CONFIDENTIAL UNTIL PUBLISHED

Tucatinib with trastuzumab and capecitabine for treating HER2-positive unresectable locally advanced or metastatic breast cancer after 2 or more anti-HER2 therapies

ERRATUM

Post factual accuracy check version with corrections

Produced by	Southampton Health Technology Assessments Centre (SHTAC)
Authors	David Alexander Scott, Principal Research Fellow, SHTAC Lorna Hazell, Senior Research Assistant, SHTAC Jill Colquitt, Senior Researcher, Effective Evidence LLP Neelam Kalita, Senior Research Fellow, SHTAC Joanne Lord, Director, SHTAC Emma Loveman, Senior Researcher, Effective Evidence LLP Jonathan Shepherd, Principal Research Fellow.
Correspondence to	Jonathan Shepherd Southampton Health Technology Assessments Centre (SHTAC) Wessex Institute Alpha House Enterprise Road, University of Southampton Science Park Southampton SO16 7NS www.southampton.ac.uk/shtac
Date completed	23/06/2021

Source of Funding: This report was commissioned by the National Institute for Health Research (NIHR) Evidence Synthesis Programme as project number HTA NIHR134065

Acknowledgements

We thank Dr Peter Simmonds, Consultant Medical Oncologist, University Hospital Southampton NHS Foundation Trust for providing expert clinical advice to the project team and for reading and commenting on a draft the report.

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

Declared competing interests of the authors and advisors

The authors and clinical advisor declare no competing interests

Rider on responsibility for report

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

This report should be referenced as follows:

Scott, D; Hazell, L; Colquitt, J; Kalita, N; Lord, J; Loveman, E; Shepherd, J. Tucatinib with trastuzumab and capecitabine for treating HER2-positive unresectable locally advanced or metastatic breast cancer after 2 or more anti-HER2 therapies - A Single Technology Appraisal. Southampton Health Technology Assessments Centre, 2021.

Contributions of authors

David Scott critically appraised the clinical effectiveness systematic review and network meta-analysis, and drafted the report; Lorna Hazell critically appraised the clinical effectiveness systematic review and network meta-analysis, and drafted the report; Jill Colquitt critically appraised the clinical effectiveness systematic review and drafted the report; Neelam Kalita critically appraised the health economic systematic review, critically appraised the economic evaluation, and drafted the report; Jo Lord critically appraised the health economic systematic review, critically appraised the economic evaluation, and drafted

the report; Emma Loveman critically appraised the clinical effectiveness systematic review and drafted the report; Jonathan Shepherd critically appraised the clinical effectiveness systematic review, drafted the report and is the project co-ordinator and guarantor.

[redacted]

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AE	Adverse event
AIC	Academic in confidence
BICR	Blinded Independent Central Review
BNF	British National Formulary
CI	Confidence interval
CIC	Commercial in confidence
CRD	Centre for Reviews and Dissemination
CrI	Credible interval
CS	Company submission
CSR	Clinical study report
DSU	Decision Support Unit
EMA	European Medicines Agency
EMC	Electronic Medicines Compendium
EPAR	European Public Assessment Report
EQ-5D-3L	European Quality of Life Working Group Health Status Measure 3 Dimensions, 3 Levels
EQ-5D-5L	European Quality of Life Working Group Health Status Measure 5 Dimensions, 5 Levels
EQ-VAS	EuroQol Visual Analogue Scale
ERG	Evidence Review Group
HER2	Human epidermal growth factor receptor 2
HER2-	Human epidermal growth factor receptor 2-negative
HER2+	Human epidermal growth factor receptor 2-positive
HER2CLIMB	A Study of Tucatinib vs Placebo in Combination with Capecitabine & Trastuzumab in Patients with Advanced HER2+ Breast Cancer
HRG	Healthcare Resource Group
HRQoL	Health-related quality of life
HTA	Health technology assessment
ICER	Incremental cost-effectiveness ratio
IPD	Individual patient level data
ITT	Intent to treat
ITC	Indirect treatment comparison
LFT	Liver function tests

mITT	Modified intent to treat
MRI	Magnetic resonance imaging
NHS	National Health Service
NMA	Network meta-analysis
NICE	National Institute for Health and Care Excellence
NR	Not reported
ORR	Objective response rate
OS	Overall survival
PFS	Progression-free survival
PSA	Probabilistic sensitivity analysis
PSS	Personal Social Services
PSSRU	Personal social services research unit
QoL	Quality of life
RCT	Randomised controlled trial
RECIST	Response Evaluation Criteria in Solid Tumours
RR	Relative risk/risk ratio
SAE	Serious adverse event
SD	Standard deviation
SE	Standard error
SLR	Systematic literature review
SmPC	Summary of product characteristics
T-DM1	Trastuzumab emtansine
TA	Technology appraisal
TEAE	Treatment-emergent adverse event
TKI	Tyrosine kinase inhibitor
TSD	Technical Support Document
UK	United Kingdom
US	United States
VAS	Visual analogue scale

1 EXECUTIVE SUMMARY

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

Section 1.1 provides an overview of the key issues. Section 1.2 provides an overview of key model outcomes and the modelling assumptions that have the greatest effect on the ICER. Sections 1.3 to 1.5 explain the key issues in more detail. Background information on the condition, health technology, evidence and information on the issues are in the main ERG report.

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

1.1 Overview of the ERG's key issues

Table 1 Summary of key issues

Issue number	Summary of issue	Report sections
1	The results of the indirect comparison between tucatinib in combination with trastuzumab and capecitabine and comparator treatments are uncertain due to clinical heterogeneity across the trials included in the network meta-analysis. The ERG uses a random effects model rather than the fixed-effect model favoured by the company. We also correct a HR typographical error in one of the trials included in the NMA Pivot et al (2015).	3.3 and 3.4
2	Lack of justification for the company's survival extrapolation model (based on the fractional polynomial NMA Weibull model for a reference arm), adjusted for indirect comparators with the HR NMA. The ERG proposes use of survival curves directly fitted to the HER2CLIMB trial.	4.2.6
3	Cost-effectiveness analysis may not reflect the prevalence of brain metastases in the clinical population. We suggest an exploratory analysis with baseline survival adjusted for the proportion of patients with brain metastases.	4.2.6
4	There is a lack of justification for the use of different health state utilities for the tucatinib combination and comparators. The analysis of EQ-5D data from the HER2CLIMB trial is an appropriate source for estimation of utilities, but the analysis is poorly reported, and potentially subject to bias from missing data.	4.2.9

1.2 Overview of key model outcomes

Table 2 reports the company's revised base case results. These estimates are based on PAS discount for tucatinib and an assumed discount for trastuzumab, and other drugs at list price. The company's model results were most sensitive to relative dose intensity for the tucatinib combination, and health state utilities for progressed health.

Table 2 Company's revised base case, deterministic

Treatment	Cost [a]	QALYs	Pairwise ICERs Tuc + tras + cap vs. comparators	ICERs fully incremental Excluding Tras + cap	ICERs fully incremental Including Tras + cap
Capecitabine	[redacted]	[redacted]	[redacted]	-	-
Vinorelbine	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]	-	[redacted]
Eribulin	[redacted]	[redacted]	£37,483	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-	[redacted]	[redacted]

Source: Clarification response Table B2c (tucatinib combination and eribulin), other results produced by the ERG
[a] PAS discount for tucatinib and assumed discount for trastuzumab, other drugs at list price
Ext dom, extendedly dominated; ICER, incremental cost-effectiveness ratio; Tras + cap, trastuzumab with capecitabine; Tuc + tras + cap, tucatinib with trastuzumab and capecitabine.

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

The ERG identified no key issues relating to the decision problem in general. Aspects of the decision problem where there is uncertainty (e.g. subgroup of patients with brain metastases) are covered by key issues for clinical and cost effectiveness.

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

Issue 1 Uncertain indirect comparison results due to study heterogeneity

Report section	3.3 and 3.4
Description of issue and why the ERG has identified it as important	The results of the network meta-analysis (NMA) which provides an indirect comparison of the tucatinib combination versus the comparators relevant to the scope of the appraisal (eribulin monotherapy, capecitabine monotherapy and vinorelbine monotherapy), are uncertain. The primary cause of uncertainty is heterogeneity between studies included in the NMA in terms of the proportion of patients with brain metastases, a likely effect modifier. The HER2CLIMB trial includes patients with and without brain metastases. The comparator trials, in contrast, include

	few or no patients with brain metastases. This creates an uneven distribution of patients with brain metastases across the trials, and there is likely to be bias in the results, though the direction and magnitude of this bias is unclear. The company's choice of a fixed-effect NMA model is inappropriate given this heterogeneity.
What alternative approach has the ERG suggested?	In the context of this heterogeneity, we suggest that a random-effects NMA model is more appropriate than a fixed-effect model. The ERG revised the company's random-effect NMA to correct for a reporting error in the paper by Pivot et al. (section 3.6.1). We use the corrected NMA random-effect HRs in the ERG preferred cost-effectiveness analysis (section 6.4). The ERG also conducted an exploratory NMA scenario analysis using data for the subgroup of patients without brain metastases from the HER2CLIMB trial (3.6.3). This reduces heterogeneity between the studies included in the evidence network and produced HRs that are less favourable for the tucatinib combination in patients with brain metastases than for the whole trial population. The results of this subgroup analysis are subject to limitations, and we did not include them in additional ERG cost-effectiveness analysis.
What is the expected effect on the cost-effectiveness estimates?	The use of random-effect NMA with Pivot correction increased the ICERs for the tucatinib combination. The implications of our exploratory NMA analysis for patients without brain metastases is unclear. It suggests that heterogeneity over the proportion of patients with brain metastases is likely to affect cost-effectiveness. All else being equal, the higher HRs for patients without brain metastases would give higher ICERs. However, it is not possible to conduct a subgroup analysis for people with brain metastases, due to the lack of evidence for the indirect comparators. It is also likely that other model parameters will differ for people with/without brain metastases (see Issue 3).
What additional evidence or analyses might help to resolve this key issue?	The lack of clinical effectiveness evidence for the comparator treatments in the subgroup of patients with brain metastases is difficult to resolve without further evidence. We request that the company revise their random-effect NMA analysis to correct for the reporting error for the Pivot et al. study. The NMA outputs should be included in their economic model.

1.5 The cost-effectiveness evidence: summary of the ERG's key issues

Issue 2 Survival extrapolations

Report section	4.2.6
Description of issue and why the ERG has identified it as important	For their base case, the company obtain OS and PFS extrapolations by applying relative effects from the NMA ([redacted]) to fractional polynomial OS and PFS survival curves for a reference treatment (lapatinib plus capecitabine). The resulting modelled OS estimates are substantially more favourable than those observed in the HER2CLIMB study. This may be due to the population in this trial, which included more patients with brain metastases than other trials in the NMA (which included few or no patients with brain metastases). The company did not explore the impact of alternative functional forms for PFS and OS in their scenario analysis. Although a scenario with alternative survival models is reported (CS Table 39), QALY estimates from this scenario did not differ from those in the base case analysis (see section 6.2). This is not surprising as the model only includes one fractional polynomial function for OS and one for PFS.
What alternative approach has the ERG suggested?	The ERG has used the 'within trial' approach, which was coded in the company's model but not used, to estimate OS and PFS curves fitted to the HER2CLIMB trial data and then adjusted for indirect comparators with HRs from the NMA (section 6.3.1). We explored the impact of alternative survival models for OS and PFS.
What is the expected effect on the cost-effectiveness estimates?	The 'within-trial' analysis produced PFS and OS curves that are reflective of outcomes in the pivotal HER2CLIMB trial. This reduced survival and QALY estimates for all treatments, and also incremental differences between them. Hence ICERs were significantly higher than with the company's NMA based modelling approach. There is residual uncertainty because several models for OS had a good fit to the trial data and appeared plausible but gave a range of ICER results.
What additional evidence or analyses might help to resolve this key issue?	Clinical opinion on the plausibility of the survival extrapolations and whether the company's NMA-based survival estimates or ERG's within-trial estimates are more reflective of the population in clinical practice.

