DNA-based, which as outlined in Section 2.2.2 have limited sensitivity to structural rearrangements. Implementing the ESMO guidelines would therefore require switching to RNA-based NGS. This might have implications on costs because RNA-based NGS is more expensive. Nonetheless, with the increased use of RNA-based NGS in clinical settings, the ESMO approach for using NGS will become more practical in the future, as fusion testing can easily be added to current testing panels for nominal costs.

Comparator Testing Costs

The ERG considers the inclusion of screening costs included in the comparator arm, as seen in Table 41, to be inappropriate. The focus of testing costs should be on the incremental testing costs associated with identifying *NTRK* fusion positive patients and therefore the ERG suggests the comparator testing costs should be removed unless current testing is able to identify *NTRK* fusions. In current practice, the NHS only reimburses *NTRK*-fusion screening assays for MASC patients. This may be subject to change over time, as molecular testing is expanded, however it is the ERG's view that decisions should be made on current practice not a possible hypothetical future. The ERG has concerns regarding the validity of the scenarios presented by the company in which testing costs are shared across two and four *NTRK* targeting therapies. The base-case analysis should represent the incremental costs of implementing entrectinib in the NHS, not the implementation of a range of hypothetical agents that may or may not be available in the future. Testing costs are significant drivers of cost-effectiveness and the implications of removing these testing costs from the comparator arm will be explored further in Section 6.3.2

Unit Costs Applied

The IHC testing unit cost was assumed to be £75.00 based on a pathologist's input obtained during a NICE committee meeting to discuss the appraisal of crizotinib for the treatment of NSCLC. In this appraisal, the cost of IHC was identified to be between £50 and £100 excluding laboratory costs. The ERG is concerned that the £75.00 unit cost applied in the CS underestimates the unit cost of IHC testing as this estimate does not include laboratory costs. Further, the ERG considers that the marginal cost of implementing IHC screening is likely to vary depending upon whether current service provision already supports the regular use of molecular testing (for further information on tumour types in which molecular testing is used see Section 2.2.2). This reflects the fact in tumour types where no testing is currently implemented will require greater investment in infrastructure, as well as additional marginal costs associated with the administration of testing. This may include costs of obtaining tissue samples, postage, and clinician time associated with interpretation of test results.

The company assumes that the cost of NGS testing is [redacted]. The ERG considers the cost reasonable and in line with clinical advice received by the ERG. The ERG, however, notes that the majority of NGS currently available on the NHS is DNA-based, which is unsuitable for testing *NTRK* fusions due to poor diagnostic accuracy. Implementation of testing regimens based around a first line NGS would therefore require the adoption of either RNA-based or hybrid DNA/RNA based NGS which is more expensive than DNA tests. Implementation of an NGS-centred testing strategy would therefore potentially incur additional costs, even where current testing includes NGS.

The ERG further notes the testing regimen proposed by the company requires the implementation of either RNA-based or DNA/RNA hybrid NGS panels as a confirmatory test. Due to its labile nature,

RNA can be easily damaged during tissue handling and preparation of the sample, most damaged samples can be detected in pre-screening tests conducted before administration of the NGS panel. However, in highly contaminated RNA, damage is not detected until the after assay has been used. Therefore, a new test is required to detect an *NTRK* fusion, and hence, increasing the cost. This additional potential cost is not accounted for in the company's calculations and therefore applied costs are likely to be underestimates.

Number Needed to Screen

In order to calculate the costs associated with testing, the company calculated the number of individuals that would need to be screened in order to identify one individual with an *NTRK* fusion. The CS reports that the NNS was estimated based on prevalence of *NTRK* figures reported in Amatu *et al* [2]and data on file. [70] The ERG, however notes that the reported estimates of NNS differ from the ERG's preferred estimates based on the FMI database which recorded the frequency of *NTRK* fusion positive patients from a sample circa 166,000 samples. The NSS to screen for each tumour type estimated by the ERG and by the company is presented in Table 42.

As can be seen from Table 42, there are a significant number of inconsistencies in the estimated NNS between the CS and ERG. For example, in pancreatic cancer the CS estimates the NNS as [redacted], while the ERG estimates a figure of [redacted]. The impact of these differences in NNS is significant, affecting both the average NNS across all sites, as well as tumour type-specific testing costs.

In addition to the above, the ERG notes that the company's estimates of the NNS, and by extension average testing costs are based upon the distribution of the 13 tumour sites represented in the integrated efficacy analysis. This is problematic for two reasons. Firstly, as discussed in section 5.2.3 the distribution of tumour types in the integrated efficacy analysis is unlikely to represent the distribution in practice with some sites over represented and other underrepresented. Secondly, the modelled population includes only 13 tumour types and excludes a number of tumour types in which it is known that *NTRK* fusions occur; NNS for unrepresented tumour types ranges from [redacted] (Infantile Fibrosarcoma) to [redacted] (Cervix Cancer). Estimated costs of testing informing the company's base case analysis are therefore unlikely to represent testing costs for the population eligible for entrectinib. The ERG implements scenario analysis in section exploring both of these issues using the ERG's preferred estimates of NNS.

Table 42 Number needed to screen by tumour type

Tumour Type	Prevalence of *NTRK* fusion (ERG)	Number Needed to Screen (ERG)	Number Needed to Screen (Company)
Salivary gland (MASC)	[redacted]	[redacted]	[redacted]
NSCLC	[redacted]	[redacted]	[redacted]
Breast cancer (not specified)	[redacted]	[redacted]	[redacted]
Secretory breast carcinoma	[redacted]	[redacted]	[redacted]
Papillary thyroid tumour	[redacted]	[redacted]	[redacted]
Thyroid Tumour (NOS)	[redacted]	[redacted]	[redacted]
Colon/colorectal	[redacted]	[redacted]	[redacted]
Neuroendocrine (NOS)	[redacted]	[redacted]	[redacted]
Cholangiocarcinoma	[redacted]	[redacted]	[redacted]
Pancreatic	[redacted]	[redacted]	[redacted]
Uterine	[redacted]	[redacted]	[redacted]
Ovarian	[redacted]	[redacted]	[redacted]
Cervix	[redacted]	[redacted]	[redacted]
Soft tissue sarcoma	[redacted]	[redacted]	[redacted]
High grade glioma	[redacted]	[redacted]	[redacted]
Paediatric high grade glioma	[redacted]	[redacted]	[redacted]
Congenital mesoblastic nephroma	[redacted]	[redacted]	[redacted]
Paediatric melanoma	[redacted]	[redacted]	[redacted]
Infantile fibrosarcoma	[redacted]	[redacted]	[redacted]
Paediatric low grade glioma	[redacted]	[redacted]	[redacted]

The impact of different testing approaches and different estimates of the population on the company's base-case ICER will be explored in Section 6.3.2.

Feasibility

The company acknowledge that the requirements for screening for *NTRK* will take into account the likely large economic impact it is expected to have. However, the company do not fully consider the feasibility of implementing additional testing to a large population.

In order to determine the impact of testing for *NTRK*, the ERG calculated the number of individuals that would require IHC and NGS testing, as proposed in the company's testing strategy. As outlined in Section 2.2.2, there are two approaches that can be used to calculate the number of individuals who would require screening. Using a top-down approach, based on the total annual incidence of cancer in England with stage 3/4 tumours, 92,524 individuals would require IHC screening every year. A

further 10,178 would also require confirmatory NGS tests assuming, 11% of individuals who receive IHC will require NGS. The ERG also used a conservative, bottom-up approach to calculate the number of patients who require testing, based on the tumour sites in which there is a known *NTRK* fusion. Under this approach, the ERG estimates that an additional 52,782 IHC would be undertaken each year along with a further 5,806 confirmatory NGS.

These figures represent a significant increase in the number of molecular tests that would be untaken annually and therefore the ERG is concerned that additional investment in current genomic services will be necessary to provide *NTRK* testing across the NHS. This may include expansion of current infrastructure that would be required to meet an increasing number of referrals, additional requirements for a larger workforce to prepare samples and process tests, as well as the need to employ and train additional clinical geneticists and bioinformaticians specialised in genetic fusions and targeted medicines. [32, 34] These costs are, however, not considered in the economic model and as described in Section 2.2.2 may have long term implications on the viability of molecular testing services which are predicted to become increasingly overwhelmed by demand for testing services. See Section 2.2.2 for further discussion of the feasibility of expanding testing services.

5.2.9 Cost effectiveness results

Entrectinib has a confidential patient access scheme (PAS), comprising a simple discount of [redacted]. A number of the interventions that comprise established management also have PAS available. The results presented below include the PAS for entrectinib, but do not include PAS available for comparators, with results including these PAS presented in a confidential appendix to this report.

Table 43 presents the base-case deterministic analysis of entrectinib. It shows that entrectinib was associated with increased costs (cost difference of [redacted]) and was more effective (gain of [redacted] QALYs), compared with established management. The company's base-case ICER was £54,646 per QALY.

Table 43 Base-case results (Adapted from CS, Table 72 and Table 73, p 132)

Technologies	Total Costs (£)	Total LYG	Total QALYs	Incremental Costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY)
Entrectinib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£54,646
Established management	£62,931	1.74	1.12	-	-	-	-
CS, company submission; ICER, incremental cost-effectiveness ratio; LYG, life-years gained; QALYs, quality-adjusted life-years; PAS, patient access scheme							

5.2.9.1 Updated base-case

Following response to clarification questions, the company presented an updated deterministic base case ICER using the latest trial data cut-off for PFS and OS for the integrated analysis population. The previous data cut was from 31st May 2018 and the updated data cut is from [redacted]. The update was implemented along with other requested model corrections suggested by the ERG. The updated results show the ICER of entrectinib compared to the established management comparator is £52,609 and is presented in Table 44.

Table 44 Updated base-case results

Technologies	Total Costs (£)	Total LYG	Total QALYs	Incremental Costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY)
Entrectinib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£52,609
Established management	£63,028	1.74	1.12	-	-	-	-
CS, company submission; ICER, incremental cost-effectiveness ratio; LYG, life-years gained; QALYs, quality-adjusted life-years; PAS, patient access scheme							

Table 44 shows that entrectinib was associated with increased costs (cost difference of [redacted]) and was more effective (gain of [redacted] QALYs), compared with established management.

5.2.9.2 Sensitivity analyses

Probabilistic sensitivity analysis

The company performed a probabilistic sensitivity analysis (PSA) using Monte-Carlo simulation with 2,000 iterations. In each iteration, the model drew inputs from defined distributions for selected parameters (CS Table 70, Pages 128-129). The probabilistic ICERs were lower than those in the deterministic analysis. Table 45 presents the results of the probabilistic sensitivity analysis.

Table 45 Company probabilistic cost-effectiveness results (including Entrectinib PAS)

Technologies	Total Costs (£)	Total LYG	Total QALYs	Incremental Costs (£)	Incremental LYG	Incremental QALYs	ICER (£/QALY)
Entrectinib	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£52,052
Established management	£64,128	1.74	1.12	-	-	-	-
CS, company submission; ICER, incremental cost-effectiveness ratio; LYG, life-years gained; QALYs, quality-adjusted life-years; PAS, patient access scheme							

The mean probabilistic ICER of entrectinib was £52,052 per QALY gained versus established management. The probability that entrectinib is the most cost-effective treatment option at WTP threshold of £30,000 is [redacted], and [redacted] at £50,000. The cost-effectiveness acceptability curve for all comparators is provided in Figure 18.

Figure 18 Cost-effectiveness acceptability curve for entrectinib and established management (including entrectinib PAS), (CS, Figure 25, pg. 134)

Figure redacted

QALY, quality-adjusted life year; SoC standard of care

The results show there is little difference between the deterministic and probabilistic ICERs. The average incremental QALYs gained with entrectinib compared to established management was [redacted], which was [redacted] QALYs more than in the updated deterministic analysis.

ERG Comment

The ERG has concerns about the uncertainty of the probabilistic ICER included in the CS. The narrow distributions of the comparator costs, total life years gained, and total QALYs appears to be unrealistic and result in a misleading level of confidence in the comparator results. The narrow confidence intervals around the comparator effectiveness results stems from the company not properly accounting from the uncertainty in the comparator effectiveness estimates.

The ERG also has concerns about the standard errors around the survival estimates used to construct the established management comparator. The standard errors are assumed, and have not been extracted from the original sources of the comparator effectiveness, utilities, and costs. For a discussion of the methods used to construct the weighted average comparator, see Section 3.3. In response to clarification questions, the company stated that for the published survival estimates, due to lack of covariance matrices and correlations reported and the use of an exponential model for the extrapolation, the extrapolated mean is varied around a normal distribution. This was to avoid any assumptions on skewness and allow for a normal range of assessments of the uncertainty around these estimates. The ERG is concerned this approach underestimates the uncertainty around the cost-effectiveness results of the comparator, particularly given the issues involved in the method used to construct the weighted average comparator.

Deterministic sensitivity analysis

The company presented a series of deterministic sensitivity analyses (DSA) in the form of univariate sensitivity analyses to assess the impact of varying key model input parameters upon the ICER.

Selection of parameters for inclusion in the analysis was conducted *a priori.* Unless otherwise stated, base case values were adjusted across a +/- 20% range. The DSA inputs are summarised in CS Table 75.

Tornado diagrams summarising the twelve most influential parameters as reported by the company are presented in Figure 19. The results indicate that varying the median OS of the comparator and the weighted screening costs had the greatest impact upon the ICER. The utility of the PFS state in the entrectinib arm was also a driver of the model's results. The DSA did not produce any ICERs less than £40,000/QALY.

Figure 19 Univariate sensitivity analysis for entrectinib vs comparator (CS, Figure 26, pg. 135)

Figure redacted

OS, overall survival; PFS, progression-free survival; HCRU, health care resource utilisation; PPS, post-progression survival

5.2.9.3 Scenario analyses

The submission and clarification response included an extensive series of scenario analyses to assess the robustness of the model results and the impact of the assumptions included in the base-case analysis. The results of the scenario analyses performed on the company's updated base case are presented in Table 46. The results were most sensitive to variations in the parametric function used to extrapolate OS which resulted in a range of ICERs from £37,217 to £81,588 per QALY. For a discussion of the choice of parametric function, see Section 5.2.6. The results were also sensitive to the tumour weighting applied to the comparator. Reweighting the comparator data to be 100% weight applied to MASC and pancreatic comparator outcomes resulted in ICERs of £31,064 and £114,524 per QALY, respectively. For a discussion of the methods and survival data used to construct the comparator, see Section 5.2.4.