Issue 3 Subgroup analysis

Report section	4.2.3
Description of issue and why the ERG has identified it as important	The HER2CLIMB trial included patients with presence or a history of brain metastases: nearly 50% of the total study population (CS Table 6). The NICE scope requested subgroup analysis for patients with brain metastases if the evidence allowed. The company did not attempt to model cost-effectiveness for this subgroup, on the basis that there is a lack of clinical evidence in this group for the scope comparators. The ERG acknowledges this lack, but we note that the economic model nevertheless relies on estimates of relative effectiveness derived from NMA comparisons across these heterogeneous studies. The ERG within-trial analysis has also demonstrated that the ICERs are sensitive to the absolute levels of survival for a reference comparator, as well as to relative treatment effects from the NMA. It is unclear whether the HER2CLIMB trial (which included a high proportion of patients with brain metastases) or other trials in the NMA (with few or no patients with brain metastases) provide a more realistic reflection of clinical practice.
What alternative approach has the ERG suggested?	Modelling a 'real-world' baseline for the survival extrapolations, based on a relevant, reliable cohort or a weighted average for HER2CLIMB patients with and without brain metastases. In the absence of subgroup-specific estimates of relative effects, NMA results could be used to model results for the direct comparators, as in the current model.
What is the expected effect on the cost-effectiveness estimates?	This is uncertain. We anticipate that the survival predictions would be less favourable than with the company's NMA-based model. But if the proportion of patients with brain metastases in practice is lower than that in HER2CLIMB, the results should be more favourable than with the ERG's within-trial analysis.
What additional evidence or analyses might help to resolve this key issue?	An exploratory analysis with a modelled reference arm that is representative of the population in clinical practice.

Issue 4 Health state utilities

Report section	4.2.9
Description of issue and why the ERG has identified it as important	The company conducted a revised analysis of the HER2CLIMB utility data as a response to clarification question B2 for which they used a repeated measures model with adjustment for baseline values. Whilst this approach is preferable to the approach in the original base case, there are concerns about the method of analysis and lack of detail in

	reporting. We note the potential for bias due to missing data, particularly at the post-treatment follow-up. The company's revised base case includes the post-progression utility of 0.496 from Lloyd et al.[1] However, the TA423 committee concluded that the most plausible post-progression utility lies between the Lloyd et al. estimate and an estimate of 0.679 (Crott and Briggs mapping of the Study 301 trial data).[2 3] In TA423 the same post-progression utility was used across treatments. By comparison, the post-progression utility for the tucatinib combination in the company's revised base case (0.698) is much higher than that assumed for eribulin, capecitabine and vinorelbine (0.496). This difference is not based on comparative evidence and seems implausible. It is not clear why such a large difference should persist after progression and treatment discontinuation. The clinical plausibility of the difference in pre-progression utility for the tucatinib combination (0.762) and comparators (0.706 for eribulin and 0.701 for capecitabine and vinorelbine) is questionable. This may well relate to differences in the trial populations (HER2CLIMB versus Study 301)[3 4] or valuation methods (crosswalk EQ-5D versus Crott and Briggs mapping), rather than to differences in treatment-related quality of life. Clinical advice to the ERG is that the adverse effects and quality of life will be similar across these treatments.
What alternative approach has the ERG suggested?	We suggest that the same utilities should be used for all treatments in the pre- and post-progression health states. We prefer the HER2CLIMB utilities, as these are derived from EQ-5D data in a relevant trial population, using NICE-recommended methods. An alternative that provides continuity between proposals for the same indication would be to use estimates from TA423.
What is the expected effect on the cost-effectiveness estimates?	ERG scenario analysis shows that the model is sensitive to the assumption of equal pre-progression utility and/or equal post-progression utilities between treatments (Table 35 and Table 39). ICERs were similar with estimates from the HER2CLIMB trial or TA423.
What additional evidence or analyses might help to resolve this key issue?	Further evidence and expert opinion on the plausibility of differences between treatments in health-related quality experienced before progression and after progression. Further information on the methods analysis of HER2CLIMB EQ-5D data and how missing data was handled.

1.6 Other key issues: summary of the ERG's view

The ERG does not have any other key issues to discuss.

1.7 Summary of ERG's preferred assumptions and resulting ICER

Based on the ERG critique of the company's (revised) cost effectiveness model, we have identified six key aspects of the company base case with which we disagree. Our preferred model assumptions are the following:

- **Within-trial analysis:** We use the OS and PFS fitted to HER2CLIMB trial data for the tucatinib combination and trastuzumab + capecitabine (see section 4.2.6).
- **Relative effects for other comparators** from the HR NMA with random effects and the ERG correction for the Pivot upper confidence limit (section 3.6).
- **Health state utilities** from HER2CLIMB EQ-5D analysis applied to all treatments (section 4.2.9.2).
- ERG scenario for the use of **subsequent treatments** (section 4.2.10.2)
- **Adjustment of utilities** for age
- Costs for **drug wastage**.

The cumulative effect of ERG preferred assumptions to the company's (revised) base case is shown in Table 3 .

Table 3 Cumulative change from company base case to ERG's preferred model assumptions

Treatment	Cost [a]	QALYs	Pairwise ICERs	Change to pairwise ICERs
Revised company base case				
Capecitabine	[redacted]	[redacted]	[redacted]	
Vinorelbine	[redacted]	[redacted]	[redacted]	
Tras + cap	[redacted]	[redacted]	[redacted]	
Eribulin	[redacted]	[redacted]	£37,483	
Tuc + tras + cap	[redacted]	[redacted]		
+ Within-trial analysis (PFS and OS, with HR NMA fixed effect)				
Capecitabine	[redacted]	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]		
+ HR NMA random effects with ERG Pivot correction				
Capecitabine	[redacted]	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]		
+ HER2CLIMB utilities (0.762 pre-progression, 0.698 post-progression)				

Capecitabine	[redacted]	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]		
+ Age-adjustment for utilities				
Capecitabine	[redacted]	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]		
+ ERG subsequent treatment scenario (50% tras, 20% cap/vin: [redacted] per person)				
Capecitabine	[redacted]	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]		
+ Include costs for drug wastage (ERG preferred analysis)				
Capecitabine	[redacted]	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]		

The ERG conducted a range of scenario analyses on our preferred base case model. These are presented in Table 39 of this report.

2 INTRODUCTION AND BACKGROUND

2.1 Introduction

This report is a critique of the company's submission (CS) to NICE from Seagen on the clinical effectiveness and cost effectiveness of tucatinib (TUKYSA®) with trastuzumab and capecitabine for treating HER2-positive unresectable locally advanced or metastatic breast cancer after two or more anti-HER2 therapies. It identifies the strengths and weakness of the CS. Clinical experts were consulted to advise the evidence review group (ERG) and to help inform this report.

Clarification on some aspects of the CS was requested from the company by the ERG via NICE on 7th May 2021. A response from the company via NICE was received by the ERG on 25th May 2021 and this can be seen in the NICE committee papers for this appraisal.

2.2 Background

2.2.1 Background information on HER2-positive locally advanced or metastatic breast cancer

The CS (Section B1.3) provides a brief overview of the condition, describing the course of the disease; global and national epidemiology; and the impact on health-related quality of life (HRQoL) and survival. The CS highlights the significant impact of metastases to the brain, stating that this affects up to 50% of people over the course of the disease. In contrast, the ERG notes that a recent meta-analysis of epidemiological studies[5] estimated the proportion of people with HER2+ breast cancer with brain metastases to be lower, at 31%. Similarly, the ERG's expert clinical advisor estimates that around a third of patients with HER2+ metastatic disease develop brain metastases.

The CS states that the brain acts as a "sanctuary site" for HER2+ disease due to the inability of current drug treatments to penetrate the blood-brain barrier. Survival after the occurrence of brain metastases in HER2+ disease is poor: 1-year survival of 50% and 3-year survival of 16%. As will become apparent in subsequent sections of this report, brain metastasis is a disease characteristic of significant importance in the clinical effectiveness and cost-effectiveness evidence submitted.

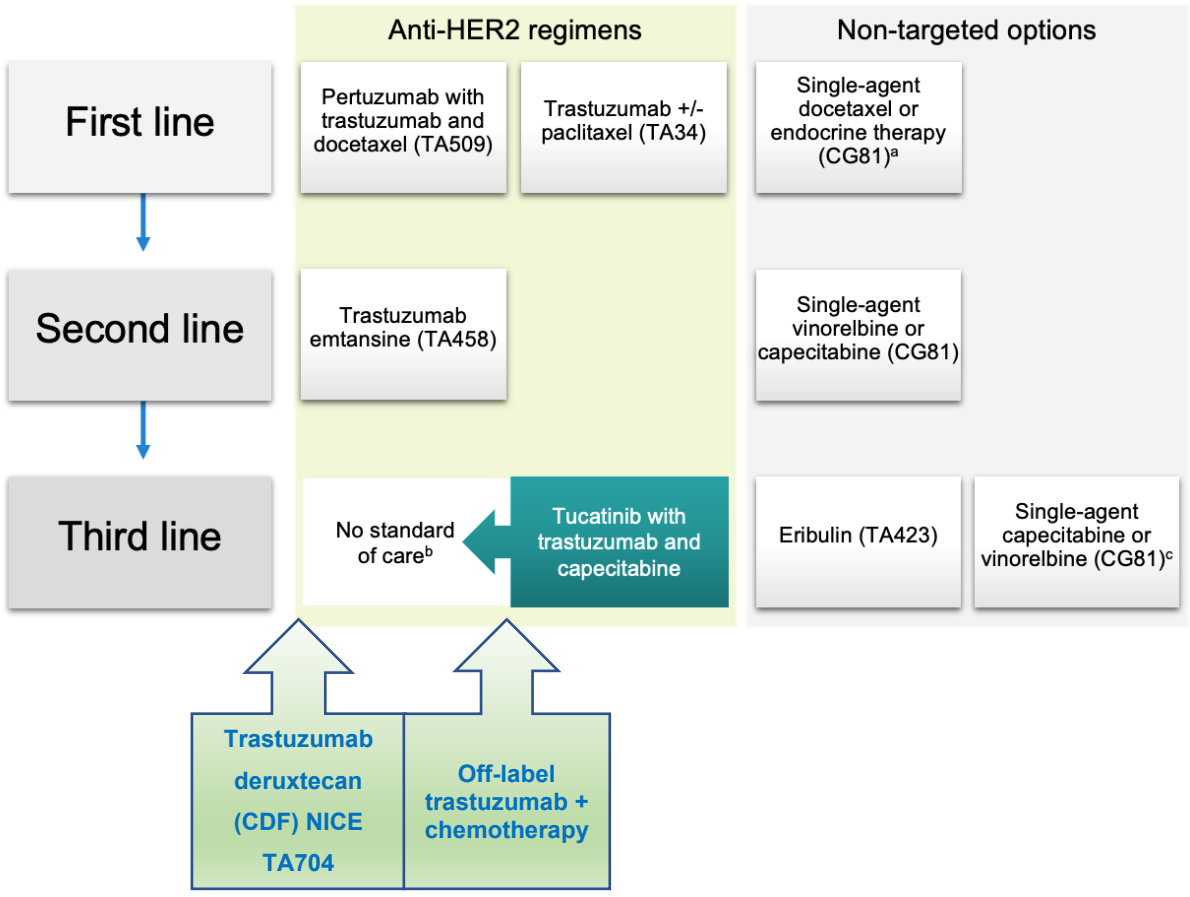
2.2.2 Background information on tucatinib

The CS describes the mechanism of action of tucatinib, an oral tyrosine kinase inhibitor (TKI), and that of two other drugs it is used in combination with: trastuzumab (a recombinant

humanised IgG1 monoclonal antibody) and capecitabine (an anti-metabolite chemotherapy). Tucatinib is an orally bioavailable, reversible small molecule TKI that is highly specific to HER2, and therefore defined as a targeted treatment. Tucatinib received its UK regulatory approval on 22 February 2021 and is indicated in combination with trastuzumab and capecitabine for the treatment of adult patients with HER2-positive locally advanced or metastatic breast cancer who have received at least two prior anti-HER2 treatment regimens.

2.2.3 The position of tucatinib in the treatment pathway

Figure 1 is the company's depiction (with minor adaptation by the ERG) of the current treatment pathway for locally advanced and metastatic HER2 positive breast cancer, and the proposed third line positioning of tucatinib in combination with trastuzumab and capecitabine.



Source: adapted by the ERG from CS Figure 1

CDF – Cancer Drugs Fund

Figure 1 Current treatment pathway in HER2+ metastatic breast cancer in England and proposed positioning of tucatinib

As patients progress through successive lines of therapy they may receive anti-HER2 regimens and/or non-targeted treatments (e.g. chemotherapies or endocrine therapy for those patients whose tumours express hormone receptors).

The CS suggests that there is no standard of care for patients progressing to third line therapy, as no anti-HER2 regimens are currently recommended. However, the CS also mentions that eribulin (a non-targeted chemotherapy) "is the most plausible standard of care" (CS page 19) as it is the only NICE recommended treatment at third line (see NICE TA423). The ERG notes that NICE clinical guideline (CG81) recommends that patients may receive treatment with non-targeted chemotherapies such as capecitabine or vinorelbine. The ERG's expert clinical advisor points out that NICE guidance on eribulin is for metastatic breast cancer irrespective of HER2 status (+/-) and that capecitabine, vinorelbine and eribulin as single agents are equally appropriate third line treatment options. Our expert advisor also notes that single agent chemotherapy in combination with trastuzumab is used (off label) at third line, and that hormone therapy may also be used at third line in people with hormone receptor positive disease (though most hormone therapy tends to be given earlier in the pathway).

Since the CS was written, NICE has published guidance on an anti-HER2 regimen for use at third line: Trastuzumab deruxtecan is recommended for use within the Cancer Drugs Fund (CDF) as an option for treating HER2-positive unresectable or metastatic breast cancer in adults after 2 or more anti-HER2 therapies (NICE TA704). After a period of further data collection based on the use of trastuzumab deruxtecan in the NHS, and availability of data from trials currently in progress, NICE will review whether this treatment is cost-effective and can be recommended for routine commissioning in the NHS. Thus, standard care may potentially include this treatment in due course.

The company argues that there is unmet need for an efficacious third line therapy that can target brain metastases. They cite the inability of systemic treatments (e.g. single agent chemotherapies) to treat brain metastases effectively and the potential for greater clinical benefit with targeted treatment in the third line setting. The ERG's expert clinical advisor agreed this is a significant unmet need and noted that trastuzumab deruxtecan is not expected to target brain metastases because it is unlikely to cross the blood-brain barrier.

Finally, the ERG's clinical expert commented that the treatment pathway does not reflect the fact many patients will have received prior adjuvant treatment for their primary tumour. Thus, only a minority will present to secondary care with de novo metastases. In this respect the requirement to have received at least two prior anti-HER2 treatment regimens before receiving tucatinib could be interpreted as including anti-HER2 adjuvant treatments.

ERG conclusion

The ERG considers that tucatinib is appropriately positioned as a third line treatment of metastatic disease given the lack of available targeted anti-HER2 treatments for patients whose disease has progressed in this setting. Expert clinical advice to the ERG is that trastuzumab (used off label) in combination with chemotherapy (e.g. capecitabine) is used at third line. This is not, however, included as a comparator in the NICE scope or the decision problem.

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

Table 4 compares the company's decision problem to the final scope for this appraisal issued by NICE. The ERG concludes that the decision problem adheres to the NICE scope, with the following exceptions:

- **Population**. The population is not restricted to people with unresectable locally advanced or metastatic breast cancer. This widens the population to also include people whose tumours are resectable, and potentially the effects of tucatinib may not necessarily be the same for them as they are for people with unresectable tumours.
- **Comparators.** The base case economic analyses includes only one of the three comparators – eribulin. The other two comparators (capecitabine and vinorelbine) are included in additional analyses "for completeness". The company regards the three comparators to be similar in efficacy and safety based on clinical advice. Expert clinical advice to the ERG suggests there is variation in clinical practice at third line. For example, some clinicians may continue trastuzumab treatment with the addition of a single agent chemotherapy (e.g. capecitabine), whilst others may use single agent capecitabine. We discuss the appropriateness of the company's approach in our critique of cost effectiveness, in section 4 of this report.
- **Subgroups.** Cost effectiveness is not estimated for the subgroup of people with brain metastases. Although the HER2CLIMB trial included patients with brain metastases, this is atypical in breast cancer treatment clinical trials. Trials of the comparator drugs tended to exclude patients with unstable brain metastases. The

company was therefore unable to conduct an indirect treatment comparison to inform this cost-effectiveness analysis. We discuss this in sections 3.3 and 3.4 of this report.

Table 4 Summary of the decision problem

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	ERG comments
Population	People with HER2-positive unresectable locally advanced or metastatic breast cancer who have had 2 or more prior anti-HER2 therapies	As per scope, except not restricted to people with unresectable cancer.	Company state this is because tucatinib's licence indication does not mention the term unresectable.	Expert clinical advice to the ERG is that resection is an option for a small number of patients and surgery would not be expected to be curative.
Intervention	Tucatinib with trastuzumab and capecitabine	As per scope	N/A	Decision problem matches the NICE scope
Comparators	• Eribulin • Capecitabine • Vinorelbine	Only eribulin is included in the base case economic analyses. Capecitabine and vinorelbine are included in additional analyses.	Expert clinical advice to the company is that eribulin is the most plausible standard of care as it is used as a single agent and is the only treatment approved by NICE for use in the third-line setting.	Expert clinical advice to the ERG is that there is variation in practice and other treatments may be given.
Outcomes	The outcome measures to be considered include: • progression-free survival	As per scope	N/A	Decision problem matches the NICE scope

	• overall survival • response rate • duration of response • adverse effects of treatment • health-related quality of life			
Subgroups	People with brain metastases (where evidence allows)	People with untreated and previously treated brain metastases	Clinical effectiveness evidence is presented for this subgroup, but not cost effectiveness estimates, due to the lack of available evidence for potential comparators including a similar subgroup of patients.	Decision problem does not completely match the NICE scope The ERG concurs that evidence for this subgroup in comparator trials is lacking (see sections 4.2.3 and 4.2.6)

Source: Adapted from CS Table 1

3 CLINICAL EFFECTIVENESS

3.1 Critique of the methods of review(s)

Evidence for the clinical effectiveness of tucatinib was identified from a single broad systematic literature review (SLR); the full methods of which are reported in CS Appendix D and summary points referred to in CS section B.2.1. This SLR also identified studies for the indirect comparisons between the treatments identified in the decision problem (referred to in CS section B.2.9.1.1). The ERG provides a critique of the methods and processes of the SLR in Table 5. We have no concerns that the search strategy or study eligibility assessments may have missed potentially eligible studies. The CS summarises the included studies appropriately and the ERG assessment of the risk of bias generally concurs with that of the CS. For the NMA, additional study eligibility criteria were applied. The ERG has no concerns with the final selection of studies, although notes that five peripheral studies were included in the NMA. The ERG has no concerns over the conduct of the NMA analysis.

Table 5 ERG appraisal of systematic review methods

Systematic review components and processes	ERG response (Yes, No, Unclear)	Comments
Was the review question clearly defined using the PICOD framework or an alternative?	Yes	PICOD framework used. Different criteria for titles and abstracts as full text review but minimal differences (CS Appendix D, Tables 10 and 11).
Were appropriate sources of literature searched?	Yes	The sources searched were comprehensive. The date of the last search was November 2020.
What time period did the searches span and was this appropriate?	Yes	All years were searched (CS Appendix D page 4)
Were appropriate search terms used and combined correctly?	Yes	Search terms were appropriate and combined correctly, the RCT search strings were somewhat sparse and not translated consistently across the databases, however, the ERG has no concerns about this.

Were inclusion and exclusion criteria specified? If so, were these criteria appropriate and relevant to the decision problem?	Yes	• Two stage approach, slightly different criteria between the two stages of review but reasonable and with criteria that were relevant to the decision problem. • The company included systematic reviews at the titles and abstract review stage and checked reference lists of 'robust systematic reviews' (CS Appendix D page 4) but no further detail was provided and it is unclear if any references were identified from these systematic reviews. • The ERG has checked the excluded studies lists in CS Appendix D (Tables 13 and 14) and exclusions appear appropriate based on the information provided. • The ERG has also cross-checked studies identified in a recent NMA [6] to validate the company's approach; no additional references were identified. • Studies assessed as meeting the PICOD criteria were then assessed on additional eligibility criteria for inclusion in the NMA. (CS Appendix Table 15). The ERG has no concerns with the final study selection for the NMA.
Were study selection criteria applied by two or more reviewers independently?	Yes	Both screening and full text assessment were undertaken by two independent reviewers, CS Appendix D, Page 23.
Was data extraction performed by two or more reviewers independently?	No	Data extraction was undertaken by one reviewer (CS Appendix D Page 24)
Was a risk of bias assessment or a quality assessment of the included	Yes	NICE criteria for RCTs (CS Appendix D1.3) and ERG assessment generally concurs

studies undertaken? If so, which tool was used?		
Was risk of bias assessment (or other study assessment) conducted by two or more reviewers independently?	Unclear	Unclear if single reviewer or double reviewer (CS Appendix D1.3)
Is sufficient detail on the individual studies presented?	Yes	CS describes trial methodology, outcomes and results of the pivotal trial in Sections B.2.2-2.6 and summarises the comparator trials in CS Section B.2.9.1.3 and CS Appendix D.
If statistical evidence synthesis (e.g. pairwise meta-analysis, ITC, NMA) was undertaken, were appropriate methods used?	Yes	The CS NMA analysis was well-conducted and appropriate studies were included. The ERG preferred the random effects NMA, discussed in Section 3.2.4.

NMA – Network meta-analysis; PICOD – population, intervention, comparator(s), outcome(s) and study design(s)

3.2 Critique of studies of the technology of interest, the company's analysis and interpretation (and any standard meta-analyses of these)

3.2.1 Included studies

The key source of clinical effectiveness evidence in the CS is the HER2CLIMB trial, a company-sponsored Phase II randomised, double-blind controlled trial of tucatinib in combination with trastuzumab and capecitabine versus placebo in combination with trastuzumab and capecitabine for unresectable locally advanced or metastatic HER2+ breast cancer. Although the trial was originally designed as a Phase II trial, the company states that the eventual sample size attained, and the conduct of the trial, made it consistent with the standards of a Phase III trial. HER2CLIMB is currently ongoing in an open-label extension where the placebo group can switch to tucatinib combination treatment. The CS does not appear to specify the trial's completion date, but the National Clinical Trial record on clinicaltrials.gov states this will be 31st May 2022.[7]

3.2.1.1 Study characteristics

The HERCLIMB2 trial was conducted in 15 countries, with 10 centres in the UK. The trial was initiated with 180 participants, but sample size was increased on two occasions to improve statistical robustness (see section 3.2.4 for details). The original study protocol was also amended after the first sample size increase to include European Quality of Life Working Group Health Status Measure 5 Dimensions, 5 Levels (EQ-5D-5L) for subsequently enrolled participants. The history and timelines of HER2CLIMB can be seen in CS Figure 2.