Table 46 Scenario analysis results (adapted from CS and clarification response)

Parameter	Value	Incremental Costs	Incremental QALYs	ICER (£/QALY)
Distribution Entrectinib NTRK+ OS	Exponential	[redacted]	[redacted]	52,609
Distribution Entrectinib NTRK+ OS	Weibull	[redacted]	[redacted]	64,149
Distribution Entrectinib NTRK+ OS	Log-normal	[redacted]	[redacted]	37,217
Distribution Entrectinib NTRK+ OS	Gamma	[redacted]	[redacted]	71,383
Distribution Entrectinib NTRK+ OS	Log-logistic	[redacted]	[redacted]	41,509
Distribution Entrectinib NTRK+ OS	Gompertz	[redacted]	[redacted]	81,588

Distribution Entrectinib NTRK+ PFS	Exponential	[redacted]	[redacted]	52,609
Distribution Entrectinib NTRK+ PFS	Weibull	[redacted]	[redacted]	52,463
Distribution Entrectinib NTRK+ PFS	Log-normal	[redacted]	[redacted]	53,571
Distribution Entrectinib NTRK+ PFS	Gamma	[redacted]	[redacted]	52,941
Distribution Entrectinib NTRK+ PFS	Log-logistic	[redacted]	[redacted]	53,566
Distribution Entrectinib NTRK+ PFS	Gompertz	[redacted]	[redacted]	52,570
Distribution Entrectinib NTRK+ TTD	Exponential	[redacted]	[redacted]	52,609
Distribution Entrectinib NTRK+ TTD	Weibull	[redacted]	[redacted]	52,609
Distribution Entrectinib NTRK+ TTD	Log-normal	[redacted]	[redacted]	52,609
Distribution Entrectinib NTRK+ TTD	Gamma	[redacted]	[redacted]	52,609
Distribution Entrectinib NTRK+ TTD	Log-logistic	[redacted]	[redacted]	52,609
Distribution Entrectinib NTRK+ TTD	Gompertz	[redacted]	[redacted]	52,609
Treatment duration assumption	Trial-observed treatment duration	[redacted]	[redacted]	50,838
Treatment duration assumption	According to label	[redacted]	[redacted]	52,609
Time horizon	5	[redacted]	[redacted]	68,849
Time horizon	10	[redacted]	[redacted]	54,807
Time horizon	15	[redacted]	[redacted]	53,011
Time horizon	20	[redacted]	[redacted]	52,684
Time horizon	25	[redacted]	[redacted]	52,621
Time horizon	30	[redacted]	[redacted]	52,609
Screening costs	Base case: 100% attributed to entrectinib	[redacted]	[redacted]	52,609
Screening costs	50% attributed to entrectinib	[redacted]	[redacted]	44,762
Screening costs	25% attribution to entrectinib	[redacted]	[redacted]	40,838
Screening costs	Screening costs excluded	[redacted]	[redacted]	36,914
Prognosis of comparator	Base case: aggregated trial reported outcomes	[redacted]	[redacted]	52,609
Prognosis of comparator	Adjustment to reflect poorer NTRK prognosis (HR= 2.33)	[redacted]	[redacted]	35,589
Prognosis of comparator	Incorporation of CNS	[redacted]	[redacted]	46,981

	metastases (comparator)			
Post-progression therapy	Base case: 0% active treatment for comparator patients; 35% for entrectinib	[redacted]	[redacted]	52,609
Post-progression therapy	0% active treatment for comparator patients; 50% for entrectinib	[redacted]	[redacted]	58,120
Post-progression therapy	0% active treatment for comparator patients; 80% for entrectinib	[redacted]	[redacted]	69,143
Post-progression therapy	Equivalent post-progression treatment (50% each)	[redacted]	[redacted]	54,868
PFS utility	Base case: Entrectinib PFS utility derived from trial data	[redacted]	[redacted]	52,609
PFS utility	Entrectinib PFS utility reduced to match comparator PFS value	[redacted]	[redacted]	59,390
Tumour-weighting	Base case – trial weighting	[redacted]	[redacted]	52,609
Tumour-weighting	100% weight applied to MASC comparator outcomes	[redacted]	[redacted]	31,064
Tumour-weighting	100% weight applied to pancreatic cancer comparator outcomes	[redacted]	[redacted]	114,524
OS, overall survival; PFS, progression-free survival; TTD,				

5.2.9.4 Additional scenarios requested at points for clarification

The company provided an updated economic model in the updated response to clarifications. The updated model includes a scenario whereby the five efficacy-evaluable adult primary CNS tumour patients and the seven paediatric patients have been added to the model, as per the ERG's request. This was requested as the ERG believes these patients with primary CNS tumours and paediatric patients fall within the population in which the company is seeking a recommendation. For a full

discussion of the population included in the economic analysis, see Section 5.2.3. The inclusion of the 12 patients resulted in a decrease in the company's base case ICER to £49,358. These results can be seen in Table 47.

The ERG also had some concerns regarding the company's assumption that a proportion of patients receiving second-line therapy following entrectinib will continue to receive the second-line therapy until death. The company acknowledged that this is a conservative assumption, and that the clinical plausibility of this is low. As a result, two alternative scenarios were provided in which the duration of subsequent therapy is limited to 6 months and 3 months. This reduced the ICER to £40,093 and £39,849, respectively.

Table 47 Additional scenarios following clarification questions

Parameter	Value	Incremental Costs	Incremental QALYs	ICER (£/QALY)
Inclusion of paediatric and CNS	Base case: paediatric and CNS patients excluded	[redacted]	[redacted]	£52,609
Inclusion of paediatric and CNS	Paediatric and CNS patients included	[redacted]	[redacted]	£49,358
Duration of subsequent therapy	Base case: until death	[redacted]	[redacted]	£52,609
Duration of subsequent therapy	6 months	[redacted]	[redacted]	£40,093
Duration of subsequent therapy	3 months	[redacted]	[redacted]	£39,849
ICER, incremental cost-effectiveness ratio; QALYs, quality-adjusted life-years; CNS, central nervous system				

5.2.10 Model validation and face validity check

The company stated that the cost-effectiveness analysis was validated in a number of ways. The internal validity of the model processes was assessed by an external consultancy company, who undertook a technical validation of the model (including pressure testing using extreme values, formula checking, and cell references). In addition, the validation of entrectinib extrapolations, comparator choice and data, and tumour type proportions presented in the integrated analysis were also described by the company (Section B.3.3.6 of the CS).

The clinical plausibility of the survival curves for entrectinib was discussed with investigators at two UK sites from the STARTRK-2 study, through their visual inspection of all six extrapolations of the PFS and OS curves for entrectinib, with emphasis placed on OS extrapolation due to its importance in the model. No details were provided as to why certain distributions were rejected. These investigators also noted that the frequencies of the tumour types seen may reflect clinical practice, with the possible exception that MASC is overrepresented.

The specific treatment choices for each tumour type were discussed with a clinical expert in each of the following tumour types: non-small cell lung cancer, breast cancer, sarcoma, thyroid cancer, neuroendocrine tumours, colorectal cancer, and pancreatic cancer. The choices of comparator were kept broadly in line with the therapies listed in NICE Pathways, with deviations based on recommendations from clinical experts. The company also stated that their clinical experts endorsed the survival data extracted for each comparator, with the caveat that some comparator OS outcomes were confounded by crossover, and therefore exhibited better-than-expected outcomes where adjusted data could not be found.

In addition to the clinical plausibility of the extrapolated outcomes, selection of the appropriate distributions has been driven by statistical fit to the data, and the company presented a comparison of modelled and trial-based OS and PFS for entrectinib. Modelled PFS appeared to represent the clinical data well throughout the trial period up to around 12 months, after which the modelled PFS appeared higher than the trial PFS, with the degree of this overestimation varying by distribution. The OS data was less mature, and therefore it was more difficult to assess its predictive validity.

As noted in Section 5.2.6, the company's selection of an exponential distribution for modelling OS and PFS resulted in patients remaining in the progression health state longer than the PFS health state (more than twice as long, in the comparator arm). The ERG did not consider this to appear plausible, given the end stage of the pathway at which patients are treated.

5.3 Conclusions of the cost effectiveness section

The modelling of a histology independent indication such as the one covered by the present decision problems generates a number of significant challenges that impact greatly on the validity of the ICERs generated in the company's presented economic analysis. The results of the economic model are therefore subject to very considerable uncertainty and may differ significantly from those presented in the company's base-case. Further, the single ICER presented by the company conceals the potentially significant variation in the tumour specific ICERs, driven by a combination of factors, particularly variability in relative effectiveness between tumour types and testing costs. As stated in Section 5.2.1 it is the ERG's general view that optimised decisions are preferable, and while the ERG acknowledges the challenges presented by the current decision problem, it considers that the company could have gone further in justifying the use of a single ICER. In particular, the ERG suggests the company could have explored variability in the treatment effect across tumour types, and how testing costs are likely to impact on the cost-effectiveness of specific tumour types. An overview of the key uncertainties identified by the ERG are presented below.

1) *Heterogeneity in the treatment effect*

A central issue of the current appraisal is the potential for heterogeneity in the treatment effect across tumour types, as well as across other clinical characteristics such as age (paediatric vs adults), fusion type, and position in the treatment pathway. As demonstrated in the ERG exploratory analyses on response data (see Section 4.3), there is evidence to suggest that the treatment effect is heterogeneous across tumour types. Furthermore, the predictive distribution, which provides an estimate of the likely response rate in an unrepresented tumour has a credible interval of [redacted], implying that mean response across all eligible tumour types could be very different to that estimated in the integrated analysis. This has significant implications for the economic analysis and suggests that the tumour specific ICER will vary significantly.

2) Uncertainties surrounding the comparability of comparator effectiveness evidence

The ERG's has several concerns about the representativeness of the modelled population, which was based on the integrated efficacy analysis. These include concerns about the distribution of tumour types modelled, which appear to over represent some tumour types, while under-representing others. Further, the modelled population includes only the 13 tumour types included in the EEA dataset, while there is evidence to suggest that *NTRK* fusions occur in at least another 11 tumour types, representing a minimum of 20% of the eligible population. The omission of these patients has a number of implications for the model and potentially impacts upon a number of inputs used to model established management, including comparator effectiveness, comparator treatment cost, testing costs, and health state utilities. The ERG is also concerned that the analysed integrated efficacy data set excluded available evidence on patients with [redacted].

3) Uncertainties surrounding the relevance of selected comparators

There are significant uncertainties regarding whether the appropriate comparators have been modelled. The anticipated marketing authorisation for entrectinib allows it to be used at multiple points in the treatment pathway, meaning there is significant uncertainty regarding the patient group in which entrectinib may be used in practice. It is therefore unclear whether the modelled comparators represent current NHS practice. Furthermore, because the model only considers 13 tumour sites and not all tumour sites in which *NTRK* fusions may occur, there are a number of relevant comparators not covered by the model. The model therefore implicitly assumes that the modelled population is representative of the eligible population, which appears to be unlikely given available evidence on the distribution of tumour types with *NTRK* fusions.

4) Uncertainties surrounding the comparability of comparator effectiveness evidence

Because the available effectiveness evidence for entrectinib was from single arm studies, it was necessary to generate an appropriate comparator dataset. The company does this by using previous NICE TAs as a source of effectiveness data, which were then weighted by the distribution of tumour

types in the integrated efficacy analysis. While the ERG considers the broad approach adopted by the company to be reasonable, there are significant challenges associated with implementing this successfully, as well as further issues resulting from the company's execution of this approach.

The ERG's principal concerns regarding the company's approach to generating a comparator is that it relies on an unadjusted naïve comparison between the weighted comparator and the integrated efficacy analysis with significant scope for confounding bias. The ERG in particular notes that a significant proportion of the patients in the integrated efficacy population (37.0%) received entrectinib as a first-line systemic therapy, while the comparator data set draws predominantly from patients in later lines of therapy. Further, the use of NICE TAs as source of effectiveness evidence means that comparator effectiveness data is being drawn from a population who are primarily *NTRK* fusion negative. This is problematic because there is evidence to suggest that *NTRK* fusions are prognostic, with variable impact upon prognosis depending upon tumour type.

Because of these significant concerns about confounding bias and the challenges of generating a truly comparable comparator data set the ERG considers that the company should have also considered other approaches to generating a comparator data set. The company could for example have considered two approaches discussed in a recent publication by Hatswell *et al.* [1] which suggest using evidence from non-responders and on patients' time to progression on previous lines of therapy to further explore the uncertainties associated generating a comparator data set.

5) Uncertainty surrounding the extrapolation of OS data for entrectinib

The ERG highlights that the observed data for entrectinib was immature, with median OS not yet met. As such, there is significant uncertainty regarding the longer-term survival benefits of entrectinib. The company base-case fits an exponential function to the available KM which was selected from a range of standard parametric functions on the basis that the exponential function has the best statistical fit to the observed data. The ERG considers that the exponential function represents a potentially plausible extrapolation of OS, but is concerned that it implies that post-progression survival is significantly longer than pre-progression survival. The ERG questions the clinical plausibility of this given that entrectinib therapy is discontinued on progression and that only [redacted] of patients received any subsequent therapy. The ERG's preference is therefore for the Weibull function, which produces a more reasonable balance between pre- and post- progression survival, while also having good statistical fit to the observed data.

6) Uncertainty surrounding the appropriate testing strategy and applied testing costs

The ERG also has substantive concerns regarding the companies approach to modelling *NTRK* fusion testing. The ERG in particular is concerned that the company appears to have included extensive

testing costs in the comparator arm of the model. The ERG considers that the focus of modelled testing costs should be on the incremental testing costs associated with identifying *NTRK* fusion positive patients.

Furthermore, it is not clear whether the primary strategy proposed by the company of using IHC followed by NGS will reflect NHS practice should *NTRK* be recommended for use on the NHS. The ERG notes that there are a range of alternative testing strategies that have been discussed in the literature with consequences for the incremental costs of implementing *NTRK* fusion testing as well diagnostic accuracy.