The population of HER2CLIMB is people with unresectable locally advanced or metastatic HER2+ breast carcinoma previously treated with trastuzumab, pertuzumab, and trastuzumab emtansine (T-DM1). The ERG's clinical expert notes that these treatments are standard practice in the UK. Key inclusion and exclusion criteria are summarised in CS Table 4. Of note, participants with brain metastases were eligible, including those with active (treated progressing or untreated) or treated stable brain metastases. Clinical trials in this setting generally exclude patients with brain metastases or include only those with stable brain metastases. People were excluded if they had previously received the following treatments in the metastatic setting: capecitabine, lapatinib within 12 months (except if lapatinib was given for ≤21 days and was discontinued for reasons other than disease progression or severe toxicity), neratinib, afatinib or other investigational HER2/ epidermal growth factor receptor or HER2 tyrosine kinase inhibitor.

Eligible participants were randomised in a 2:1 ratio to one of the following arms:

Tucatinib combination

- Tucatinib: 300 mg orally twice daily continuously during treatment period
- Trastuzumab: 8 mg/kg intravenous loading dose then 6 mg/kg (or 600 mg subcutaneous injection) once every 21 days
- Capecitabine: 1000 mg/m^2 orally twice daily days 1–14 of each 21-day cycle

Placebo combination

- Placebo orally twice daily continuously during treatment period
- Trastuzumab: 8 mg/kg intravenous loading dose then 6 mg/kg (or 600 mg subcutaneous injection) once every 21 days
- Capecitabine: 1000 mg/m^2 orally twice daily days 1–14 of each 21-day cycle

Treatments continued until unacceptable toxicity, disease progression, withdrawal of consent or study closure.

The key features of HER2CLIMB are summarised in Table 6.

Table 6 Summary of HER2CLIMB trial characteristics

Trial characteristic	Description
Study design	Phase II randomised, double-blind, controlled clinical trial (originally registered as Phase II but amended to be consistent with Phase III)
Number and location of centres	155 sites across 15 countries: US, Canada, Austria, Belgium, Czechia, Denmark, France, Germany, Italy, Portugal, Spain, Switzerland, Israel, Australia and UK (10 sites; [redacted] patients)
Study population	Unresectable locally advanced or metastatic HER2+ breast carcinoma previously treated with trastuzumab, pertuzumab and T-DM1
Intervention	Tucatinib in combination with trastuzumab and capecitabine (Primary endpoint population n=320) (Total population n=410)
Comparator	Placebo in combination with trastuzumab and capecitabine (Primary endpoint population n=160) (Total population n=202)
Primary outcome	PFS in the primary endpoint population (n=480)
Key secondary outcomes	PFS in all randomised patients with brain metastases (n=291) Overall survival in total population (n=612) ORR in total population (n=612)
Randomisation stratification factors	Known history of treated or untreated brain metastases (yes or no), ECOG performance status (0 or 1) geographic region (US, Canada or rest of world).
Status	Ongoing (open-label extension)
Latest available data	4th September 2019 Median follow-up 14.0 months (total population)
Pre-specified sub-groups	• Age (≥65 or <65 years) • Race (white or non-white) • Hormone receptor status (HmR+ or HmR-) • Baseline brain metastases (yes or no) • ECOG performance-status score (0 or 1) • Geographic region (US and Canada or rest of world)
Source: CS section B.2.3.1, CS Table 3 and 4.	

ECOG, Eastern Cooperative Oncology Group; ORR, objective response rate; PFS progression-free survival T-DM1, trastuzumab emtansine.

3.2.1.2 Patients' baseline characteristics

Baseline characteristics in HER2CLIMB are summarised in Table 7 below. The arms are generally well balanced for both the primary endpoint population and the total study population, although the ERG notes there is a slight imbalance in the proportion of white participants and those with liver metastases in the primary endpoint population, both of which are slightly higher in the placebo-combination arm. People with liver metastases generally have a worse prognosis than those without, but the extent this may affect treatment outcomes in the trial arms is unclear. A subgroup analysis was not undertaken for presence/absence of liver metastases.

[redacted] The ERG's clinical expert notes that very few patients have locally advanced disease at third-line treatment. These patients have biologically distinct disease which has a propensity not to metastasise and which has a better response to treatment compared to distant (particularly visceral) metastases. Around 63% of participants had Stage 0 to III disease at initial diagnosis and would have already received treatment for their primary tumour, although details of types of treatment were not provided. The proportion with stage IV disease at initial diagnosis (36%) is high compared with UK practice, where around 10-15% present with stage IV disease (estimated by the ERG's clinical expert). Some of the participants were extensively pre-treated, with up to 17 prior lines of therapy (median 4 lines), and up to 14 lines of prior therapy in the metastatic setting (median 3 lines).

The proportion of patients with brain metastases was 45.6% in the primary endpoint population and 47.5% in the total study population. The ERG clinical expert considers that this is higher than expected in UK clinical practice, possibly due to study screening identifying asymptomatic cases or to the trial attracting participants with active brain metastases. Approximately 60% of the brain metastases were active (i.e. either treated and progressing or untreated).

Table 7 Baseline characteristics in HER2CLIMB

	Primary endpoint population (N=480)		Total study population (N=612)	
Variable, no. (%) if not otherwise stated	**Tucatinib combination (n=320)**	**Placebo combination (n=160)**	**Tucatinib combination (n=410)**	**Placebo combination (n=202)**
Female sex	317 (99.1)	158 (98.8)	407 (99.3)	200 (99.0)
Age				
<65 years	252 (78.8)	132 (82.5)	328 (80.0)	168 (83.2)
≥65 years	68 (21.3)	28 (17.5)	82 (20.0)	34 (16.8)
Median – years	54.0	54.	55.0	54.0
Race				
Asian	17 (5.3)	3 (1.9)	18 (4.4)	5 (2.5)
Black/African American	30 (9.4)	13 (8.1)	41 (10.0)	14 (6.9)
White	225 (70.3)	125 (78.1)	287 (70.0)	157 (77.7)
Unknown/other	48 (15.0)	19 (11.9)	64 (15.6)	26 (12.9)
Region				
US/Canada	204 (63.8)	103 (64.4)	246 (60.0)	123 (60.9)
Rest of world	116 (36.3)	57 (35.6)	164 (40.0)	79 (39.1)
Hormone receptor status				
ER and/or PgR-positive	190 (59.4)	99 (61.9)	243 (59.3)	127 (62.9)
ER and PgR-negative	126 (39.4)	61 (38.1)	161 (39.3)	75 (37.1)
Other	4 (1.3)	0	6 (1.5)	0
ECOG performance status				
0	159 (49.7)	76 (47.5)	204 (49.8)	94 (46.5)
1	161 (50.3)	84 (52.5)	206 (50.2)	108 (53.5)
Disease status at study entry				
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Stage IV at initial diagnosis	108 (33.8)	67 (41.9)	143 (34.9)	77 (38.1)
Presence or history of brain metastases	148 (46.3)	71 (44.4)	198 (48.3)	93 (46.0)
Previously treated stable	Not reported	Not reported	80 (40.4)	37 (39.8)
Previously treated progressing	Not reported	Not reported	74 (37.4) [a]	34 (36.6) [a]
Untreated	Not reported	Not reported	44 (22.2) [a]	22 (23.7) [a]
Location of other metastases				
Lung	160 (50.0)	82 (51.3)	200 (48.8)	100 (49.5)
Liver	108 (33.8)	64 (40.0)	137 (33.4)	78 (38.6)
Bone	178 (55.6)	85 (53.1)	223 (54.4)	111 (55.0)
Prior lines of therapy, median (range)	4.0 (2, 14)	4.0 (2,17)	4.0 (2, 14)	4.0 (2,17)

	Primary endpoint population (N=480)		Total study population (N=612)	
Variable, no. (%) if not otherwise stated	**Tucatinib combination (n=320)**	**Placebo combination (n=160)**	**Tucatinib combination (n=410)**	**Placebo combination (n=202)**
Prior lines of therapy in the metastatic setting, median (range)	3.0 (1, 14)	3.0 (1, 13)	3.0 (1, 14)	3.0 (1, 13)
Prior therapies				
Trastuzumab	320 (100)	160 (100)	410 (100)	202 (100)
Pertuzumab	320 (100)	159 (99.4)	409 (99.8)	201 (99.5)
T-DM1	320 (100)	160 (100)	410 (100)	202 (100)
Lapatinib	22 (6.9)	10 (6.2)	24 (5.9)	10 (5.0)

Source: CS Table 6, CS Appendix E Table 1

ECOG, Eastern Cooperative Oncology Group; ER, oestrogen receptor; ITT, intent to treat; no, number; OS, overall survival; PgR, progesterone receptor; T-DM1, trastuzumab emtansine; US, United States

[a] Previously treated progressing and Untreated rows are transposed in CS Table 6. [b] CSR Table 4 and Table 8.

3.2.1.3 Ongoing studies

No additional ongoing studies are reported in the CS. The ERG is not aware of any relevant ongoing studies.

ERG conclusion on included studies

The ERG considers the population of the HER2CLIMB trial to be generally representative of the target population with unresectable locally advanced or metastatic HER2+ breast cancer previously treated with two or more anti-HER2 therapies. We note, however, that the trial participants have experienced more previous lines of therapy than typical in UK practice. Furthermore, the proportion of patients with brain metastases included in the trial appears to be higher than in clinical practice. The comparator trial arm (which includes trastuzumab and capecitabine in combination) is not licensed for use at third line and is not included as a comparator treatment in the NICE scope for this appraisal.

3.2.2 Risk of bias assessment

The company assessed the quality of HER2CLIMB using criteria recommended by NICE. This is presented in CS Table 7 and CS Appendix D; the company's comments on quality

assessment differ slightly between these sources, but the overall judgements are the same. A comparison of the company's and the ERG's assessment of HERCLIMB2 can be seen in 9.1. The ERG generally agrees with the company's assessment, although we note the following minor points:

- There is a slight imbalance in the proportion of white participants and those with liver metastases, both of which are slightly higher in the placebo-combination arm. The implications of this are unclear.
- There were no unexpected imbalances in dropouts between groups. A higher proportion discontinued treatment in the placebo-combination arm (86.3%) than in the tucatinib-combination arm (70.8%), mostly due to progressive disease in the placebo-combination arm (68% vs 50%) (CS Appendix D Figure 11). Discontinuations due to adverse events were higher with tucatinib (5.7% vs 3.0%).
- The protocol [8] lists duration of response (DOR) and clinical benefit rate (CBR) determined by the investigator as well as blinded independent central review (BICR) as secondary endpoints, however only BICR-determined DOR and CBR are reported in the CS. Investigator results are reported on the National Clinical Trial record [7] (although there is a discrepancy between the BICR DOR reported on the trial record and that reported in the CS and CSR).

Overall, the ERG considers there is a low risk of bias in HER2CLIMB.

3.2.3 Outcomes assessment

The efficacy and safety outcome measures included in the HER2CLIMB trial are reported in CS Section 2.3 and CS Table 4. These are standard outcomes used in trials of cancer drugs and their definitions appear appropriate. An overview of all the outcomes and ERG comments is provided in Appendix 9.2.

The key outcomes were assessed by blinded independent central review (BICR) using RECIST (Response Evaluation Criteria in Solid Tumours) criteria version 1.1. In the protocol, the CSR and the National Clinical Trial record [7] secondary outcomes as determined by the investigator assessments were also provided. Investigator assessments were included predominantly to ensure treatment decisions could be made in a timely manner.

The trial's primary outcome was PFS in the 'primary endpoint population' (see Section 3.2.4 of this report for further detail on the trial's statistical analysis populations). Secondary outcomes were classed as being 'key' or 'other'. The key secondary outcomes included

overall survival (OS) and the objective response rate (ORR) in the total population, and PFS in patients with baseline brain metastases at baseline. The latter outcome was included in response to promising results from an early phase I dose-escalation trial of tucatinib, trastuzumab and capecitabine in people with brain metastases.[9] Outcomes were assessed every six weeks in the initial 24 weeks and then once every nine weeks until disease progression, initiation of a new therapy, withdrawal of consent, or study closure (CS Figure 3).