7) Uncertainty surrounding broader infrastructure requirements

The implementation of an appropriate testing regime to identify patients with *NTRK* gene fusions would likely require a significant increase in molecular testing with between 50 and 92 thousand patients potentially eligible for testing. Regardless of the testing strategy adopted for *NTRK* fusions, this is likely to require a significant expansion of current testing service capacity, which would potentially require further investment infrastructure and/or training. These costs are, however, not considered in the company's economic analysis.

6 Impact on the ICER of additional clinical and economic analyses undertaken by the ERG

6.1 Overview

This section details the ERG's further exploration of the assumptions and uncertainties raised in the review and critique of the company's cost-effectiveness analysis, presented in Section 5. This section is organised in four parts. Section 6.2 details the impact of correction of errors identified in ERG's validation of the executable model and other amendments to the company base-case analysis. Section 6.3 details a series of scenario analyses exploring the robustness of the cost-effectiveness results under specific assumptions and additional uncertainties identified by the ERG. These analyses were conducted within the company-corrected base-case analysis as presented in Table 44 in Section 5.2.9.1 . The scenario analyses presented in Section 6.3 focus on exploring the following issues:

- An alternative distribution of tumour types;
- Testing costs to identify *NTRK* fusion patients;
- Estimation of treatment-related costs.

In Section 6.4, the ERG alternative base-case is presented, which combines a number of exploratory analyses presented in Section 6.3 and alternative assumptions provided in the company exploratory analyses.

Further exploratory analyses in the context of the ERG alternative base-case analysis are presented in Section 6.5. This section presents the implementation of an alternative model structure for estimating outcomes in the established management arm. In addition, the ERG presents additional statistical analyses of the results of the economic model, including estimating the value of heterogeneity and net population benefit in the ERG's alternative base-case economic model.

Due to time constraints, ICERs based on deterministic analyses are presented throughout this section.

There are a number of treatment options in the established management arm that are associated with a confidential PAS. These include eribulin, everolimus, nintedanib, nab-paclitaxel, trabectedin, and trifluridine and tipiracil. The results of these analysis with the cPAS applied are presented in a confidential appendix to this report.

6.2 ERG corrections and adjustments to the company's base case model

The ERG identified a minor error in the executable model, pertaining to the estimation of post-progression second-line treatment costs in the entrectinib arm. The model effectively applied the discount rate twice to these costs in addition to the omission of the drug administration costs. At the clarification stage, the company provided a corrected model, which also incorporated the results of a

survival analysis for entrectinib based on a more recent data cut. The impact of this correction was minor: the company base-case ICER was reduced from £54,646 to £52,609 (discussed and presented in Section 5.2.9.1, Table 44).

Subsequent analyses in this section are based on this corrected, updated analysis.

6.3 Additional ERG analyses

6.3.1 Alternative distribution of tumour types

The company base-case analysis uses the distribution of tumour types from the integrated efficacy analysis to estimate a weighted set of outcomes for established management. The ERG presents a scenario based on a plausible alternative distribution of tumour types, which was estimated using observed *NTRK* fusion frequencies provided in the FMI data set and published cancer statistics for England. The FMI database is considered by the ERG to be more representative as it is based on a large sample of 166,000 patients. The alternative distribution used by the ERG, along with the original proportion of tumour types provided in the CS, can been seen in Table 29, Section 5.2.3 The method used to estimate this can be found in Appendix A: ERG estimates of eligible population.

The impact of incorporating this alternative distribution of tumour types resulted in a re-estimation of the weighted outcomes for the comparator arm: it was not possible to reweight the outcomes in the entrectinib arm as OS and PFS data were not available by tumour type. This approach therefore implicitly assumes homogeneous PFS and OS across tumour types.

The results of this analysis are presented in Table 48. This scenario was associated with greater incremental costs and lower incremental QALYs than the base-case analysis. The difference in the cost was mostly driven by the large decrease in the proportion of patients with sarcoma: a tumour type associated with higher total costs than the other tumour types in the established management arm.

Table 48 Results of the ERG analyses on the distribution of tumours

	Inc costs	**Inc QALYs**	**ICER**
Base case	[redacted]	[redacted]	£52,609
Scenario 1: Alternative distribution of tumour types	[redacted]	[redacted]	£69,747
Inc., incremental; QALYs, quality adjusted life years; ICER, incremental cost effectiveness ratio			

6.3.2 Testing costs to identify *NTRK* fusion+ patients

Marginal costs of testing

The ERG considers that the focus of applied testing costs should be on the incremental testing costs associated with identifying *NTRK* fusion positive patients. The ERG therefore implements a scenario in which testing costs are removed from the established management arm. As WGS is currently funded for sarcoma and paediatric patients, zero incremental costs are assumed for sarcoma and paediatric patients in this scenario. Similarly, the testing costs for MASC patients were removed from the model for this scenario as current testing already identities *NTRK* fusions in these patients. The results of this analysis are presented as Scenario 2 in Table 49.

Removal of testing costs of NGS for lung cancer patients

As discussed in Section 5.2.8, the cost of adding a new *NTRK* panel to an RNA-based NGS test is negligible. Currently, lung cancer is the only tumour type of those included in the efficacy evaluable data set where RNA-based NGS is available for a specific subgroup of patients with NSCLC. A scenario is therefore presented where no additional costs would apply for lung cancer patients. The results of this analysis is presented as Scenario 3 in Table 49. In this scenario, only marginal costs of testing are applied (as per Scenario 2).

Confirmatory NGS following WGS

The ERG received clinical advice that WGS cannot be used to confirm the presence of *NTRK* fusions at present (see Section 2.2.2) and that a confirmatory NGS test would be required for patients receiving WGS. Scenario 5 in Table 49 presents the results of including a confirmatory RNA-based NGS test in patients who already receive WGS. It is assumed WGS will remove 89% of *NTRK* fusion negative samples, reducing the requirement for RNA-based NGS confirmatory testing to 11% of the NNS population. This figure is based on the company's assumptions for IHC as the ERG were unable to identify any statistics on the performance of WGS. In this scenario (Scenario 4 in Table 49), only marginal costs of testing are applied (as per Scenario 2).

Numbers needed to screen

The discovery of the *NTRK* gene fusion is a relatively recent one and the frequency of the fusion in tumour types is still being established. As a result, there remains a degree of uncertainty regarding the exact frequencies used in the model. The number of patients who require screening to identify one individual with an *NTRK* fusion varies depending on the frequency of the gene fusion. The ERG estimated an alternative set of prevalence rates for each tumour type, with details provided in Appendix B. The results of this analysis are presented as Scenario 5 in Table 49.

Cost of testing in whole NTRK population

As outlined in Section 3.1 a number of tumour types are not represented in the model, as such the testing costs represent this population rather the [redacted]. Using data from the FMI database, the ERG implemented a scenario where testing costs are estimated based on all tumour types know to harbour *NTRK* fusions. In this scenario the distribution of tumour types is also assumed to align with ERG estimates of the NNS presented in Appendix B. In unrepresented tumour types, the ERG assumes that patients will receive IHC followed by confirmatory RNA-based NGS, unless WGS is already available on the NHS. The results of this analysis are presented as Scenario 6 in Table 49.

Removal of testing costs

The ERG also included the scenario in which all testing costs were removed (Scenario 8). This represents a future practice scenario where screening is routinely carried out on NHS. Clinical advice to the ERG, however, suggests that this is not likely to happen in the near future. The inclusion of this scenario represents a potential lower bound for the estimate of cost-effectiveness.

Identifying paediatric glioma patients

At the clarification stage, the ERG requested that the company present an analysis that includes primary CNS and paediatric patients. For the purposes of the model, the company grouped the paediatric primary CNS patients with the adult primary CNS patients for the weighted comparator costs and outcomes, since common comparators were assumed for these patients.

Following the factual accuracy check, the company highlighted that screening costs for glioma are overestimated since the costs represent a mixture of adult (five) and paediatric (four) primary brain tumours; screening costs for paediatric gliomas are significantly lower to due to inclusion in the genomic test directory.

The company analysis presented following the clarification stage applied the cost of IHC and confirmatory NGS to identify these patients (Scenario 9 in Table 50). The ERG has implemented a scenario whereby the costs of testing for paediatric glioma patients were based on WGS, which had the impact of removing the cost in these patients. As a result, the ICER was reduced from £49,358 to £48,860 per QALY.

Results

The results of the scenarios described above are presented in Table 49.

Table 49 Results of the ERG analysis on testing costs

Scenario	Inc cost	Inc QALY	ICER
Base case	[redacted]	[redacted]	£52,609
Scenario 2: Remove testing costs in comparator arm	[redacted]	[redacted]	£63,329

Scenario 3: Remove lung cancer cost of testing	[redacted]	[redacted]	£59,465
Scenario 4: Confirmatory RNA-based NGS in WGS patients	[redacted]	[redacted]	£64,608
Scenario 5: Prevalence of *NTRK* fusions (tumour types represented in the trial)	[redacted]	[redacted]	£56,914
Scenario 6: Prevalence of *NTRK* fusions (based on the whole *NTRK* population)	[redacted]	[redacted]	£65,981
Scenario 7: Cumulative impact of 2, 3, 4, 6	[redacted]	[redacted]	£64,115
Scenario 8: No testing costs	[redacted]	[redacted]	£36,914

Table 50 Results of the ERG analysis on testing costs of paediatric patients

	Inc costs	Inc QALYs	ICER (£/QALY)
Company base-case scenario including CNS patients and children (Table 47 in Section 5.2.9.4)	[redacted]	[redacted]	£49,358
Scenario 9: WGS for identifying *NTRK* tumours in paediatric patients	[redacted]	[redacted]	£48,860

6.3.3 Treatment costs

eMIT costs for therapies in established management arm

Drug acquisition costs for the comparator therapies were obtained from the BNF. However, many of these therapies are generic products that are widely available to the NHS at discounted prices. The Department of Health's eMIT database provides information on the average price paid by the NHS for pharmaceuticals, which can differ from the list prices listed in the BNF and are a more representative estimate of drug expenditure. Unit costs from eMIT (presented in Table 36 in Section 5.2.8) are generally considerably lower than those in the BNF, and the use of the BNF costs will overestimate drug expenditure, biasing the analysis in favour of entrectinib. The results of this analysis are presented as Scenario 10 in Table 51.

Drug wastage

Wastage has the potential to significantly impact upon drug expenditure, and the ERG is concerned that the company's model, which excludes drug wastage in the base-case analysis, underestimates the drug costs that would be incurred by the NHS. The ERG explored a scenario that allowed for drug wastage. The results of this analysis are presented as Scenario 11 in Table 51.

Results

The results of the scenarios described above are presented in Table 51. The analysis was not sensitive to the inclusion of eMIT unit costs for the therapies in the established management arm; however, the

inclusion of drug wastage for entrectinib resulted in an increase to the ICER of approximately £2,750, due to the additional drug costs in this arm.

Table 51 Results of the ERG analysis on treatment costs

Scenario	Inc costs	Inc QALYs	ICER
Base case	[redacted]	[redacted]	£52,609
Scenario 10: eMIT costs for comparator therapies	[redacted]	[redacted]	£52,081
Scenario 11: With drug wastage	[redacted]	[redacted]	£55,357

6.4 ERG alternative base-case

Table 52 presents the results of the ERG alternative base-case analysis. These incorporate a number of changes to key model parameters and assumptions, which were previously explored individually in Section 6.3, as well as a number of additional assumptions which were previously explored by the company in scenario analyses (Section 5.2.9.3). Most notably, this includes an alternative extrapolation of available PFS and OS data for entrectinib. As discussed in Section 5.2.6, the ERG considered the Weibull function to be a more preferable model for OS and PFS, as it uses the more reasonable assumption of increasing hazards over time, and resulted in a more plausible estimate of time spent in the post-progression health state.

The ERG alternative base-case analysis includes the following changes to the company base-case analysis:

- Inclusion of children and primary CNS tumours in the population (see Section 5.2.3),
- Weibull distribution for entrectinib OS and PFS (Section 5.2.6),
- Inclusion of marginal testing costs only,
- Confirmatory RNA-based NGS test after WGS test, and removal of NGS testing costs for lung cancer patients,
- WGS test to identify *NTRK* tumours in paediatric patients,
- Testing costs estimated using the number needed to screen based on the whole *NTRK* population,
- Second-line therapy following discontinuation of entrectinib, limited to 6 month duration,
- eMIT costs for therapies in the established management arm,
- Inclusion of drug wastage for entrectinib.

Under the ERG's alternative set of assumptions, the ICER for entrectinib versus established care is £77,120 per QALY.

Table 52 ERG alternative base-case analysis

	Total costs	Total QALYs	Incremental costs	Incremental QALYs	ICER (£/QALY)
Entrectinib *	[redacted]	[redacted]	[redacted]	[redacted]	£77,109
Established management	£19,853	1.03	-	-	-
* Note that these results have changed following the factual accuracy check to include the change made in Scenario 9					

6.5 Exploratory analysis on ERG base-case

6.5.1 Estimation of comparator outcomes based on a response model

Motivation

As discussed in Section 5.2.6.1, the ERG considers that the company should have also considered alternative methods of generating comparator effectiveness estimates. These could include a dual-partitioned response-based model, which distinguishes between responders to treatment and non-responders. This additional model complexity allows for a distinction to be made in the health-related quality of life (HRQoL) of responders and non-responders, as well as allowing for potential differences in the costs of care.[1] An alternative would be based around a surrogate relationship between response and PFS and OS. In the FDA evaluation of larotrectinib, it was considered that these surrogate relationships were reasonably likely to predict meaningful benefit. However, this approach has its own drawbacks: a review of the relationship between the more long-term outcomes of PFS and OS suggested that it varies considerably by cancer type and is not always consistent even within one specific cancer type.[53] In the absence of specific guidance on the surrogate relationship between response and survival, the use of this type of model structure would need to be accompanied by a review of studies in *NTRK* fusion patients to consider the extent to which response-based outcomes can be considered a robust surrogate endpoints for PFS and OS, and to establish how these relationships might be quantified in a modelling approach.

The ERG implemented an exploratory responder-based approach, which uses effectiveness data on non-responder patients as a proxy for patients not receiving an active treatment. The ERG recognises that such an approach is subject to limitations particularly regarding the maturity of the data and the number of patients included in the analysis, but believes that presenting the results of this analysis can provide some degree of reassurance regarding the predicted comparator effect estimates, given the large amount of uncertainty in the approach taken in the company base case. This approach ensures that the population used to model the comparator and the intervention arms are consistent with each other, which is of particular importance given that the prognostic status of *NTRK* fusion tumours is unknown and likely to differ based on each tumour type. However, by ensuring that the efficacy in

both arms is reflective of the trial, it does limit the applicability of the findings to the general eligible population.

As described in Section 5.2.1, a further advantage of the response-based model is that it is easier to generate ICERs specific to each tumour type. This was implemented in the response-based model by altering the rate of response in the model and subsequently changing the survival predictions for the entrectinib arm. As in the model above, survival in the established management arm was modelled assuming a 0% response in the comparator arm. This analysis therefore makes the strong assumption that effectiveness for established management is the same across all tumour types. The assumed response for each tumour type was based on the Bayesian hierarchical analysis presented in Section 4.3.1.

Methods

The ERG constructed a response-based model using the heterogeneous response rates across tumour types, estimated by the BHM in Section 4.3.1, and linked these to OS and PFS. This method facilitated linking response to costs and QALYs and so create histology specific estimates of cost effectiveness. It should be emphasised that this model is for illustrative purposes and given the data made available it has been necessary to make strong assumptions to explore heterogeneity.

At the clarification stage, the ERG requested that the company provide KM plots for PFS and OS for non-responders and responders to entrectinib. The company provided this information for the population in the original analysis, and the population that also includes paediatric patients and patients with primary CNS tumours.

The ERG reconstructed the individual participant data (IPD) for the population including paediatric patients and patients with primary CNS tumours, and fit standard parametric models (exponential, Weibull, log-logistic, lognormal, Gompertz, and generalised gamma) to the dataset for each outcome, using response status as a covariate. Survival in the established management arm was assumed to be equivalent to that of the non-responder patients, and survival in the entrectinib was estimated as a weighted average of survival in the responder and non-responder patients, weighted by the estimated response rate of [redacted] from the BHM described in Section 4.3.1.

To determine the most appropriate model, the ERG referred to fit statistics (AIC and BIC, Table 53 below), visual fit to the observed KM curves, and clinical plausibility of survival estimates. Figure 20 a graphical summary of each curve and their fit to the observed KM data. For PFS, generalised gamma had the best statistical fit; however, it appears to produce long-term projections that were considered overly optimistic and, therefore, implausible. The fit statistics for OS favoured the

lognormal and the exponential. However, the lognormal did not appear to fit well when compared to the responder population.

Figure 20 Survival extrapolations

Figure redacted

Table 53 Fit statistics for survival models fit to the whole population in the integrated analysis

	Overall survival		**Progression-free survival**	
	AIC	**BIC**	**AIC**	**BIC**
Generalised gamma	209.5179	218.2766	275.3453	284.1039
Weibull	210.2889	216.8579	284.6180	291.1870
Exponential	209.3053	213.6846	283.5421	287.9214
Loglogistic	208.5645	215.1334	281.5299	288.0988
Lognormal	208.2129	214.7819	279.6231	286.1920
Gompertz	210.9277	217.4967	285.5215	292.0904

On the basis of the plausibility of the long-term predictions, comparisons with the KM plots for the responder and non-responder population, and for consistency with the assumptions made in the ERG alternative base-case analysis, the Weibull survival function was selected to model both PFS and OS in this exploratory cost-effectiveness analysis. Figure 21 compares the predicted survival for entrectinib and established management (using the Weibull distribution) compared with the predicted survival for each arm as used in the ERG alternative base-case analysis in Section 6.4. The two methods result in similar survival extrapolations for entrectinib; however the survival curve for established management estimates higher survival in the response-based model.

Figure 21 Comparison of survival functions used in the ERG base case and the responder-based cost-effectiveness analysis

Figure redacted

Results

The analysis was based on the assumptions made under the ERG alternative base-case analysis set out in Section 6.4.

The ICER in this analysis was £95,723 per QALY (Table 54). The costs and QALYs generated for the entrectinib arm were similar to that of the ERG base-case analysis in Table 44. However, this method produces higher QALYs and costs in the established management arm as a result of the higher rate of survival for these patients. Consequently, the ICER that was estimated using this method is higher than that using the same assumptions under the model structure presented by the company.

Table 54 Results of responder-based cost-effectiveness analysis

	Total costs	**Total QALYs**	**Incremental costs**	**Incremental QALYs**	**ICER (£/QALY)**
Entrectinib	[redacted]	[redacted]	[redacted]	[redacted]	£95,709
Established management	£28,507	1.32	-	-	-
* Note that these results have changed following the factual accuracy check to include the change made in Scenario 9					

The results of the analyses by tumour type are presented in Table 55. The ICERs ranged from £57,451 per QALY for sarcoma patients, to £128,663 for thyroid tumours.

Table 55 Results of the responder model by tumour type

Tumour type	**ICER ***
CRC	£98,493
MASC	£111,464
Thyroid	£128,663
NSCLC	£89,668
Pancreatic	£89,770
Sarcoma	£57,451
Neuroendocrine	£108,634
Breast	£86,697
Glioma	£117,456
IFS	£119,787
Melanoma	£114,868
Other	£98,164
All tumours	*£95,709*
* Note that these results have changed following the factual accuracy check to include the change made in Scenario 9	

6.5.2 Value of heterogeneity and net population benefit

The costs and health consequences associated with using entrectinib in colorectal cancer (CRC) with *NTRK* fusions are used to illustrate the importance of taking account of heterogeneity in histology independent assessments. CRC was chosen as it had a low predicted response rate compared to other tumour types, and accounted for a large proportion of patients eligible for treatment with entrectinib. From the responder-based model, entrectinib was associated with [redacted] in extra costs and [redacted] additional QALYs in CRC. The additional QALYs from entrectinib are lower in CRC as the estimated response rate for these patients, as estimated from the hierarchical model, is [redacted], which is below the pooled response rate across all tumour types, which is [redacted]%. This results in an ICER of approximately £95,451 per QALY for CRC. This can be compared to the additional costs and QALYs associated with using entrectinib across all tumour types. For this pooled *NTRK* population (which includes CRC) entrectinib is associated with [redacted] in extra costs and [redacted] additional QALYs. This results in an ICER of approximately £93,532 per QALY.

This simple comparison of a subgroup specific ICER to a pooled population ICER illustrates that the cost-effectiveness of entrectinib could vary significantly between individual tumour types. This also means that the 'average' ICER could be more favourable if the subgroup with a CRC histology were excluded. To understand the implications of this for population health requires that benefits and costs are expressed as net health effects (NHE). The NHE is the difference between any health gained with the intervention and the health forgone elsewhere in the health-care system, all expressed in QALY terms. With an ICER in CRC of approximately £96,451 per QALY, the incremental NHE at a threshold of £50,000 is -0.45 QALYs per patient, that is, the additional health gained with the intervention is more than offset by health forgone elsewhere. This means that for every CRC patients who receives entrectinib, 0.45 QALYs could potentially be lost elsewhere in the health system.

Table 56 Value of heterogeneity – an illustrative example using the CRC population

	Total cost	Total QALYs	ICER	NHE (QALYs)	Incremental NHE (QALYs)
Entrectinib	[redacted]	[redacted]	[redacted]	[redacted]	-0.45
Established management	£32,460	1.43	-	0.785	-
Note, a threshold of £50,000 was used to estimate NHE					

The advantage of NHE is that they can be used to help understand the population level consequences of decisions. The number of CRC patients with *NTRK* fusions in the UK was estimated by the ERG to be approximately 29 per year (Appendix A). This means that an 'non-optimised' recommendation which includes CRC might result in an additional 12.99 QALYs per year to the health system

compared to established management. In other tumour types, entrectinib may provide positive QALYs to the health system but further analysis would be required to identify these tumour types, if they exist. As the cost effectiveness of entrectinib could depend on the tumour type treated, this analysis also illustrates the importance of understanding the distribution of tumour types expected to receive the treatment in practice.

6.6 Conclusions from ERG analyses

The ERG has presented a number of additional analyses carried out in stages. These exploratory analyses were undertaken on a model provided by the company at the clarification stage, which addressed an error identified by the ERG, and included a more recent data cut of the survival data from the integrated efficacy analysis. The impact of these changes was to decrease the ICER from £54,646 to £52,609 per QALY.

Using the corrected and updated model, the ERG then presented a number of analyses considering a range of issues raised in Section 5.2. These scenario analyses addressed the following issues:

- An alternative distribution of tumour types;
- Testing costs to identify *NTRK* fusion+ patients;
- Estimation of treatment-related costs.

The scenarios associated with the greatest impact on cost-effectiveness outcomes related to changes made by the ERG, which involved the removal of testing costs in the comparator arm to more accurately reflect the incremental cost of testing to identify *NTRK* fusions. Testing costs comprise a significant proportion of the total costs, and removing the testing costs for the comparator resulted in the ICER increasing from £52,609 to £63,329. A scenario analysis that explored the impact of an alternative distribution of tumour types demonstrated that the results of the model are sensitive to this assumption. This sensitivity was a consequence of tumour types being associated with different QALYs or costs, highlighting the heterogeneity in this patient population.

The ERG alternative base-case implemented a number of alternative assumptions that were included in the company exploratory analyses. The assumptions that had the largest impact on the ICER was the restriction of duration of second-line therapy to 6 months following discontinuation of entrectinib, and the implementation of a Weibull survival model to estimate overall and progression-free survival of entrectinib. This analysis estimated entrectinib to be more costly (cost difference [redacted]) and more effective ([redacted] QALY gain) compared with established management, and suggests that the ICER for entrectinib compared with established management is £77,109 per QALY.

The final part of this section carried a further series of exploratory analyses that explored the impact of an alternative method to estimate survival. This method used the survival of non-responder patients

to estimate survival predictions in the established management arm. The entrectinib arm was based on a weighted average of responder and non-responder survival predictions, which allowed for the exploration of cost-effectiveness in different tumour types by varying the response rate used to estimate the weighted average. The ICER for the pooled group was £95,705 per QALY. This was higher than the ICER estimated in the ERG analysis, as a result of the higher survival rates predicted by the response-based model for the established management arm. When varied by tumour type, the ICERs ranged from £57,451 per QALY for sarcoma patients, to £128,663 for thyroid tumours.

7 End of life

In the CS and clarification response, the company state that entrectinib meets the end-of-life criteria compared to the current established management across all patients potentially eligible for entrectinib, on the basis of the results of the integrated efficacy analysis.

The conventional application of end-of-life (EoL) criteria to a highly heterogeneous population with no established comparators is challenging in two respects. Firstly, the EoL criteria may apply across some tumour types and not others. Secondly, there is a great deal of uncertainty around estimates of both life expectancy and extension to life which may vary widely by tumour type, this is further exacerbated by the uncertainty around the positioning of entrectinib in the treatment pathway for each tumour type. While there is little precedent for decision optimisation on the basis of the eligibility of sub-populations for EoL, it does not appear appropriate to apply a higher willingness-to-pay threshold to sub-populations that do not otherwise meet the necessary criteria based on the 'unmet need' of other included cancer types. Application of the higher willingness-to-pay threshold in such cases necessarily implies that patients are able to access therapy that otherwise would be considered cost-ineffective based on conventional thresholds and potentially raise issues about equity of access treatment, as QALY generated in *NTRK* positive and *NTRK* negative patients are being valued differently.

Application of the higher EoL threshold across all tumour types regardless of whether they all meet EoL also potentially offers as a commercial advantage to histology independent products as competitor products for tumour types not meeting EOL, which are required to be priced in accordance with conventional £20,000 to £30,000 thresholds, and as a consequence potentially distorts investment incentives towards histology independent therapies.

Criterion 1: The treatment is indicated for patients with a short life expectancy, normally less than 24 months.

In the ERG's base-case analysis, the population who are anticipated to meet the eligibility criteria in the product license have an average mean OS of 20.89 months (median 15.7 months). In the ERG's

response-based model, the mean OS of those patients who did not respond to entrectinib was 24.72 months (median 19.9 months) using a Weibull function. The ERG favours the use of the mean to represent life expectancy, as it better represents the distribution of OS, and measures of health benefit upon which decision about cost-effectiveness are made on the basis of mean values (mean QALYs). While the base-case mean survival falls under the two years stipulated in Criterion 1, the ERG does not consider this figure appropriate for decision-making for a number of reasons.

Firstly, as discussed in Section 3.3, the comparability of the comparator population with patients eligible for entrectinib in clinical practice is highly uncertain, particularly given the company's pooling of often very different life expectancy data from TAs covering multiple lines of therapy within the same indication. As a testing strategy will dictate when entrectinib is made available, it will always be at the same point in the pathway, making one comparator life expectancy estimate more appropriate than the other.

Secondly, the company anticipate NICE's recommendation to cover all tumour types affected by *NTRK* gene fusions. However, the majority of tumour types are not represented in the company's trial or comparator searches, and therefore the life expectancy of these populations is unknown, and may be significantly different to those included in the CS.

The company also stated in their clarification response that the prognostic implications of *NTRK* gene fusions mean these patients are likely to have a lower OS than the population considered in the NICE appraisals. As previously discussed, the ERG does not consider existing literature to support the concept of *NTRK* as being consistently prognostic of a shorter life expectancy.

It is highly uncertain whether the presented average OS estimate represents life expectancy at the time patients would become eligible for entrectinib. Furthermore, it does not reflect the heterogeneity of life expectancy across tumour types when they reach eligibility for treatment. A summary of mean and median OS estimates by tumour type used in the ERG base-case analysis are presented in Table 57.

Table 57 Average SoC OS by tumour type

Tumour type	Median OS (months)	Exp. mean OS (months)
Breast	12.18	17.56
Colorectal	9.07	13.08
MASC	13.80	19.91
Neuroendocrine	39.61	57.14
NSCLC	10.65	15.36
Cholangiocarcinoma	17.23	24.86
Pancreatic	8.80	12.70
Sarcoma	14.30	20.63
Thyroid	30.95	44.65
CNS	7.95	11.46
Infantile fibrosarcoma	17.23	24.86
Melanoma	6.40	9.23
Total (EEA weights)	**15.74**	**22.71**
Total (ERG weights)	**16.39**	**23.64**

These OS estimates suggest that patients with thyroid and neuroendocrine tumours would not meet the first of the EoL criteria, and represent approximately 31% of the incident *NTRK* fusion population (see Section 3.1).