The patient-reported outcome of health-related quality of life (HRQoL) was measured using the EQ-5D-5L in a subset of the total population following a protocol amendment (see Section 3.2.5). As such, the outcomes from the EQ-5D should be considered as exploratory. The CS reports baseline and endpoint data for the EQ-5D descriptive system and the EQ-5D Visual analogue scale (VAS) but not for the EQ-5D index scores, however, these were provided in response to clarification question A2 (see Section 3.2.5.4).

The outcomes that inform the economic model are PFS, OS, adverse events and HRQoL (CS Table 4).

ERG conclusion on outcomes assessment

Overall, the outcome measures included in the HER2CLIMB trial are appropriate for a cancer treatment trial and the ERG has no concerns over their definition or measurement.

3.2.4 Statistical methods in the HER2CLIMB trial

The CS presented only summary descriptions of the statistical analyses of the HER2CLIMB trial, however, the trial publication [4 8] includes the trial protocol and the statistical analysis plan (SAP) which have been checked by the ERG.

The original sample size (n=180) was increased prior to unblinding or knowledge of the trial results on two occasions (CS B.2.3.1). The first increase, to n=480, was necessary for the upgrade of the trial's status from phase 2 to phase 3 to support the tucatinib licensing registration. This was undertaken at approximately 12 months and the n=480 became the primary endpoint population. The second increase, to a target of n=600, was to improve power for the key secondary endpoint of PFS in patients with brain metastases (described in

CS B.2.4) and was undertaken at approximately 3 years. The ERG does not have any concerns with the sample size or the power analysis.

The primary analysis of PFS was conducted for the first 480 participants recruited. The CS says (CS B.2.3.1) this was to avoid potential bias from early progression in the overall population owing to the short follow-up of some participants. Additional analyses sets were as described in Table 8. The statistical approach for the primary outcome and key secondary outcomes is described in Appendix 9.2, the ERG has taken information from CS B.2.4 and the HER2CLIMB trial publication, protocol and SAP. [8] Multiplicity was controlled using a group sequential procedure described in Appendix 9.2. The additional secondary outcomes were not subject to type 1 error control and are deemed as exploratory. The handling of missing data, sensitivity analyses and analysis of pre-specified subgroups were described in the trial protocol / SAP and are summarised in Appendix 9.2. The ERG does not have any significant concerns with the analysis plan or the statistical analyses.

Table 8 HER2CLIMB analysis populations

Analysis population	Definition	Analyses
Primary endpoint population	First 480 patients randomised	Primary endpoint of PFS per BICR
ITT-OS	Total study population (N=612)	Secondary endpoints of OS and confirmed ORR
ITT-PFS brain metastases	All randomised patients with brain metastases (N=291)	Secondary endpoint of PFS per BICR in brain metastases subgroup
Safety	All randomised who received at least 1 dose of study treatment (N=601)	Safety analyses
Source: CS Table 5		

BICR, blinded independent central review; BM, brain metastases; ITT, intent to treat; OS, overall survival, ORR, objective response rate; PFS, progression-free survival.

3.2.5 Efficacy results of the intervention studies

The CS presents results from the HER2CLIMB trial up to 4th September 2019. The ERG assumes this is the final analysis for the trial as no further planned analyses are mentioned in the CS.

In this section we focus on the clinical effectiveness outcomes from the trial that inform the economic model, as follows:

- PFS (primary endpoint); assessed in the first 480 randomised patients, the ITT-PFS population
- OS (key secondary endpoint); assessed in the total trial population (n=612), the ITT-OS population
- HRQoL (exploratory endpoint) assessed in a subset of patients (n=331) from the total trial population. The subset was used because baseline data collection for HRQol measures did not start until part way through the trial following a protocol amendment.

The HER2CLIMB trial was also designed to assess the effect of the tucatinib combination on PFS in patients with brain metastases; a subgroup relevant to the NICE scope for this appraisal. Thus, PFS was assessed as a key secondary endpoint in a subset of the total trial population with active or stable brain metastases at baseline, the 'ITT-PFS brain metastases' population (N=291). We summarise the results from this analysis below along with results from an exploratory analysis of PFS in patients without brain metastases.

The outcome measures ORR and DOR do not inform the economic model but are listed in the NICE scope and are summarised briefly in this section. Additional outcomes which do not inform the economic model (clinical benefit rate and time to new brain lesions or death). are not reproduced here but can be found in CS sections B.2.6.6 and B.2.6.8. In addition, exploratory analyses were conducted within the ITT-PFS brain metastases population (central nervous system PFS, intracranial objective response rate, overall survival, duration of response and HrQoL) and are presented in CS Appendix E.

Safety data, including adverse events and treatment exposure, are summarised in section 3.2.5.6.

3.2.5.1 Progression-free survival

Primary endpoint population

The primary endpoint in the HER2CLIMB trial was PFS measured by BICR according to RECIST 1.1. criteria in the primary endpoint population (ITT-PFS, n=480). A statistically significant reduction in the risk of progression or death was observed for the tucatinib-combination group compared to the placebo-combination group with a hazard ratio (HR) of 0.54 (95% CI: 0.42, 0.71; p<0.0012) and a 2 month increase in median PFS (Figure 2). Results were similar when PFS was measured by investigator assessment ([redacted]) in this study population (CSR section 11.1.4.1).

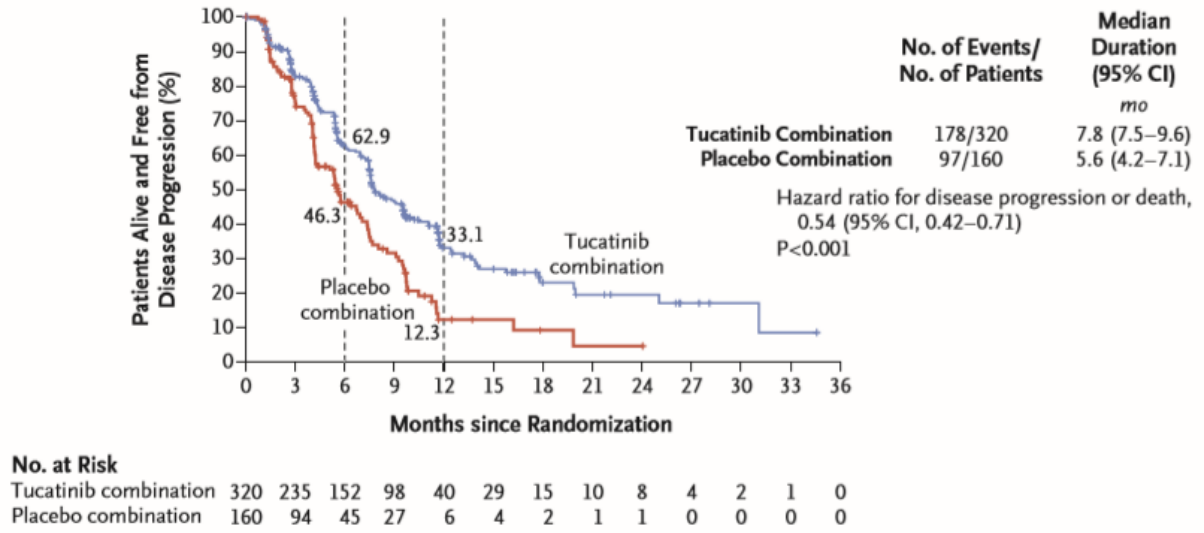


Figure 2 Kaplan-Meier estimate of PFS per BICR (primary endpoint population)

BICR, blinded independent central review; CI, confidence interval; HR, hazard ratio; mo, months; No, number
Source**:** CS Figure 4

Total population

In the total study population (ITT-OS, n=612), PFS with BICR results were comparable with those in the primary endpoint population with a HR of 0.54 (95% CI: 0.42, 0.68; CS section B.2.6.1).

3.2.5.2 Overall survival

A statistically significant reduction in the risk of death (key secondary endpoint) was observed for the tucatinib-combination group compared to the placebo-combination group in the total study population with a HR of 0.66 (95% CI 0.50, 0.88; p=0.005) and a 4.5 month improvement in OS (Figure 2).

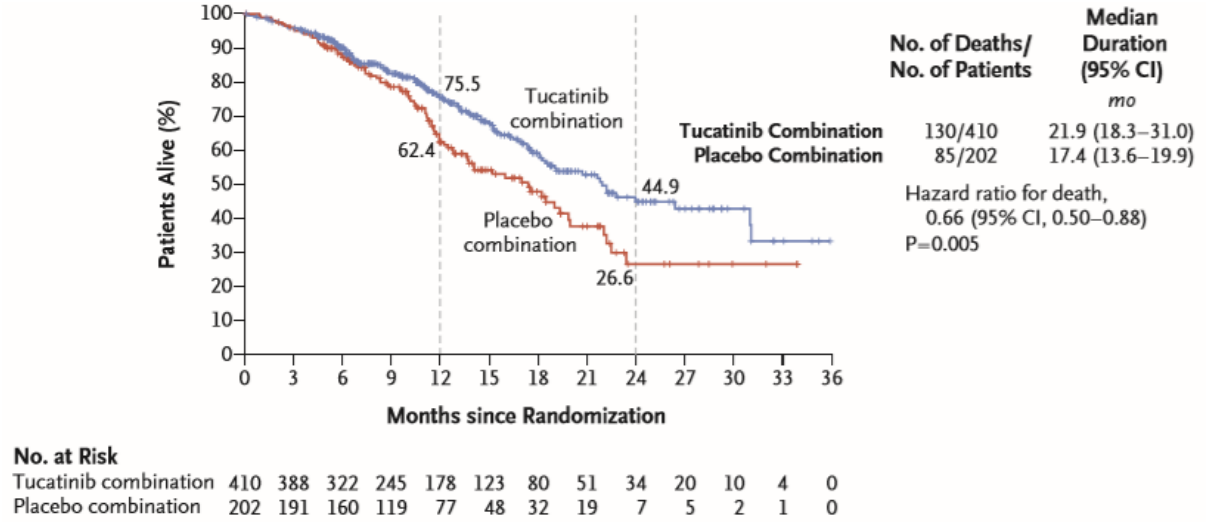


Figure 3 Kaplan-Meier estimate of OS per BICR (total study population; ITT-OS)

BICR, blinded independent central review; CI, confidence interval; mo, months; ITT, intent to treat; No, number; OS, overall survival; source: CS Figure 5

3.2.5.3 Objective response rate and duration of response

A higher proportion of patients had an objective response in the tucatinib-combination group compared to the placebo-combination group (46.0% vs 22.8%, p<0.00008; Table 9) in patients with measurable disease by BICR at baseline (n=511). Median DOR was longer in the tucatinib-combination group compared to placebo-combination group (median 8.3 vs 6.3 months) but the 95% confidence intervals overlapped suggesting no evidence of a difference between groups. (Table 9).

Table 9 Confirmed objective response per BICR and duration of response in patients with measurable disease at baseline

Outcome	Tucatinib combination (N=340)	Placebo combination (N=171)	p-value
Objective response, n (%)	138 (40.6)	39 (22.8)	<0.00008
95% CI	35.3, 46.0	16.7, 29.8	
Median duration of response (months)	8.3[a]	6.3[a]	Not reported[b]
95% CI	6.2, 9.7	5.8, 8.9	
Source: CS Table 8 and CS section B.2.6.7			

[a] in patients with a confirmed response
[b] nominal only

3.2.5.4 HRQoL outcomes

Baseline EQ-5D-5L data were available for a subset of the trial's total population (217 patients in the tucatinib-combination group and 112 patients in the placebo-combination group) as collection of baseline HRQoL data was only introduced after the start of the trial following a protocol amendment. The CS states that the baseline characteristics were similar between those patients with and without available baseline HRQoL data but does not present these data.

The CS presents graphical EQ-5D-5L VAS and subscale data only (CS Figures 10 and 11). The company additionally provided EQ-5D-5L index scores in response to clarification question A2 for the total population and for the subgroup of patient with brain metastases (Table 10). No comparative statistical analysis is provided (exploratory only). In the total population, [redacted]. In the subgroup of patients with brain metastases, [redacted]. The ERG clinical expert noted that these are meaningful changes with index scores rising to closer to population norms.

In both trial populations [redacted]. Similarly, the proportion of patients with moderate, severe or extreme problems were generally higher than baseline at the post-treatment 30 day follow up point, particularly in the tucatinib combination arm (CS Figure 11).

Table 10 EQ-5D-5L index scores in the total population and subgroup of patients with brain metastases

Time point	Total population		Patients with brain metastases	
	Tucatinib-combination	Placebo-combination	Tucatinib-combination	Placebo-combination
Baseline				
n	213	112	104	57
Mean (SD)	0.817 (0.158)	0.807 (0.190)	[redacted]	[redacted]
95% CI	0.796, 0.839	0.771, 0.842	[redacted]	[redacted]
Median	0.838	0.859	[redacted]	[redacted]
Cycle 3				
n	175	89	84	43
Mean (SD)	0.823 (0.164)	0.845 (0.155)	[redacted]	[redacted]
95% CI	0.799, 0.848	0.812, 0.878	[redacted]	[redacted]

Median	0.859	0.887	[redacted]	[redacted]
Cycle 5				
n	152	71	76	33
Mean (SD)	0.835 (0.185)	0.835 (0.157)	[redacted]	[redacted]
95% CI	0.806, 0.865	0.798, 0.872	[redacted]	[redacted]
Median	0.887	0.859	[redacted]	[redacted]
Cycle 7				
n	130	54	68	27
Mean (SD)	0.859 (0.143)	0.808 (0.188)	[redacted]	[redacted]
95% CI	0.834, 0.884	0.757, 0.860	[redacted]	[redacted]
Median	0.89	0.859	[redacted]	[redacted]
Cycle 9				
n	86	38	46	14
Mean (SD)	0.872 (0.129)	0.810 (0.246)	[redacted]	[redacted]
95% CI	0.845, 0.900	0.729, 0.891	[redacted]	[redacted]
Median	0.904	0.869	[redacted]	[redacted]
30 Day Follow Up				
n	72	42	29	21
Mean (SD)	0.738 (0.287)	0.778 (0.207)	[redacted]	[redacted]
95% CI	0.670, 0.805	0.713, 0.842	[redacted]	[redacted]
Median	0.835	0.835	[redacted]	[redacted]
Source: Clarification response Tables A2a and A2b; table drawn by ERG				

SD: standard deviation; CI: confidence interval

The ERG notes that only 50.6% and 47.0% of patients with baseline data and remaining in the study completed the EQ-5D-5L survey at the end of cycle 9 in the tucatinib- and placebo-combination groups respectively (CS Figure 10). Thus, data beyond cycle 9 are likely to be less reliable due to increased attrition. The CS does not describe any methods to impute missing HRQoL data. Section 4.2.9.1 of this report describes how the index scores from HER2CLIMB are used in the economic model.

3.2.5.5 Subgroup analyses

Results for PFS per BICR in patients with and without brain metastases

The CS reports results from the total population separately for patients with brain metastases (PFS brain metastases population) at baseline (key secondary endpoint, CS section B.2.6.3) and for patients without brain metastases (prespecified exploratory endpoint, CS section B.2.6.4). A statistically significant improvement in PFS was observed in patients with brain metastases in the tucatinib-combination group compared to the placebo-combination group Table 11. An improvement in PFS was also observed in patients without brain metastases (exploratory only; nominal p value<0.001). The effect size was

slightly greater in the subgroup of patients with brain metastases (52% reduction in progression or death) versus those without brain metastases (43% reduction), however, a formal comparison between these two groups was not intended or performed.

Table 11 PFS by BICR in patients with and without brain metastases

Outcome	With brain metastases[a] (key secondary endpoint)		Without brain metastases[a] (exploratory endpoint)	
	Tucatinib - combination (n=198)	Placebo-combination (n=93)	Tucatinib-combination (n=211)	Placebo-combination (n=108)
No. of PFS events (%)	106 (53.5)	51 (54.8)	91 (43.1)	60 (55.6)
Median PFS (95% CI)	7.6 (6.2, 9.5)	5.4 (4.1, 5.7)	9.6 (7.6, 12.4)	6.8 (4.3, 9.3)
HR (95% CI)	0.48 (0.34, 0.69)		0.57 (0.41, 0.80)	
p-value	p<0.001		Nominal p<0.001	
Table drawn by ERG. Source: CS sections B.2.6.3 and B.2.6.4				

[a] [redacted]

Results for PFS per BICR and OS for pre-specified subgroups

The effects of the tucatinib combination on PFS by BICR in the primary endpoint population and OS in the total population across pre-specified subgroups were generally consistent with the overall treatment effect for these outcomes (CS Figures 12 and 13). The 95% confidence intervals crossed the line of no effect indicating no evidence of a difference in PFS/OS between treatment arms for some subgroups . Slight differences in point estimates were observed for some subgroups, however, these should be interpreted with caution as these analyses were exploratory and no formal tests for interaction were performed.

3.2.5.6 Safety outcomes

3.2.5.6.1 Treatment exposure

Treatment exposure was assessed in all randomised patients who received at least one dose of study treatment in the primary endpoint safety population (n=474) and in the total safety population (n=601). The median duration of exposure for the tucatinib-combination group was 7.3 months and 4.4 months for the placebo-combination group in the primary endpoint safety population (CS Table 11).

In the total safety population, the median duration of treatment was 5.8 months for the tucatinib component of the intervention combination (5.7 months for capecitabine, 6.0 months for trastuzumab) and 4.4 months for the placebo component of the control combination (4.4 months for capecitabine, 4.6 months for trastuzumab).

In the total population (n=612), 118 (28.8%) patients in the tucatinib-combination group and 27 (13.4%) patients in the placebo-combination group were still receiving treatment at the data cut off.

3.2.5.6.2 Treatment emergent adverse events

Adverse events were assessed in all randomised patients who received at least one dose of study treatment (n=601). Treatment emergent adverse events (TEAEs) were defined as events that were new or worsened on or after receiving the first dose of study treatment (tucatinib/placebo, capecitabine or trastuzumab) and up to 30 days after the last dose of study treatment. Rates of TEAEs, serious TEAEs and discontinuations due to TEAEs were comparable between trial arms (Table 12). A slightly higher proportion of patients experienced ≥Grade 3 severity TEAEs in the tucatinib-combination group (55.2%) compared to the placebo-combination group (48.7%).

CS Table 12 provides additional detail regarding dose modifications due to adverse events including doses withheld and dose reductions. Rates of dose modifications were generally higher for tucatinib (54.5%), capecitabine (77.5%) and trastuzumab (25.7%) in the tucatinib-combination group compared to the respective components of the placebo-combination group (41.1%, 61.9% and 19.3%).

Table 12 Summary of TEAEs (safety analysis population)

Adverse event, n (%)	**Tucatinib-combination (N=404)**	**Placebo-combination (N=197)**
Any TEAE	401 (99.3)	191 (97.0)
TEAE leading to discontinuation of:		
Any study treatment	45 (11.1)	19 (9.6)
Tucatinib/placebo	23 (5.7)	6 (3.0)
Capecitabine	41 (10.1)	18 (9.1)
Trastuzumab	18 (4.5)	5 (2.5)
Grade ≥3 TEAE	223 (55.2)	96 (48.7)

Adverse event, n (%)	Tucatinib-combination (N=404)	Placebo-combination (N=197)
Any TE serious adverse events	104 (25.7)	53 (26.9)
TEAE leading to death	8 (2.0)	6 (3.0)
Table drawn by ERG. Source: CS Tables 12 and 13		

TE, treatment-emergent; TEAE, treatment-emergent adverse event

The most frequently reported TEAEs in patients in the tucatinib-combination group were diarrhoea, hand-foot syndrome, nausea, fatigue and vomiting (Table 13). Most of these events were of Grade 1 or 2 severity.

Table 13 Most common (≥20% in the tucatinib-combination) adverse events (safety analysis population)

	Tucatinib-combination (N=404)		Placebo-combination (N=197)	
Adverse event	Any (N, %)	Grade ≥3 (N, %)	Any (N, %)	Grade ≥3 (N, %)
Diarrhoea	327 (80.9)	52 (12.9)	105 (53.3)	17 (8.6)
Hand-foot/PPE syndrome	256 (63.4)	53 (13.1)	104 (52.8)	18 (9.1)
Nausea	236 (58.4)	15 (3.7)	86 (43.7)	6 (3.0)
Fatigue	182 (45.0)	19 (4.7)	85 (43.1)	8 (4.1)
Vomiting	145 (35.9)	12 (3.0)	50 (25.4)	7 (3.6)
Stomatitis	103 (25.5)	10 (2.5)	28 (14.2)	1 (0.5)
Decreased appetite	100 (24.8)	2 (0.5)	39 (19.8)	0
Headache	87 (21.5)	2 (0.5)	40 (20.3)	3 (1.5)
AST increased	86 (21.3)	18 (4.5)	22 (11.2)	1 (0.5)
ALT increased	81 (20.0)	22 (5.4)	13 (6.6)	1 (0.5)

Source: CS Table 14

AE, adverse event; ALT, alanine transaminase; AST, aspartate transaminase; PPE, palmar-plantar erythrodysesthesia

3.2.5.6.3 Adverse events of special interest

CS Appendix F, Table 1 describes the frequencies of prespecified adverse events of special interest including diarrhoea, elevations in liver enzymes, elevations in serum creatinine, cerebral oedema and left ventricular dysfunction. Events of diarrhoea and elevations in liver enzymes were more commonly reported in the tucatinib-combination group, however, most events were <Grade 3 severity, were managed with supportive care and/or dose

modification and ≤1% of patients discontinued treatment due to these events. Elevations in serum creatinine were also more common in the tucatinib-combination group but were reversible and no patients discontinued therapy due to these events. Left ventricular systolic dysfunction leading to dose modification or discontinuation and cerebral oedema events were infrequent (≤2% of patients) with no cerebral oedema events reported in the tucatinib-combination group.

3.2.5.6.4 Use of adverse event data in the economic model

The frequencies of grade 3/4 severity TEAEs reported in ≥2% of patients in the HER2CLIMB trial are used in the company's base case economic model to calculate costs and health resources. Trial-derived utilities are assumed to capture utility loss due to adverse events (CS B.3.4.4). Further details on the sources of adverse event data for the company's chosen comparators are described in Section 4.2.8.

3.2.6 Pairwise meta-analysis of intervention studies

As only one clinical trial of tucatinib was reported in the CS, a pairwise meta-analysis of clinical effectiveness studies was not possible.

3.3 Critique of studies included in the network meta-analysis (NMA)

3.3.1 Rationale for indirect comparisons

The HER2CLIMB trial provides a direct comparison between tucatinib in combination with trastuzumab and capecitabine versus trastuzumab and capecitabine (plus placebo). The control arm of the pivotal HER2CLIMB trial is not, however, a relevant comparator in the NICE scope for this appraisal and the decision problem. The NICE scope states eribulin, capecitabine or vinorelbine as relevant comparators at third-line treatment of patients with HER2+ locally advanced or metastatic breast cancer. Therefore, in the absence of direct head-to-head comparisons the company conducted a network meta-analysis (NMA) to provide indirect comparisons between tucatinib and these treatments. Hazard ratios for PFS and OS analyses from these indirect comparisons directly inform estimates of clinical effectiveness used in the economic model.

3.3.2 Identification, selection and feasibility assessment of studies for NMA

The company's SLR of clinical effectiveness studies (section 3.1 above) was used to inform HTA reimbursement submissions across multiple national markets including England (clarification response A2). The SLR therefore includes a broader range of treatments than

those currently licensed for use in the UK for locally advanced or metastatic HER2+ breast cancer.

The SLR identified [redacted] of which [redacted] were included in the evidence network. Reasons for exclusion of studies from the NMA are listed in CS Appendix D Table 14.