Criterion 2: There is sufficient evidence to indicate that the treatment offers an extension to life, normally of at least an additional 3 months, compared with current NHS treatment.

The CS states that the median OS has not yet been reached in the EEA dataset, however, based on the company's extrapolation of the latest data cut, the estimated mean OS is predicted to be [redacted] This suggests a mean OS benefit of [redacted]; therefore, the company conclude that entrectinib offers at least a three-month extension to life. The ERG base-case predicted a mean OS on treatment with entrectinib of 31.1 months, suggesting a mean OS benefit of 10.2 months.

The ERG's response-based model presented in scenario analysis demonstrates that this extension to life may not be consistent between tumour types, which was also the case in company scenario analyses in their clarification response (Appendix E Table 76). Patients who responded to entrectinib had a median OS (Weibull) of 25.2 months, suggesting an OS benefit of 5.3 months (non-responder median OS = 19.9 months). Mean OS benefit by tumour type in the ERG's base-case ranged between 6.40 months (CRC) and 8.80 months (MASC).

While the ERG notes the significant challenges in obtaining a robust estimate of the extension to life generated by entrectinib, these ranges of values suggest that extension to life across the majority of tumour sites is likely to be greater than 3 months. The benefit of entrectinib in unrepresented tumour types is unknown and cannot be assumed to be equal to that seen in the trials. Utilising the predictive distribution estimated in section 4.3.1 in the response based model, extension to life in unobserved tumour sites could potentially range from [redacted] months to [redacted] months. Some tumour sites may therefore not meet the 3-month criteria.

8 Overall conclusions

The clinical evidence for entrectinib is very limited. Most of the efficacy evidence comes from an *NTRK* positive subgroup of patients of a phase 2, uncontrolled basket trial. A total of only 66 *NTRK*-fusion positive patients across 13 tumour types were included in the efficacy evidence, and each of the tumour types was represented by between one and 13 patients. Overall, the trial evidence showed a clinically meaningful overall response rate across tumour types. However, there is considerable uncertainty regarding the extent to which the response observed translates into clinically meaningful survival benefits. OS, PFS and DOR data presented were immature. Despite substantial censoring and the small number of patients at risk in the tails of the Kaplan-Meier curve, the crossing of OS curves between entrectinib responders and non-responders is of some concern.

Due to limited evidence, there is considerable uncertainty about the precision of response and survival benefit estimates and heterogeneity by tumour type and line of therapy. The ERG explored heterogeneity in response rates between 13 tumour types using a Bayesian hierarchical model, which assumes the response probabilities are similar (i.e. exchangeable) across tumour types, rather than identical (the company's preferred assumption). Although the ERG's analyses found that response rates obtained were similar to those presented in the company submission, there was considerable uncertainty in the level of heterogeneity of response rates across tumour types. Therefore, the possibility that some tumour types could have response rates that differ significantly from the pooled [redacted] response rate cannot be excluded. Due to small numbers of patients and subgroups there was insufficient evidence to explore formally whether response and survival may differ by *NTRK* fusion subtype or line of therapy.

The company's updated base-case ICERs for entrectinib compared with established management and presented single ICER of £52,609 per QALY (inclusive of the confidential PAS) to cover all the [redacted] anticipated marketing authorisation.

The ERG's review of the company's presented analysis centred round the challenges associated with assessing cost-effectiveness in a histology independent indication and the uncertainties associated with the limit evidence effectiveness available. The ERG proposed an alternative base-case to address several of the key uncertainties identified. These included uncertainties associated with testing and identify patients with NTRK fusions. This analysis explored alternative estimates of the NNS, as well as testing costs in tumour types not represented in the trial. The ERG also considered a number of plausible modifications to the testing strategy proposed by the company. Uncertainty surrounding the extrapolation of survival data for entrectinib was explored, with the ERG preferring to model PFS and OS using a Weibull function instead of an exponential function proposed by the company. The ERG base-case also included a number of minor alterations to costs as a scenario analysis presented by the

company as part their clarification response which explored alternative assumptions regarding the duration of post progression therapy.

Despite the ERG's attempt to address all the relevant uncertainties, data limitations imply that some key uncertainties could not be fully explored. These unresolved uncertainties could potentially have a profound impact on the cost-effectiveness of entrectinib and would require further data to fully address.

First, the cost-effectiveness estimates are based on an uncontrolled comparison used, which used data from previous NICE TA's as a source of effectiveness data. The ERG, however, found that the population included in the comparator trials is unlikely to match the entrectinib efficacy population, notably due to the unknown prevalence of NTRK fusions in most of the comparator evidence, and the mismatch of previous lines of therapy with the treatment pathway. The ERG therefore has substantive concerns about the validity of the comparator effectiveness data and explored alternative methods of generating a comparator data set by modifying the company model structure so that PFS and OS outcomes were determined based on response to treatment. In this scenario analysis comparator outcomes were generated assuming that all patients were non-responders.

Second the ERG has concerns about the implicit assumption of a homogenous treatment effect across all tumour types and that the presentation of a single ICER conceals the potential for significant variation in tumour specific ICERs. To explore this uncertainty, the ERG, utilised the response base model to integrate the results of the Bayesian hierarchal analysis discussed above, and generate tumour type specific ICERs. This exploratory analysis showed that the tumour type specific ICER's varied significantly from £57, 451 per QALY in sarcoma to £128,663 per QALY in Thyroid cancer (ICER for all tumour types £95,723 per QALY). Methods for further exploring the heterogeneity in the ICER in using population NHE were also present in brief with an illustrative example presented in CRC. This considered the implications of an optimised decision in which CRC was excluded from any NICE recommendation.

Third there are a number of uncertainties relating to the population treated and the positioning of entrectinib in the treatment pathway. This has implications for the modelled comparators that are assumed to represented established management. The company stated that they expect entrectinib to be positioned towards the end of a patient's treatment pathway and this was reflected in the comparator data selected to represent established management. However, the anticipated marketing authorisation for entrectinib is ambiguous in this regard and potentially permits entrectinib to be used as a first-line therapy in several tumour types. Alterative assumptions about the position of entrectinib will necessarily have implications on the model including on comparator costs, effectiveness and HRQoL.

8.1 Implications for research

Exploring outstanding uncertainties differential response rates and survival benefits across tumour sites represents potentially valuable aim of any further research. While evidence from a large RCT would be preferred, it is acknowledged that it is unlikely to be feasible to conduct one in this population. However, a mature and appropriately powered basket trial recruiting patients with a wide range of tumour types in statistically sufficient numbers, and at clinically appropriate and consistent positions in treatment pathways will be necessary to assess heterogeneity of response to entrectinib to inform optimised decision making in future.

9 References

1. Hatswell AJ, Sullivan WG. Creating historical controls using data from a previous line of treatment - Two non-standard approaches. *Stat Methods Med Res* 2019:962280219826609. Available from: https://www.ncbi.nlm.nih.gov/pubmed/30698076
2. Amatu A, Sartore-Bianchi A, Siena S. NTRK gene fusions as novel targets of cancer therapy across multiple tumour types. *ESMO Open* 2016;**1**:e000023. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27843590
3. Pietrantonio F, Di Nicolantonio F, Schrock AB, Lee J, Tejpar S, Sartore-Bianchi A, et al. ALK, ROS1, and NTRK Rearrangements in Metastatic Colorectal Cancer. *J Natl Cancer Inst* 2017;**109**. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29370427
4. Musholt TJ, Musholt PB, Khaladj N, Schulz D, Scheumann GF, Klempnauer J. Prognostic significance of RET and NTRK1 rearrangements in sporadic papillary thyroid carcinoma. *Surgery* 2000;**128**:984-93. Available from: https://www.ncbi.nlm.nih.gov/pubmed/11114633
5. Prasad ML, Vyas M, Horne MJ, Virk RK, Morotti R, Liu Z, et al. NTRK fusion oncogenes in pediatric papillary thyroid carcinoma in northeast United States. *Cancer* 2016;**122**:1097-107. Available from: https://www.ncbi.nlm.nih.gov/pubmed/26784937
6. Vokuhl C, Nourkami-Tutdibi N, Furtwängler R, Gessler M, Graf N, Leuschner I. ETV6–NTRK3 in congenital mesoblastic nephroma: A report of the SIOP/GPOH nephroblastoma study. *Pediatr Blood Cancer* 2018;**65**:e26925.
7. Gatalica Z, Xiu J, Swensen J, Vranic S. Molecular characterization of cancers with NTRK gene fusions. *Mod Pathol* 2019;**32**:147-53. Available from: https://www.ncbi.nlm.nih.gov/pubmed/30171197
8. Okamura R, Boichard A, Kato S, Sicklick JK, Bazhenova L, Kurzrock R. Analysis of NTRK Alterations in Pan-Cancer Adult and Pediatric Malignancies: Implications for NTRK-Targeted Therapeutics. *JCO Precis Oncol* 2018;**2018**. Available from: https://www.ncbi.nlm.nih.gov/pubmed/30637364 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6329466/
9. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med* 2017;**23**:703-13. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28481359
10. Chmielecki J, Bailey M, He J, Elvin J, Vergilio JA, Ramkissoon S, et al. Genomic Profiling of a Large Set of Diverse Pediatric Cancers Identifies Known and Novel Mutations across Tumor Spectra. *Cancer Res* 2017;**77**:509-19. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28069802
11. Pavlick D, Schrock AB, Malicki D, Stephens PJ, Kuo DJ, Ahn H, et al. Identification of NTRK fusions in pediatric mesenchymal tumors. *Pediatr Blood Cancer* 2017;**64**. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28097808
12. Solomon JP, Hechtman JF. Detection of NTRK Fusions: Merits and Limitations of Current Diagnostic Platforms. *Cancer Res* 2019.
13. Bozzetti C, Nizzoli R, Tiseo M, Squadrilli A, Lagrasta C, Buti S, et al. ALK and ROS1 rearrangements tested by fluorescence in situ hybridization in cytological smears from advanced non-small cell lung cancer patients. *Diagn Cytopathol* 2015;**43**:941-6.
14. Selinger CI, Rogers T-M, Russell PA, O'toole S, Yip P, Wright GM, et al. Testing for ALK rearrangement in lung adenocarcinoma: a multicenter comparison of immunohistochemistry and fluorescent in situ hybridization. *Mod Pathol* 2013;**26**:1545.
15. England N. *National Genomic Test Directory*. 2019. Available from: https://www.england.nhs.uk/publication/national-genomic-test-directories/ [accessed 15th May 2019].
16. Penault-Llorca F, Rudzinski ER, Sepulveda AR. Testing algorithm for identification of patients with TRK fusion cancer. *J Clin Pathol* 2019. Available from: https://www.ncbi.nlm.nih.gov/pubmed/31072837
17. Marchiò C, Scaltriti M, Ladanyi M, Iafrate AJ, Bibeau F, Dietel M, et al. ESMO recommendations on the standard methods to detect NTRK fusions in daily practice and clinical research. *Ann Oncol* 2019.

18. Liu J, Weng L, Ming Y, Yin B, Liu S. Fusion Genes and Their Detection through Next Generation Sequencing in Malignant Hematological Diseases and Solid Tumors. *Diagnostic Pathology: Open Access* 2015;**1**:108.
19. Williams HL, Walsh K, Diamond A, Oniscu A, Deans ZC. Validation of the Oncomine™ focus panel for next-generation sequencing of clinical tumour samples. *Virchows Archiv: European Journal of Pathology*;**473**:489-503.
20. Bishop JA, Yonescu R, Batista DA, Westra WH, Ali SZ. Cytopathologic features of mammary analogue secretory carcinoma. *Cancer Cytopathol* 2013;**121**:228-33. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23042752
21. Excellence NIfHaC. *Gene expression profiling and expanded immunohistochemistry tests to guide selection of chemotherapy regimes in breast cancer management*: National Institute for Health and Clinical Excellence; 2011.
22. England N. *National Genomic Test Directory for Rare and Inherited Disease*. In: England Ni, editor.; 2019. Available from: https://www.england.nhs.uk/publication/national-genomic-test-directories/
23. Cross P, Chandra A, Giles T, Johnson S, Kocjan G, Poller D, et al. *Guidance on the reporting of thyroid cytology specimens*. London: Royal College of Pathologists; 2016.
24. Abulafi M, Ross P, Gaya A, Riddle P, Monahan K, Burling D, et al. *London Cancer Alliance Colorectal Cancer Clinical Guidlines* London; 2014.
25. Reid R, Bulusu R, Carroll N, Eatock M, Geh I, Judson I, et al. *Guidelines for the management of gastrointestinal stromal tumours (GIST)*. In; 2009. Available from: http://www.augis.org/wp-content/uploads/2014/05/GIST_Management_Guidelines_180809.pdf
26. Ganesan R, Singh N, McCluggage G. *Dataset for histological reporting of endometrial cancer* London: Royal College of Pathologists; 2017.
27. Fotopoulou C, Hall M, Cruickshank D, Gabra H, Ganesan R, Hughes C, et al. British Gynaecological Cancer Society (BGCS) epithelial ovarian/fallopian tube/primary peritoneal cancer guidelines: recommendations for practice. *Eur Addict Res* 2017;**213**:123-39.
28. Hirschowitz LA, Alex; Brown, Laurence; Das, Nagindra; Ganesan Raji; Herrington, Simon; Lopes, Tito; McCluggage, Glen; Meechan, David; Singh, Naveena; Smith, John; Thackeray, Alexandra; Wild, Sue. *Histopathology Reporting in Cervical Screening – an Integrated Approach*. Sheffield; 2012.
29. Oncology DfHaN. *National Head and Neck Cancer Audit 2014 Report 7*. Leeds: NHS Digital; 2014. Available from: https://digital.nhs.uk/data-and-information/publications/statistical/national-head-and-neck-cancer-audit/national-head-and-neck-cancer-audit-2014-dahno-tenth-annual-report#related-links
30. Oxley JV, Murali; Berney, Dan. *Dataset for histopathology reports for prostatic carcinoma*. London; 2016.
31. Alliance LC. *LCA Brain/CNS Cancer Clincial Guidelines*. London; 2014. Available from: http://www.londoncanceralliance.nhs.uk/media/74514/LCA_BCNS_Guidelines_FINAL.pdf
32. UK CR. *Cancer Research UK Policy Statement: Patient Access to Molecular Diagnostics and Targeted Medicines in England*. London; 2018. Available from: https://www.cancerresearchuk.org/sites/default/files/access_to_molecular_diagnostic_tests_and_targeted_medicines_in_england_0.pdf
33. Wright SJ, Newman WG, Payne K. Accounting for Capacity Constraints in Economic Evaluations of Precision Medicine: A Systematic Review. *Pharmacoeconomics* 2019;**37**:1011-27.
34. Bainbridge S CR, Meredith M, Furness P, Gordon B. *Testing Times to Come? An Evaluation of Pathology Capacity Across the UK*: Cancer Research UK; 2016. Available from: https://www.cancerresearchuk.org/sites/default/files/testing_times_to_come_nov_16_cruk.pdf
35. Brada M, Stenning S, Gabe R, Thompson LC, Levy D, Rampling R, et al. Temozolomide versus procarbazine, lomustine, and vincristine in recurrent high-grade glioma. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2010;**28**:4601-8.
36. Batchelor TT, Mulholland P, Neyns B, Nabors LB, Campone M, Wick A, et al. Phase III randomized trial comparing the efficacy of cediranib as monotherapy, and in combination with lomustine, versus lomustine alone in patients with recurrent glioblastoma. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2013;**31**:3212-8.

37. Middleton MR, Grob JJ, Aaronson N, Fierlbeck G, Tilgen W, Seiter S, et al. Randomized phase III study of temozolomide versus dacarbazine in the treatment of patients with advanced metastatic malignant melanoma. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2000;**18**:158-66.
38. 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**:377-84.
39. Thall PF, Wathen JK, Bekele BN, Champlin RE, Baker LH, Benjamin RS. Hierarchical Bayesian approaches to phase II trials in diseases with multiple subtypes. *Stat Med* 2003;**22**:763-80.
40. Neuenschwander B, Wandel S, Roychoudhury S, Bailey S. Robust exchangeability designs for early phase clinical trials with multiple strata. *Pharmaceutical Statistics* 2016;**15**:123-34.
41. Freidlin B, Korn EL. Borrowing information across subgroups in phase II trials: is it useful? *Journal of Clinical Cancer Research* 2013;**19**:1326-34.
42. Hyman DM, Puzanov I, Subbiah V, Faris JE, Chau I, Blay J-Y, et al. Vemurafenib in multiple nonmelanoma cancers with BRAF V600 mutations. 2015;**373**:726-36.
43. Heinrich MC, Joensuu H, Demetri GD, Corless CL, Apperley J, Fletcher JA, et al. Phase II, open-label study evaluating the activity of imatinib in treating life-threatening malignancies known to be associated with imatinib-sensitive tyrosine kinases. 2008;**14**:2717-25.
44. Chugh R, Wathen JK, Maki RG, Benjamin RS, Patel SR, Myers PA, et al. Phase II Multicenter Trial of Imatinib in 10 Histologic Subtypes of Sarcoma Using a Bayesian Hierarchical Statistical Model. *J Clin Oncol* 2009;**27**:3148-53. Available from: https://ascopubs.org/doi/abs/10.1200/JCO.2008.20.5054
45. Cunanan KM, Iasonos A, Shen R, Gonen M. Variance prior specification for a basket trial design using Bayesian hierarchical modeling. *Clin Trials* 2019;**16**:142-53.
46. Lunn D, Jackson C, Best N, Thomas A, Spiegelhalter D. *The BUGS book*. Boca Raton, FL: CRC Press; 2013.
47. R Core Team. *R: A language and environment for statistical computing (https://www.R-project.org/)*. [program] Vienna, Austria: R Foundation for Statistical Computing; 2019.
48. Sturtz S, Ligges U, Gelman A. R2WinBUGS: A package for running WinBUGS from R. *Journal of Statistical Software* 2005;**12**:1-16 (http://www.jstatsoft.org/v2/i03). Available from: http://www.jstatsoft.org/v12/i03
49. National Institute for Health and Care Excellence. *Guide to the methods of technology appraisal 2013*. London: NICE; 2013. Available from: https://www.nice.org.uk/process/pmg9/
50. National Instiute for Health and Care Excellence. *Eribulin for treating locally advanced or metastatic breast cancer after 1 chemotherapy regimen*. London: NICE; 2018. Available from: https://www.nice.org.uk/guidance/ta515
51. National Instiute for Health and Care Excellence. *Atezolizumab for treating locally advanced or metastatic non-small-cell lung cancer after chemotherapy*. London: NICE; 2018.
52. National Instiute for Health and Care Excellence. *Paclitaxel as albumin-bound nanoparticles with gemcitabine for untreated metastatic pancreatic cancer*. London: NICE; 2017. Available from: https://www.nice.org.uk/guidance/ta476
53. Latimer NR. Survival analysis for economic evaluations alongside clinical trials--extrapolation with patient-level data: inconsistencies, limitations, and a practical guide. *Med Decis Making* 2013;**33**:743-54. Available from: https://www.ncbi.nlm.nih.gov/pubmed/23341049
54. Center for Drug Evaluation and Research. *NDA Multidisciplinary Review and Evaluation NDA 210861 and NDA 211710 VITRAKVI (larotrectinib)*: U.S. Food and Drug Administration, ; 2016. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2018/210861Orig1s000_211710Orig1s000MultidisciplineR.pdf
55. National Institute for Health and Care Excellence. *Lutetium (177Lu) oxodotreotide for treating unresectable or metastatic neuroendocrine tumours*. London: NICE; 2018. Available from: https://www.nice.org.uk/guidance/ta539
56. Excellence NIfHaC. *Eribulin for treating locally advanced or metastatic breast cancer after 1 chemotherapy regimen*; 2018.

57. National Institute for Health and Care Excellence. *Denosumab for the prevention of skeletal-related events in adults with bone metastases from solid tumours. Technology appraisal guidance.* London: NICE; 2012.
58. Phillippo DM, A. E. Ades, S. Dias, S. Palmer, K. R. Abrams and N. J. Welton. . *NICE DSU Technical Support Document 18: Methods for population-adjusted indirect comparisons in submissions to NICE*. Sheffield, UK, ScHARR, University of Sheffield.; 2016. Available from: https://scharr.dept.shef.ac.uk/nicedsu/wp-content/uploads/sites/7/2017/05/Population-adjustment-TSD-FINAL.pdf
59. National Institute for Health and Care Excellence. *Trifluridine–tipiracil for previously treated metastatic colorectal cancer*. London: NICE; 2016. Available from: https://www.nice.org.uk/guidance/ta405
60. National Institute for Health and Care Excellence. *Lenvatinib and sorafenib for treating differentiated thyroid cancer after radioactive iodine*. London: NICE; 2018. Available from: https://www.nice.org.uk/guidance/ta535
61. National Institute for Health and Care Excellence. *Pembrolizumab for treating PDL1-positive non-small-cell lung cancer after chemotherapy*. London: NICE; 2017. Available from: https://www.nice.org.uk/guidance/ta428
62. National Institute for Health and Care Excellence. *Olaratumab in combination with doxorubicin for treating advanced soft tissue sarcoma*. London: NICE; 2017. Available from: https://www.nice.org.uk/guidance/ta465
63. Curtis L, Burns A. *Unit Costs of Health & Social Care 2018*. Canterbury, Kent: Personal Social Services Research Unit, University of Kent; 2018. Available from: https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2018/
64. Hodgson R BM, Morgan P, Mebrahtu T, Harden M, Woolacott N. *Crizotinib for untreated anaplastic lymphoma kinase-positive advance non-small-cell lung cancer: A Single Technology Appraisal*: CRD and CHE, University of York; 2016.
65. Council SSA. *Informing the Future of Genomic Medicine in Scotland*; 2019.
66. Department of Health. *Drugs and pharmacuetical electronic market information (eMit)*. 2011. Available from: https://www.gov.uk/government/publications/drugs-and-pharmaceutical-electronic-market-information-emit [accessed 4th July 2019
67. National Institute for Health and Care Excellence. *Lung cancer: diagnosis and management. draft for consultation, October 2018*. London: NICE; 2018. Available from: https://www.nice.org.uk/guidance/ng122/documents/short-version-of-draft-guideline
68. Excellence NIfHaC. *Atezolizumab for treating locally advanced or metastatic non-small-cell lung cancer after chemotherapy*; 2018.
69. Georghiou T, Bardsley M. *Exploring the cost of end of life care*. London: Nuffield Trust; 2014.
70. Ltd. FH-LR. *Data on file*. In; 2019.
71. Hsiao SJ, Zehir A, Sireci AN, Aisner DL. Detection of tumor NTRK gene fusions to identify patients who may benefit from TRK inhibitor therapy. *J Mol Diagn* 2019. Available from: https://www.ncbi.nlm.nih.gov/pubmed/31075511
72. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med* 2017;**23**:703-13. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28481359 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5461196/
73. Sarah Caul JB. *Cancer Registration Statistics, England: 2017*. In: Statistics OfN, editor.; 2019.
74. Nicky Coombes LE-B, Sam Johnson, Jane Lyons, Clara Mackay, Helen Morement, Claudia Oehler and Mick Peake. *Rare and less common cancers: Incidence and Mortality in England, 2010 to 2013*. In: Network NCI, editor. London, UK: Public Health England; 2015.
75. Tracey Genus CB, Kwok Wong, Rajaventhan Srirajaskanthan, Brian Rous, Denis Talbot, Juan Valle, Mohid Khan, Neil Pearce, Mona Elshafie, Nicholas Reed, Tu Vinh Luong, Alia Munir, John Ramage. *Incidence and prevalence of neuroendocrine tumours in England*. In: UKI NETS. London, UK: Endocrine Abstracts; 2017.
76. Cancer Research UK. *Soft tissue sarcoma incidence statistics*. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/soft-tissue-sarcoma/incidence#heading-Four [accessed 27th February 2019].

77. Care Quality Improvement Department RCoP. *National Lung Cancer Audit annual report 2018 (for the audit period 2017)*. In: Physicians RCo, editor. London, UK: Healthcare Quality Improvement Partnership (HQIP) 2019.
78. UK CR. *Uterine cancer incidence by stage at diagnosis* Cancer Research UK; Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/uterine-cancer/incidence#heading-Three [accessed Jun 2019].
79. UK CR. *Pancreatic cancer incidence by stage at diagnosis* Cancer Research UK; Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/pancreatic-cancer/incidence#heading-Three [accessed Jun 2019].
80. UK CR. *Ovarian cancer incidence by stage at diagnosis* Cancer Research UK; Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer/incidence [accessed Jun 2019].
81. UK CR. *Cervical cancer incidence by stage at diagnosis* Cancer Research UK; Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/cervical-cancer/incidence#heading-Three [accessed Jun 2019].
82. UK CR. *Bowel cancer incidence by stage at diagnosis* Cancer Research UK; Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/soft-tissue-sarcoma/incidence#heading-Zero [accessed Jun 2019].
83. Cancer Research UK. *Breast cancer incidence (invasive) statistics*. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/incidence-invasive [accessed 15th May, 2019
84. Board PATE. *PDQ Gastrointestinal Stromal Tumors Treatment*. National Cancer Institute; 2018. Available from: https://www.cancer.gov/types/soft-tissue-sarcoma/hp/gist-treatment-pdq. [accessed June 2019].
85. Austin M, Xing Y, Hayes-Jordan A, Lally K, Cormier JJJops. Melanoma incidence rises for children and adolescents: an epidemiologic review of pediatric melanoma in the United States. 2013;**48**:2207. Available from: https://www.sciencedirect.com/science/article/pii/S0022346813005113?via%3Dihub
86. Berner AM, Morrissey GJ, Murugaesu NJCGMR. Clinical Analysis of Whole Genome Sequencing in Cancer Patients. 2019:1-8.
87. Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 1998;**7**:434-55.
88. Gelman A, Rubin DB. Inferences from iterative simulation using multiple sequences. *Statistical Science* 1992;**7**:457-72.

10 Appendices

Appendix A: ERG estimates of eligible population

As the population is defined in the CS as people with *NTRK* fusion-positive [redacted], the following formula was used to estimate the eligible population for each solid tumour type, x:

$$Annual\ eligible\ population = \sum_x FNTRK_x * \ I_x * \ s_x * \ p_x$$

Where FNTRK is the frequency of *NTRK* fusions in a specific tumour type; I is the annual incidence of the tumour type in England; s is the % of that specific tumour type with stage III/IV cancer at diagnosis and p is the position in the treatment pathway. This is done for all of those solid tumours in which an NTRK fusion has been found in the literature. [71]

The frequency of *NTRK* fusions in each specific tumour type are taken from the Foundation Medicine Inc. (FMI) dataset provided to the ERG as a response to clarification questions. This dataset was chosen, as it is a comprehensive set of over [redacted] tumour samples. The *NTRK* gene fusion frequencies for sinonasal adenocarcinoma and renal cell carcinoma were not available in the dataset so it was assumed the frequency was equivalent to that seen in head and neck squamous cell carcinoma and kidney cancer, respectively.

As the FMI data set does not provide sufficient granularity on the frequency of *NTRK* gene fusions within tumours types included in the efficacy evaluable analysis set, further estimates of the frequency of *NTRK* fusions were required. Estimates for MASC, secretory breast carcinoma, papillary thyroid cancer, gastro-oesophageal junction adenocarcinoma, congential mesoblastic nephroma and infantile fibrosarcoma were obtained from the Larotrectinib NDA Multidisciplinary review and evaluation document submitted to the FDA. [54] Estimates of *NTRK* gene fusions for paediatric melanoma and paediatric high and low grade glioma were obtained from Okamura et al. [8]

[72737171](56)The annual incidence for each cancer were primarily obtained from the Office for National Statistics Cancer Registration Statistics [73] and the Rare and Less Common Cancer Statistics. [74757373](58)(48) Annual Incidence data for neuroendocrine tumours, NSCLC and soft-tissue sarcomas were obtained from other sources. [75-77] Stage at diagnosis data were obtained from Cancer Research UK. [78-83] If data were not available for specific subtypes then estimates were obtained from a pragmatic literature search. [75, 77, 84, 85] For tumour types in which a known proportion of the patient population had an unidentified stage at diagnosis, the unidentified proportion was assumed to follow the same distribution as the known proportion.