[redacted] however, in response to an ERG query, the company clarified that these five studies do not provide additional connectivity for the comparators relevant to the decision problem and due to their positioning as "terminal" nodes in the evidence network. Therefore, they do not influence the relative effect estimates of the relevant intervention and comparator treatments in the NMA (i.e. the tucatinib combination; eribulin; vinorelbine; and capecitabine) (clarification response A3). The ERG is satisfied that these five studies do not affect comparisons relevant to the decision problem. Thus, the NMA includes seven "relevant" trials (HER2CLIMB, CEREBEL, ELTOP, GBG26, EGF100151, Study 301, NCT02225470) that provide direct or indirect comparisons of interest. Henceforth, we have therefore confined our review to these seven studies (Figure 4).

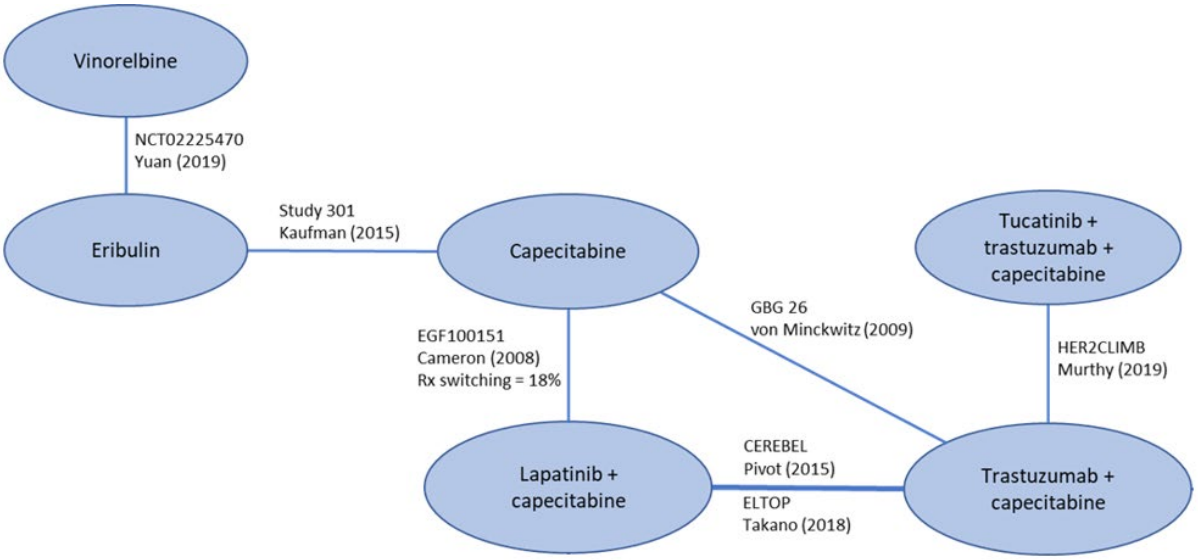


Figure 4 Company evidence network with seven studies relevant to the decision problem

Adapted from CS Figure 14

All seven studies were included in the NMA for the PFS outcome and six of the seven studies were included in the NMA for the OS outcome (Table 14). The GBG 26 study was excluded by the company from the NMA for OS [redacted]. At the

ERG's request the company reported a scenario analysis which retained the GBG 26 study using the unadjusted OS HR (clarification response A17). (see section 3.4.1)

Table 14 Studies contributing to NMA for PFS and OS outcomes

Study identifier	PFS	OS
HER2CLIMB[4]	Yes	Yes
ELTOP[10]	Yes	Yes
CEREBEL[11]	Yes	Yes
GBG 26[12]	Yes	No
EGF100151[13]	Yes	Yes
Study 301[3]	Yes	Yes
NCT02225470[14]	Yes	Yes

The primary objective of the SLR was to identify studies in the population of patients with progression after previous treatment with at least two prior anti-HER2 regimens or at least two prior chemotherapy treatments before eribulin therapy in HER2+ patients.

[redacted] to include studies in patients who had received one or more prior anti-HER2 regimens (or one or more chemotherapies prior to eribulin). The ERG considers this approach reasonable given the lack of available trials. This applied to two of the seven studies (EGF100151 and CEREBEL) but the trial publications do not report the actual numbers of prior lines of therapy used by patients in these trials so it is unclear whether these differ substantially from the other trials in the network. The ERG note that this may introduce heterogeneity in the NMA but the impact is uncertain.

A subgroup analysis for patients with brain metastases was not included in the CS. Whilst the HER2CLIMB trial permitted inclusion of participants with active brain metastases, such participants were not generally eligible for inclusion in trials of the comparator treatments. The ERG agrees there is insufficient data are available to perform an NMA subgroup analysis by brain metastases status. Although it was not feasible to conduct a subgroup analysis for patients with brain metastases, the ERG present an exploratory scenario analysis using the HRs from the subgroup of patients without metastases from HER2CLIMB (see section 3.6).

Finally, the ERG's clinical expert considered that the HER2CLIMB trial direct evidence is available for the subgroup of people with brain metastases for a comparison between the

tucatinib combination versus trastuzumab plus capecitabine. As mentioned earlier (section 2.2.3), in practice, trastuzumab plus capecitabine is given to some patients at third line, even though it is not licensed for this indication. However, to reiterate, trastuzumab plus capecitabine is not included as a comparator in the NICE scope for this appraisal.

3.3.3 Assessment of heterogeneity

3.3.3.1 Clinical heterogeneity

The company proposed several prognostic factors in HER2+ metastatic breast cancer (clarification response Table A13) including age, hormone receptor status, presence of visceral disease, ECOG performance score and prior endocrine therapy. Prior treatment with pertuzumab or trastuzumab, the number of previous lines of therapy and presence of brain metastases were identified as potential treatment effect modifiers. The ERG's clinical expert considers the company's proposed effect modifiers to be reasonable, but he also noted that hormone receptor status and presence of visceral disease could be effect modifiers as well as prognostic factors.

Baseline characteristics of the seven studies included in the NMA are compared in CS Table 10 and Appendix D Table 17. The ERG considered several potential sources of heterogeneity include:

- **Brain metastases.** Almost half of patients in HER2CLIMB trial had brain metastases at baseline, including active brain metastases. In the other studies, patients with brain metastases were either excluded altogether, or were included if brain metastases were stable, or represented a smaller proportion of the trial population. The company states that any bias introduced by these differences in trial populations may understate the relative benefit of tucatinib (clarification response A13). This would assume that patients with brain metastases are less likely to benefit from treatment. However, in the HER2CLIMB trial, the OS and PFS HRs for the subgroup of patients with brain metastases were numerically but not statistically significantly more favourable for the tucatinib combination than for those patients without brain metastases (see section 3.2.5.5). It is unclear if this observation is unique to the HER2CLIMB trial or whether this would apply to other treatment comparisons. The ERG clinical expert noted that such patients generally have a much worse prognosis and only rarely respond to chemotherapy alone or in combination with targeted treatments (trastuzumab or lapatinib). The ERG considers that the direction and magnitude of any bias in the NMA is uncertain because it is unclear whether patients with brain metastases are more or less likely to benefit from the tucatinib-combination

or any other treatment than patients without brain metastases. Our NMA scenario analysis (section 3.6) excluding the subgroup of patients with brain metastases from HER2CLIMB resulted in slightly less favourable survival estimates for the tucatinib combination.

- **Previous lines of treatment.** Two studies (EGF100151 and CEREBEL) included patients with one or more prior lines of therapy while most other studies included patients with two or more prior lines of therapy. Time since diagnosis was shorter in the latter study. The CS does not report the number of lines of prior therapy per trial, but mentions that patients in the HER2CLIMB trial were more heavily pre-treated than in the other studies included in the NMA and that many of the studies in the NMA did not include patients previously treated with agents such as pertuzumab as this drug was not approved at the time the studies were conducted (clarification response A13).
- **Previous exposure to anti-HER2 treatment**. The two eribulin studies (Study 301 and NCT02225470) included patients who had previously received chemotherapy only.

  [redacted]

  (Appendix D Table 17).
- **Mixed HER2 status**. Study 301 and NCT02225470 included a mixture of HER2 positive and negative patients while the other studies included HER2 positive patients only. Only the HER2 positive subgroup data were used in the NMA except for the OS outcome in the NCT02225470 trial where only data for the mixed population were available. The ERG notes that only a fifth of patients in this study were HER2+.
- **Performance status**. The proportion of patients with ECOG score of 1 varied from 28% to 98% across the seven studies. Although ECOG score is a likely prognostic factor, it is unclear whether this is also an effect modifier.
- **Race**. Most studies included mainly Caucasian patients, but two studies included only Chinese and Japanese patients respectively (NCT02225470 and ELTOP). It is unclear whether treatment effect may vary by race or ethnicity.

Other potential sources of heterogeneity may have arisen from differences in study methodology:

[redacted]

[redacted]

The company considers that the main source of heterogeneity in the network is differences in prior exposure to specific anti-HER2 therapies. However, they did not consider methods such as covariate adjustment to account for this to be appropriate. The ERG agrees as the number of studies available would likely preclude reliable covariate adjustment. The company state that there was no evidence of high heterogeneity and did not therefore consider it appropriate to exclude studies from the NMA e.g. in sensitivity analyses. The ERG considers this decision appropriate since exclusion of studies such as ELTOP and CEREBEL would remove the closed loop within the network and potentially weaken the NMA. The company did attempt to evaluate heterogeneity by using a random effects model in their NMA scenario analyses. We critique this further in Section 3.4.2.

3.3.3.2 Statistical heterogeneity

CS section B.2.9.1.4 reports an I^2 value of 30.6%, suggesting a moderate level of statistical heterogeneity in the overall network, although Cochran's Q test for heterogeneity was not significant at the 5% level (p=0.237). It is not clear which outcome (PFS or OS) this value refers to and limited details are provided in the CS on the method used to calculate heterogeneity statistics for the overall network. Outcome-specific heterogeneity parameters are reported as follows:

[redacted]).

ERG conclusion on heterogeneity

The ERG considers the differences between trials in the percentage of patients with active brain metastases to be an important potential source of heterogeneity. The HER2CLIMB trial may over-represent these patients while the comparator trials under-represent them. The magnitude and direction of any bias is uncertain, and thus it remains unclear whether the treatment effect from the NMA adequately represents that in a typical UK population.

3.3.4 Similarity of treatment effects between direct and indirect evidence

There was one closed loop for the PFS evidence network comprising three treatments (from four trials): capecitabine, lapatinib plus capecitabine, and trastuzumab plus capecitabine. Node-splitting analysis found no statistically significant inconsistency ($p>0.05$) between the direct and indirect evidence in this loop (CS section B.2.9.1.4), however, the exact p value is not provided in the CS. The company compared the treatment effects within individual pair-wise comparisons for the two trials (ELTOP and CEREBEL) in this loop that compared lapatinib plus capecitabine with trastuzumab plus capecitabine. Point estimates for PFS and OS from these two trials were in opposing directions and although Cochran's Q test did not provide strong evidence of statistical heterogeneity between these two trials for PFS (p[redacted]), some evidence of statistical heterogeneity was evident for OS ([redacted]; assuming a 10% significance level).

[redacted].

3.3.5 Risk of bias assessment for studies included in the NMA

The company assessed the risk of bias for all seven studies included in the NMA using criteria recommended by NICE (CS Appendix D, Tables 22 to 28). This assessment included an appraisal of the method of randomisation, allocation concealment, balance in baseline study characteristics across trial arms and selective reporting of results. The ERG independently appraised the studies using the NICE criteria and our judgements generally concur with those in the CS. We regard these trials as being generally at a low risk of bias, with the main exception being high risk of detection and performance bias due to lack of blinding in most studies

ERG conclusion on the studies included in the NMA

The company have selected appropriate studies for inclusion in the NMA and the ERG is not aware of any other eligible studies missing. The HER2CLIMB trial and comparator trials in the NMA were generally considered to be at low risk of bias. The ERG has concerns regarding evidence of clinical heterogeneity in the NMA which likely arise from differences in patient characteristics between the trials. In particular, the trials vary considerably in the proportion of patients with active brain metastases. It is unclear, however, whether patients with active brain metastases are more likely to benefit from treatment with tucatinib than those without brain metastases, or with stable brain metastases, and therefore whether relative effect

estimates are biased. The company views the differences between trials in the number of previous treatments patients had received as the main source of heterogeneity. However, the company has not formally addressed these concerns using sensitivity analyses or covariate adjustment methods. The ERG considers this to be appropriate given the limited number of studies in the network. Thus, at present, the clinical effectiveness of the tucatinib combination versus the decision problem comparators is uncertain due to unexplained heterogeneity in the NMA.