Finally, to reflect the influence the TRK-inhibitor's proposed/estimated position in the systemic therapy pathway will have on the eligible population, the position was specified for each tumour type and was used to adjust the eligible population. Entrectinib's position in the treatment pathway is proposed for people [redacted]. For those tumour types represented in the efficacy evaluable analysis and with a clear position outlined, the position was assumed to be the same as the one provided in the CS. For those tumour types represented in the efficacy evaluable analysis set but without a clear position, i.e. those with comparator data from multiple lines of therapy, it was conservatively assumed the positioning of the drug was the earliest possible position out of the options provided in the CS.

For those tumours types not represented in the entrectinib CS, it is assumed that entrectinib is positioned as a 3rd line systemic therapy. This decision was made following advice from the ERG's CA and from the identification of the position of chemotherapy, hormone therapy or best supportive care in NICE pathways.

Based on the company's assumption of patients fit enough for treatment, it was assumed for those patients in which entrectinib was first-line, 90% of patients would be eligible. For those using entrectinib as a second-line and third-line therapy, it was assumed 60% and 30% of patients respectively would be eligible.

The annual population eligible for TRK-inhibitors based on tumour types in which an *NTRK* gene fusion has been identified in the literature is 196 patients per year.

Table 58 Summary of data used to estimate annual eligible population

Tumour Type (Low Level)	**Frequency of NTRK fusion**	**Cancer Incidence (England)**	**% with Stage III/IV Cancer**	**Position of Entrectinib in line of systemic therapy**	**Annual TRK-inhibitor eligible population**
MASC	100.00%	11	22%	1	2
NSCLC (Adenocarcinoma & squamous cell carcinoma)	[redacted]	32576	57%	2	28
Breast cancer	[redacted]	46102	15%	1	27
Secretory breast carcinoma	91.70%	7	9%	2	0

Papillary thyroid tumour	13.30%	1057	31%	2	26
Thyroid tumour	[redacted]	3254	31%	2	8
Colon/colorectal	[redacted]	34825	55%	2	29
Melanoma	[redacted]	13740	10%	3	1
Neuroendocrine	[redacted]	4363	53%	2	5
Gastrointestinal stromal tumour	[redacted]	734	40%	1	2
Cholangiocarcinoma	[redacted]	556	60%	1	1
Pancreatic	[redacted]	8388	78%	1	12
Appendix	[redacted]	540	74%	3	0
Uterine	[redacted]	7862	18%	1	2
Ovarian	[redacted]	2724	55%	1	4
Cervix	[redacted]	2591	24%	1	1
Soft tissue sarcoma	[redacted]	2740	32%	1	14
Head and neck squamous cell carcinoma	0.24%	9946	63%	3	5
Salivary gland (non MASC)	2.69%	517	63%	3	3
Sinonasal adenocarcinoma	0.24%	4	63%	3	0
Gastro-esophageal junction	0.10%	7569	73%	3	2
Prostate cancer	0.23%	41201	43%	3	12
Renal cell carcinoma	0.07%	7438	43%	3	1
Low-grade glioma	0.42%	929	0%	3	0
High grade glioma (inc. glioblastoma multiforme)	0.42%	2781	100%	3	4
Paediatric high grade glioma	5.30%	67	100%	3	1
Congenital mesoblastic nephroma	60.70%	0	0%	3	0
Paediatric melanoma	11.11%	56	34%	3	1
Infantile fibrosarcoma	90.90%	59	51%	3	8
Paediatric low grade glioma	2.50%	723	0%	3	0
Total:					**196**

*The frequency NTRK fusions in appendix tumours in the FMI data set was reported to be 0%, however it has been reported to be higher than 0% in the literature [2]

Note, totals in the annual TRK-inhibitor eligible population column may add up to greater less than 196 due to rounding.

MASC, mammary analogue secretory carcinoma; NSCLC non-small cell lung cancer

Appendix B: Numbers needed to screen

The number of patients who require screening to identify one individual with an *NTRK* fusion varies depending on the prevalence of gene rearrangement. Table 59 presents the number of patients who need to be screened to identify one individual with a *NTRK* fusion. This is calculated by the following equation:

$$\text{NNS: } \frac{1}{\textit{NTRK} \text{ fusion rate}}$$

According to the company submission, the NNS to identify one patient with an *NTRK* rearrangement varies between 1 (MASC, *NTRK* prevalence = 100%) and 1250 (Pancreatic Cancer, *NTRK* prevalence = 0.08%). ERG estimates of the NNS to identify one *NTRK*+ patient for all tumour types mentioned in the CS range from 1 (MASC, Secretory Breast Carcinoma and Infantile Fibrosarcoma) to 2000 (High Grade Glioma, *NTRK* prevalence = 0.05%). Due to discrepancies in the recorded prevalence of NTRK fusions, the NNS reported in the CS differ to the NNS calculated by the ERG.

Table 59 Number needed to screen: company and ERG estimates

Tumour Type	Prevalence of *NTRK* fusion (ERG)	Number Needed to Screen (ERG)	Number Needed to Screen (Company)
Salivary gland (MASC)	[redacted]	[redacted]	[redacted]
NSCLC	[redacted]	[redacted]	[redacted]
Breast cancer (not specified)	[redacted]	[redacted]	[redacted]
Secretory breast carcinoma	[redacted]	[redacted]	[redacted]
Papillary thyroid tumour	[redacted]	[redacted]	[redacted]
Thyroid Tumour (NOS)	[redacted]	[redacted]	[redacted]
Colon/colorectal	[redacted]	[redacted]	[redacted]
Neuroendocrine (NOS)	[redacted]	[redacted]	[redacted]
Cholangiocarcinoma	[redacted]	[redacted]	[redacted]
Pancreatic	[redacted]	[redacted]	[redacted]
Uterine	[redacted]	[redacted]	[redacted]
Ovarian	[redacted]	[redacted]	[redacted]
Cervix	[redacted]	[redacted]	[redacted]
Soft tissue sarcoma	[redacted]	[redacted]	[redacted]
High grade glioma	[redacted]	[redacted]	[redacted]
Paediatric high grade glioma	[redacted]	[redacted]	[redacted]
Congenital mesoblastic nephroma	[redacted]	[redacted]	[redacted]
Paediatric melanoma	[redacted]	[redacted]	[redacted]
Infantile fibrosarcoma	[redacted]	[redacted]	[redacted]

Paediatric low grade glioma	[redacted]	[redacted]	[redacted]

Using the calculated estimates of the annual population eligible for a TRK inihibitor, the ERG estimated the total patient population that would require IHC and NGS screening to identify individuals eligible for entrectinib. This is equal to the annual population of individuals in England with [redacted].

$$\text{Population Requiring IHC screening}_x = I_x \times s_x \times p_x$$

where *x* is the tumour type in which an *NTRK* fusion has been identified; I_x is the annual incidence of the tumour type in England; s_x is the % of patients with that specific tumour type who have stage III/IV cancer at diagnosis and p_x is the position of the therapy in the treatment pathway.

ERG calculation of the annual population who require IHC screening, based on the tumours where an *NTRK* fusion has been identified, indicate that approximately 51,958 patients would need IHC screening to identify potential individuals eligible for entrectinib.

According to the CS, IHC will identify 89% of *NTRK* fusion negative individuals, hence 11% of the population screened with IHC will require confirmatory NGS screening. Whole genome sequencing also requires confirmatory RNA-based NGS; as the diagnostic accuracy of WGS is unclear, it was assumed that 11% of individuals screened with WGS would require confirmatory RNA-based NGS. [86]

$$\text{Population Requiring NGS screening}_x = (I_x \times s_x \times p_x) \times 0.11$$

The ERG calculated annual population requiring NGS screening, based on the tumours where an *NTRK* fusion has been identified, indicates that approximately 5,806 patients would need confirmatory NGS to identify the patients eligible for entrectinib.

Table 60**: Annual population requiring IHC or NGS screening in order to identify patients with an NTRK fusion.**

Tumour Type (Low Level)	Annual TRK-inhibitor eligible population	Population requiring IHC screening	Population Requiring NGS Screening
Salivary gland (MASC)	2	-	-
NSCLC Lung (Adenocarcinoma & squamous cell carcinoma)	9	11141	1226
Breast cancer (not specified)	4	6224	685
Secretory breast carcinoma	1	0	0

Papillary thyroid tumour	24	197	22
Thyroid tumour (NOS)	5	605	67
Colon/colorectal	12	11492	1264
Melanoma (NOS)	2	412	45
Neuroendocrine (NOS)	4	1387	153
Gastrointestinal stromal tumour	3	264	29
Cholangiocarcinoma	0	300	33
Pancreatic	15	5888	648
Appendix	9	120	13
Uterine	1	1274	140
Ovarian	3	1348	148
Cervix	2	560	62
Soft tissue sarcoma	4	-	87
Head and neck squamous cell carcinoma (NOS)	21	1880	207
Salivary gland (non MASC)	3	98	11
Sinonasal adenocarcinoma	0	1	0
Gastro-oesophageal junction	5	1658	182
Prostate cancer	24	5315	585
Renal cell carcinoma	4	960	106
High grade glioma (inc. glioblastoma multiforme)	1	834	92
Paediatric high grade glioma	3	-	2
Congenital mesoblastic nephroma	0	-	0
Paediatric melanoma	0	-	1
Infantile fibrosarcoma	25	-	1
Paediatric low grade glioma	0	-	0
Total	**188**	**51958**	**5806**

Appendix C: Comparator evidence

See Excel file.

Appendix D: STARTRK-2 Quality assessment checklist

Checklist		**STARTRK-2 (Interim CSR)**
Reporting	Is the hypothesis/aim/objective of the study clearly described?	yes
	Are the main outcomes to be measured clearly described in the Introduction or methods section?	yes
	Are the characteristics of the patients included in the study clearly described?	yes
	Are the interventions of interest clearly described?	yes
	Are the distributions of principal confounders in each group of subjects to be compared clearly described?	no
	Are the main findings of the study clearly described?	yes
	Does the study provide estimates of the random variability in the data for the main outcomes?	yes
	Have all important adverse events that may be a consequence of the intervention been reported?	yes
	Have the characteristics of patients lost to follow-up been described?	yes
	Have actual probability values been reported for the main outcomes except when the probability is less than 0.001?	no
External validity	Were the subjects asked to participate in the study representative of the entire population from which they were recruited?	no
	Were those subjects who were prepared to participate representative of the entire population from which they were recruited?	no
	Were the staff, places, and facilities where the patients were treated representative of the treatment the majority of patients receive?	yes
Internal validity - bias	Was an attempt made to blind study subjects to the intervention they have received?	no
	Was an attempt made to blind those measuring the main outcomes of the intervention?	no
	If any of the results of the study were based on "data dredging", was this made clear?	yes
	In trials and cohort studies, do the analyses adjust for different lengths of follow-up of patients, or in case-control studies, is the time period between the intervention and outcome the same for cases and controls?	yes
	Were the statistical tests used to assess the main outcomes appropriate?	yes
	Was compliance with the interventions reliable?	yes
	Were the main outcome measures used accurate?	yes
	Were the patients in different intervention groups recruited from the same population?	no

Internal validity - confounding (selection bias)	Were study subjects in different intervention groups recruited over the same period of time?	yes
	Were study subjects randomised to intervention groups?	no
	Was the randomised intervention assignment concealed from both patients and health care staff until recruitment was complete and irrevocable?	no
	Was there adequate adjustment for confounding in the analyses from which the main findings were drawn?	no
	Were losses of patients to follow-up taken into account?	yes
Power	Did the study have sufficient power to detect a clinically important effect where the probability value for a difference being due to chance is less than 5%?	no

Appendix E: Analysis of response heterogeneity - methods and additional results

For all analyses 55,000 iterations were run on 2 parallel chains and the first 5,000 iterations discarded as "burn-in". Convergence was assessed by visual inspection of the Brooks-Gelman-Rubin plots and assessment of the $\hat{R}$ statistic.[87, 88]

Table 61 shows the posterior probabilities estimated by the base-case BHM, using BIRC-assessed data with imputation and the prior distribution is equation (1).

Table 61 Posterior probabilities of response for all tumour types (BIRC-assessed data with imputation)

	Tumour type	mean	median	95%CrI
1	Sarcoma	[redacted]	[redacted]	[redacted]
2	NSCLC	[redacted]	[redacted]	[redacted]
3	CRC	[redacted]	[redacted]	[redacted]
4	Neuroendocrine tumours	[redacted]	[redacted]	[redacted]
5	Pancreatic	[redacted]	[redacted]	[redacted]
6	Gynaecological	[redacted]	[redacted]	[redacted]
7	Cholangiocarcinoma	[redacted]	[redacted]	[redacted]
8	MASC	[redacted]	[redacted]	[redacted]
9	Breast	[redacted]	[redacted]	[redacted]
10	Thyroid	[redacted]	[redacted]	[redacted]
11	CNS Primary	[redacted]	[redacted]	[redacted]
12	Paediatric CNS Primary	[redacted]	[redacted]	[redacted]
13	Paediatric (non-CNS)	[redacted]	[redacted]	[redacted]

Prior distribution for log-odds of response centred on a probability of 0.3;
Uniform prior distribution for the between-tumour standard deviation

Table 62 has the model fit statistics for the base-case and all sensitivity analyses, which show that all models fit the data well. Inspection of box-plots of individual groups' contributions to the residual deviance (not shown) support this.

Table 62 Model fit statistics for the base-case and sensitivity analyses.

	Posterior mean of the residual deviance	DIC
Base-case (all patients, prior for response rate centred on 0.3, Uniform prior for heterogeneity)	11.8*	40.1
Sensitivity analysis 1 (all patients, prior for response rate centred on 0.3, inverse-gamma prior for heterogeneity)	10.9*	44.3
Sensitivity analysis 2 (all patients, prior for response rate centred on 0.5, Uniform prior for heterogeneity)	11.9*	40.0
Sensitivity analysis 3 (no primary CNS or paediatric patients, prior for response rate centred on 0.3, Uniform prior for heterogeneity)	9.1†	31.3

* compare to 13 groups; † compare to 10 groups

Sensitivity analysis 1

An Inverse Gamma(2, 20) prior distribution for the between-tumour variance was used, instead of the Uniform prior on the between-tumour standard deviation. This means the between-tumour precision has prior mean 0.10 and variance 0.005, which implies that the between-tumour standard deviation has a prior mean ≈ 3.97 and variance ≈ 4.33. Note that both the prior mean and variance are higher than those implied by the Uniform(0,5) prior distribution, which are 2.5 and 2.08, respectively. This results in a much larger estimate of the between-tumour heterogeneity with median [redacted] and 95% CrI ([redacted]). Figure 22 shows the prior and posterior distributions for the between-tumour heterogeneity. The prior and posterior distributions in the base-case are also included for comparison. We can see that the inverse-gamma prior distribution places much more weight on large values of heterogeneity and does not allow for values close to zero, which is then reflected in the posterior distribution, which also excludes small values.

Figure 22 Sensitivity analysis 1: [redacted]

[redacted]*Figure redacted*

The very large estimated heterogeneity means that the 95% CrI for the probability of response is much wider than in the base-case, and the predictive interval for the response rate for an unrepresented tumour type covers nearly the whole range of probabilities from zero to 1 (Table 63).

Table 63 Sensitivity analysis 1: Probabilities of response according to the BHM.

	Overall posterior probability of response		
	mean	median	95% CrI

Posterior probability of response	[redacted]	[redacted]	[redacted]
Predictive probability of response	[redacted]	[redacted]	[redacted]

Prior distribution for log-odds of response centred on a probability of 0.3;
Inverse-gamma prior distribution for the between-tumour variance.

Given that the inverse-gamma prior is not derived from genuine prior beliefs that low levels of heterogeneity are not plausible, the ERG caution against the results from this analysis. However, for completeness, the distribution of the response rates for each tumour type are shown in Table 64 and Figure 23.

Table 64 Sensitivity analysis 1: Probabilities of response for all tumour types (IRC-assessed data with imputation)

	Tumour type	observed response (%)	Estimated mean response based on BHM (%)	Prob of response rate at least 30%	Prob of response rate at least 10%
1	Sarcoma	[redacted]	[redacted]	[redacted]	[redacted]
2	NSCLC	[redacted]	[redacted]	[redacted]	[redacted]
3	CRC	[redacted]	[redacted]	[redacted]	[redacted]
4	Neuroendocrine tumours	[redacted]	[redacted]	[redacted]	[redacted]
5	Pancreatic	[redacted]	[redacted]	[redacted]	[redacted]
6	Gynaecological	[redacted]	[redacted]	[redacted]	[redacted]
7	Cholangiocarcinoma	[redacted]	[redacted]	[redacted]	[redacted]
8	MASC	[redacted]	[redacted]	[redacted]	[redacted]
9	Breast	[redacted]	[redacted]	[redacted]	[redacted]
10	Thyroid	[redacted]	[redacted]	[redacted]	[redacted]
11	CNS Primary	[redacted]	[redacted]	[redacted]	[redacted]
12	Paediatric CNS Primary	[redacted]	[redacted]	[redacted]	[redacted]
13	Paediatric (non-CNS)	[redacted]	[redacted]	[redacted]	[redacted]

Prior distribution for log-odds of response centred on a probability of 0.3; Inverse-gamma prior distribution for the between-tumour variance

Figure 23 Sensitivity analysis 1: [redacted]

Figure redacted

Sensitivity analysis 2

Results using a more favourable prior distribution for the log-odds of response, centred on an *a priori* probability of response of 50%, are presented in this section. The prior distributions used in this sensitivity analysis are

$$\mu \sim \text{Normal}(0,10)$$
$$\sigma \sim \text{Uniform}(0,5)$$

The BHM estimates moderate between-group heterogeneity, similar to the base-case (posterior median [redacted] 95% CrI ([redacted])).

The estimated mean response rate across all tumour types and the predictive probabilities are similar to the base-case (Table 65) and the estimated probabilities of response for each tumour type in Table 66 are almost identical to the results obtained in the base-case (Table 61). We therefore conclude that the prior distribution for the mean probability of response does not have any meaningful impact on the results.

Table 65 Sensitivity analysis 2: Probabilities of response according to the BHM.

	Overall posterior probability of response		
	mean	median	95% CrI
Posterior probability of response	[redacted]	[redacted]	[redacted]
Predictive probability of response	[redacted]	[redacted]	[redacted]

Prior distribution for log-odds of response centred on a probability of 0.5;
Uniform prior distribution for the between-tumour standard deviation

Table 66 Sensitivity analysis 2: Posterior probabilities of response for all tumour types

	Tumour type	**mean**	**median**	**95%CrI**
1	Sarcoma	[redacted]	[redacted]	[redacted]
2	NSCLC	[redacted]	[redacted]	[redacted]
3	CRC	[redacted]	[redacted]	[redacted]
4	Neuroendocrine tumours	[redacted]	[redacted]	[redacted]
5	Pancreatic	[redacted]	[redacted]	[redacted]
6	Gynaecological	[redacted]	[redacted]	[redacted]
7	Cholangiocarcinoma	[redacted]	[redacted]	[redacted]
8	MASC	[redacted]	[redacted]	[redacted]
9	Breast	[redacted]	[redacted]	[redacted]
10	Thyroid	[redacted]	[redacted]	[redacted]
11	CNS Primary	[redacted]	[redacted]	[redacted]
12	Paediatric CNS Primary	[redacted]	[redacted]	[redacted]
13	Paediatric (non-CNS)	[redacted]	[redacted]	[redacted]

Prior distribution for log-odds of response centred on a probability of 0.3;
Uniform prior distribution for the between-tumour standard deviation

Sensitivity analysis 3

Results excluding primary CNS and paediatric patients are presented in this section. Note that this analysis includes only BICR-assessed data. The prior distributions used in this sensitivity analysis are given in equation (1).

The BHM estimates moderate between-group heterogeneity, similar to the base-case although with a wider 95%CrI due to less tumour types being included (posterior median [redacted] 95% CrI ([redacted])).

The estimated mean response rate across all tumour types and the predictive probabilities are similar to the base-case (Table 67) and the estimated probabilities of response for each tumour type in Table 68 are only slightly larger than the results obtained in the base-case (Table 61, including all tumour types). We therefore conclude that there are similar amounts heterogeneity across adult non-primary CNS tumours as across all tumour types, and a similar amount of uncertainty in the response rate expected in an unrepresented adult solid tumour.

Table 67 Sensitivity analysis 3: Probabilities of response according to the BHM.

	Overall posterior probability of response		
	mean	median	95% CrI
Posterior probability of response	[redacted]	[redacted]	[redacted]
Predictive probability of response	[redacted]	[redacted]	[redacted]

Prior distribution for log-odds of response centred on a probability of 0.5;
Uniform prior distribution for the between-tumour standard deviation

Table 68 Sensitivity analysis 3: Posterior probabilities of response for all tumour types

	Tumour type	**mean**	**median**	**95%CrI**
1	Sarcoma	[redacted]	[redacted]	[redacted]
2	NSCLC	[redacted]	[redacted]	[redacted]
3	CRC	[redacted]	[redacted]	[redacted]
4	Neuroendocrine tumours	[redacted]	[redacted]	[redacted]
5	Pancreatic	[redacted]	[redacted]	[redacted]
6	Gynaecological	[redacted]	[redacted]	[redacted]
7	Cholangiocarcinoma	[redacted]	[redacted]	[redacted]
8	MASC	[redacted]	[redacted]	[redacted]
9	Breast	[redacted]	[redacted]	[redacted]
10	Thyroid	[redacted]	[redacted]	[redacted]

Prior distribution for log-odds of response centred on a probability of 0.3;
Uniform prior distribution for the between-tumour standard deviation

OpenBUGS code

Bayesian Hierarchical Model: Uniform(0,5) prior distribution for the between-tumour standard deviation

```
# CODE ADAPTED FROM: Thall et al (2003)
# Hierarchical Bayesian approaches to phase II trials in diseases with multiple subtypes.
# Statist. Med., 22: 763-780. doi:10.1002/sim.1399
#
# Uniform prior distribution for between-group SD, as recommended by Cunanan et al. (Clinical
Trials, 2019)
#
model{
for (i in 1:numGroups){  # numGroups is k, the number of different probabilities
  x[i] ~ dbin(p[i],n[i]) # In each group, x is the number of responses and n is the number of
patients
  # set up deviance code with correction for zero cells
  x1[i] <- max(x[i],0.1) # zero cell correction
  xhat[i] <- p[i] * n[i] # expected value of the numerators
  xhat1[i] <- max(xhat[i], 0.1) # zero cell correction
  # Deviance contribution with zero cell correction
  dev1[i] <- 2 * (x1[i] * (log(x1[i])-log(xhat1[i]))
             +  (n[i]-x1[i]) * (log(n[i]-x1[i]) - log(n[i]-xhat1[i])))
  # deviance contribution for for zero cells
  dev0[i] <- 2 * n[i] * log(n[i]/(n[i]-xhat[i]))
  # deviance contribution
  dev[i] <- dev1[i] * (1-equals(x[i],0)) + dev0[i] * equals(x[i],0)
  # logit model for p
  logit(p[i]) <- rho[i]
  rho[i] ~ dnorm(mu,tau) # RE for log-odds
  # Probability that the response rate for each group is > than targetResp (given as data)
  pg[i] <- step(p[i] - targetResp)
  pg2[i] <- step(p[i] - targetResp2)
 }
totresdev <- sum(dev[])            # total residual deviance
# Priors
mu ~ dnorm(mean.Mu, perc.Mu)       # pooled mean of log-odds
#tau ~ dgamma(tau.alpha, tau.beta) # used in Thall (2003)
#sd <- 1/sqrt(tau)                 # between-group sd (log-odds scale)
sd ~ dunif(0,5)                    # recommended by Cunanan (2019)
tau <- pow(sd,-2)
# predictive distribution
rho.new ~ dnorm(mu,tau)            # log-odds response across groups
# convert to probabilities
logit(p.pooled) <- mu    # mean probability of response across groups
logit(p.new) <- rho.new  # probability response across groups
# predictive probabilities of response rates > targetResp (given as data)
pg.new <- step(p.new - targetResp)
pg2.new <- step(p.new - targetResp2)
}
```

Data

list(x=c(), n=c(), numGroups=13, mean.Mu=-0.847298, perc.Mu=0.1, targetResp=0.3, targetResp2=0.1)

Appendix F: Drummond Checklist

Table 69 presents the quality assessment.

Table 69 Quality assessment of included CEA study using Drummond et al. checklist completed by the ERG

	CEA quality assessment questions	Answer (Yes/No/Unclear)	Notes/Explanation for No or Unclear
1	Was the research question stated?	Yes	-
2	Was the economic importance of the research question stated?	Yes	-
3	Was/were the viewpoint(s) of the analysis clearly stated and justified?	Yes	-
4	Was a rationale reported for the choice of the alternative programmes or interventions compared?	Yes	-
5	Were the alternatives being compared clearly described?	No	Many of the interventions used to construct the weighted average comparator were described but no comparators were provided in the 'other TBC' resulting in a lack of clarity as to what the alternatives are.
6	Was the form of economic evaluation stated?	Yes	-
7	Was the choice of form of economic evaluation justified in relation to the questions addressed?	Yes	-
8	Was/were the source(s) of effectiveness estimates used stated?	Yes	-
9	Were details of the design and results of the effectiveness study given (if based on a single study)?	Yes	-
10	Were details of the methods of synthesis or meta-analysis of estimates given (if based on an overview of a number of effectiveness studies)?	Yes	-
11	Were the primary outcome measure(s) for the economic evaluation clearly stated?	Yes	-

12	Were the methods used to value health states and other benefits stated?	Yes	-
13	Were the details of the subjects from whom valuations were obtained given?	Yes	-
14	Were productivity changes (if included) reported separately?	No	Excluded from base-case analysis. Model did provide option to include productivity losses.
15	Was the relevance of productivity changes to the study question discussed?	No	Not mentioned
16	Were quantities of resources reported separately from their unit cost?	Yes	-
17	Were the methods for the estimation of quantities and unit costs described?	Yes	-
18	Were currency and price data recorded?	Yes	-
19	Were details of price adjustments for inflation or currency conversion given?	Unclear	End of life costs were adjusted for inflation, however no other parameters seem to have been adjusted for inflation/currency conversion.
20	Were details of any model used given?	No	-
21	Was there a justification for the choice of model used and the key parameters on which it was based?	Yes	-
22	Was the time horizon of cost and benefits stated?	Yes	-
23	Was the discount rate stated?	Yes	-
24	Was the choice of rate justified?	Yes	-
25	Was an explanation given if cost or benefits were not discounted?	No	-

26	Were the details of statistical test(s) and confidence intervals given for stochastic data?	No	CI for stochastic data were provided however, the CIs for the comparator data were all assumed and not taken from the original sources. Formal significance tests were not performed.
27	Was the approach to sensitivity analysis described?	Yes	-
28	Was the choice of variables for sensitivity analysis justified?	No	Justified for some of the variables, but testing costs were not explored thoroughly and this was a big driver of the cost-effectiveness.
29	Were the ranges over which the parameters were varied stated?	Yes	-
30	Were relevant alternatives compared? (That is, were appropriate comparisons made when conducting the incremental analysis?)	No	It is unclear what the relevant comparators are for the population considered in the marketing authorisation as not only are there unrepresented tumour types but it's unclear where entrectinib will be used in the treatment pathway.
31	Was an incremental analysis reported?	Yes	-
32	Were major outcomes presented in a disaggregated as well as aggregated form?	Yes	-
33	Was the answer to the study question given?	Yes	-
34	Did conclusions follow from the data reported?	Yes	-
35	Were conclusions accompanied by the appropriate caveats?	No	-
36	Were generalisability issues addressed?	No	Issues of generalisability were not fully addressed. Unclear whether the entrectinib population and the comparator population are generalisable due to unrepresented tumour types; proportion of tumour types used to weight comparator costs and utilities; prior therapies; underrepresentation of *NTRK2* fusions in the CS and issues of where in the treatment pathway entrectinib will be used.

wan