3.4 Critique of the network meta-analysis (NMA) statistical methods

3.4.1 Trial data inputs to the NMA

The ERG inspected the OS and PFS data values from the seven trials included in the NMA (CS Appendix D Table 18), and note the following issues:

- In some studies, more than one HR is reported per outcome measure, but it is not explicit which HR was used in the NMA model.
- One study (study NCT02225470) was included in the OS network (CS Figure 15) but no HR for OS was reported in CS Appendix D Table 18.
- Conversely, a HR for OS is reported for the GBG 26 trial but this study was excluded from CS Figure 15.

The company resolved the above issues in response to a clarification question by the ERG (clarification response A4a). However, there was one remaining issue with the data which the company did not resolve; the ERG noted the OS HR for the CEREBEL study is reported as 1.18 (95% CI of 0.760, 1.183) in the CS. The company confirmed this was the correct figure, taken from the supplement to the Pivot et al (2015) journal paper[11] (clarification response A4b). However, to the ERG the upper bound CI does not seem plausible and we suggest the correct value should be 1.83.

[redacted] Hence, the ERG contacted the principal author of the trial manuscript to request clarification. Professor Xavier Pivot subsequently confirmed that 1.83 is the correct value and that the published paper contained a typographical error (personal communication, 16/06/21). An updated figure from Professor Pivot is provided in Appendix 9.3. We use the correct value in the ERG's additional NMA analyses, presented below in section 3.6.

All seven studies provided data for the PFS outcome measure, however, only six studies were included in the OS analysis

[redacted]. At the ERG's request the company reported a scenario analysis which retained the GBG 26 study using the unadjusted OS HR (clarification response A17). This analysis resulted in HRs which were less favourable than those obtained in the base case for the tucatinib combination compared to eribulin, capecitabine and vinorelbine in fixed effect and random effects models.

3.4.2 Statistical methods used in the NMA

The company conducted NMA using two contrasting approaches:

1. A **Bayesian hazard ratio (HR) NMA** reporting results for both fixed effect and random effects models. This approach assumes proportional hazards between treatments, and the relative treatment effect is represented by a constant HR.
2. A **fractional polynomial NMA** to account for potential violation of the proportional hazards assumption in some of the included trials. The fractional polynomial analysis generates results which reflect the time course of the log-hazard function and the relative treatment effect is represented by a time-varying HR.

As will be explained, the ERG considers that the proportional hazards assumption cannot necessarily be rejected and therefore we consider the HR NMA is suitable for the decision problem. We focus on the HR NMA approach in the following sub-sections, and provide a brief appraisal of the fractional polynomial NMA in Appendix 9.4.

[redacted]. This approach is in line with a recent published NMA by Paracha et al (2020),,[6] although Paracha et al (2020) used slightly different informative priors from an earlier version of the Turner paper.[16] Paracha et al (2020) favoured random effects to account for between-study heterogeneity.

The company validated the Bayesian HR NMA by repeating the analysis using frequentist methods (CS Figures 19 & 21). It was found that the two methods provided consistent results. The deviance information criterion (DIC) was reported for the Bayesian models to compare relative model fit (CS sections B.2.9.2.1 & B.2.9.2.2). Random effects results were not provided in the CS but were reported in response to an ERG clarification question (A5).

Meta-regression to address heterogeneity was precluded due to the limited number of available studies.

3.4.2.1 Proportional hazards assessment

An examination of log (-log) plots appears to show some potential violation of the proportional hazards assumption (clarification response A7; NMA report sections 6.1.1 and 6.1.2). The ERG asked the company to provide Schoenfeld residuals plots to inform an assessment of proportional hazards (clarification question A7), however these were not provided. ([redacted]).

[redacted]

The ERG concludes that there is insufficient evidence to reject the proportional hazards assumption and therefore the HR NMA is acceptable to estimate relative treatment effects for the tucatinib combination and comparators in this appraisal.

3.4.2.2 External validation

The random effects results of the company's HR NMA were generally consistent with those of the recently published NMA by Paracha et al (2020). (N.B. Paracha et al reported a smaller of network of studies, excluding HER2CLIMB, NCT02225470, and Study 301 but including GBG 26, EGF100151, ELTOP, and CEREBEL). Outcomes for OS and PFS comparing capecitabine, lapatinib + capecitabine, and trastuzumab + capecitabine (i.e., the comparisons common to both NMAs) were similar apart from the OS for capecitabine versus trastuzumab + capecitabine. However, results are similar when compared to the company's scenario analysis including the GBG 26 trial (clarification response A17). The ERG is satisfied with the external validity of the results, albeit we note that both the company and Paracha et al used the OS HR with the typographical error in the upper bound of the CI from Pivot et al [11].

Computer programming code was provided by the company for all NMA models including fractional polynomials (CS Appendix D, Figures 2-8). The trial data input values for the HR NMA provided in CS Appendix D, Tables 18 & 19 are superseded by Table A4a clarification question response). Reconstructed IPD data for use in the fractional polynomial model were provided in clarification response A8. The IPD data formatted to use in with the

JAGS/WinBUGS code were requested by the ERG but was not provided, nor were the initial values used with the code (clarification question A8). The ERG found no issues with the code and are satisfied the analysis is appropriate.

3.4.2.3 Choice between random effects and fixed-effect model

The company used a fixed-effect model for both OS and PFS due to random effects results showing "inconsistencies" in terms of statistically significant results compared with the trial data (sections B.2.9.2.1 & B.2.9.2.2, clarification response A6). Whilst the respective trials reported statistically significant differences between treatments, the random effects NMA model did not. Of note, in the HER2CLIMB trial the tucatinib combination arm showed a statistically significant OS benefit over the control arm ([redacted] which was not observed in the Bayesian random effects results ([redacted]; Clarification responses Figure A5a).

[redacted] (CS Appendix D, pages 83-84). [redacted]

Whilst the Bayesian HR NMA fixed-effect model was used as the base case, the company's scenario analyses investigated the Bayesian HR NMA random effects model, and fixed-effect and random effects fractional polynomials.

In contrast to the company, the ERG favours use of the random effects model with an informative prior; this avoids the over-estimate of uncertainty from using a vague prior in a network with few studies per comparison, and the underestimate of uncertainty from using a fixed-effect model due to between-study heterogeneity.

3.4.3 Summary of ERG critique of the NMA

- The company performed NMA based on a comprehensive SLR of clinical effectiveness studies. The ERG has no concerns that relevant studies are missing from the SLR and hence the NMA.
- Two contrasting approaches were used: a Bayesian constant hazards NMA, based on the assumption of proportional hazards between treatment comparisons; and a fractional polynomial NMA with time-varying hazards, based on the assumption that the proportional hazards assumption may not hold in all studies. The ERG's view is

there is insufficient evidence to reject the proportional hazards assumption, and the Bayesian HR NMA is more appropriate than the fractional polynomial NMA.

- Potential violations of proportional hazards are most likely to affect studies in the wider evidence network, rather than the seven trials directly relevant to the decision problem. Hence, the HR NMA is acceptable as a source of comparative evidence in this appraisal.
- The company chose the fixed effect model for their NMA base case. As discussed earlier in section 3.3.3, we have concerns about the unexplained heterogeneity in the evidence network with regard to likely effect modifying specific patient characteristics at baseline which are likely to be effect modifying (e.g. presence of active brain metastases, and number of previous treatment lines received). A heterogeneous evidence network is incompatible with the assumption of a fixed-effect model (i.e. that all trial effect estimates are estimating the same underlying intervention effect). We therefore consider a random effects model is more appropriate in this instance as it takes into account heterogeneity, though it does not remove it.
- The upper bound of the CI containing the OS HR in the journal publication of the trial by Pivot et al (2015) contains a confirmed typographical error. Hence, the narrow CI used by the company gives this study a disproportionately higher weight in the company's base case fixed-effect HR NMA. The ERG has corrected this error in our NMA analysis (section 3.6).

3.5 Results of the indirect comparison

In the fixed effect analysis (company base case), the [redacted] (Table 15). In the random effects (company scenario), [redacted] (Table 16).

Table 15 Company Bayesian HR NMA - fixed effect results (base case)

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

[redacted]

Source: reproduced from CS p45, document B

Table 16 Company Bayesian HR NMA – random effects results (scenario)

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

[redacted]

Source: Reproduced from Figures A5a, A5c clarification responses

3.6 Additional work on clinical effectiveness undertaken by the ERG

The ERG conducted the following additional analyses based on the company's NMA:

1. Correction of the Pivot et al (2015) study upper bound CI in the OS HR NMA analysis (ERG "base case")
2. Use of an alternative informative prior for PFS (random effects only)
3. Use of the HRs from the subgroup of patients without brain metastases in the HER2CLIMB trial.

The ERG used the same number of burn-in and iterations as the company for the fixed effect and random effects models. Both models were thinned, although on inspection of autocorrelation plots we considered less thinning of consecutive samples was required (we thinned observations by a factor of 10 in the fixed effect, and 50 in random effects).

3.6.1 Correction of the Pivot et al (2015) study upper bound CI for OS

As noted above, the company used the incorrect upper CI bound for the OS HR [1.18 (95% CI 0.76, 1.183)] reported by Pivot et al (2015). We use the corrected upper bound [1.18 (95% CI 0.76, 1.83)]. As stated above, a narrower CI gives relatively more weight to the study in a fixed effect meta-analysis (the company base case). Table 17 shows the impact of using the corrected figure: in the fixed-effect NMA, the tucatinib combination is no longer statistically significantly better than eribulin and vinorelbine. The ERG prefers the random effects due to heterogeneity between studies as noted above.

Table 17 ERG Bayesian HR NMA – OS scenarios (corrected Pivot et al HR)

Tucatinib combination versus	Fixed effect OS HR (95% CrI)	Random effects OS HR (95% CrI)
Eribulin	0.53 (0.26, 1.06)	0.55 (0.17, 2.05)
Capecitabine	0.51 (0.28, 0.94)	0.53 (0.19, 1.63)
Vinorelbine	0.55 (0.26, 1.14)	0.57 (0.15, 2.42)

[redacted]

3.6.2 Use of an alternative informative prior for PFS (random effects only)

The company chose the "Cause-specific mortality/major morbidity event/composite" outcome from Turner 2015[15] to define the informative prior for PFS [Lognormal (μ=-3.95, σ=1.79^2)] where μ and σ are the mean and standard deviation on the log scale. Previous NICE appraisals have used an alternative category "Internal/external structure-related outcomes" [Lognormal (μ=-2.94, σ=1.79^2)] to represent PFS (e.g. NICE TA492, TA584). The ERG thus considered this alternative informative prior as a scenario for random effects PFS. Results were similar compared to the company model with a slightly wider 95% credible interval (Table 18).

Table 18 ERG Bayesian HR NMA – random effects results (alternative informative PFS prior)

Tucatinib combination versus	PFS HR (95% CrI)
Eribulin	0.25 (0.08, 0.77)
Capecitabine	0.34 (0.14, 0.82)
Vinorelbine	0.23 (0.06, 0.91)

CrI, credible interval

3.6.3 Subgroup analysis of patients without brain metastases in the HER2CLIMB trial

We conducted a scenario using the HRs from the subgroup of patients without brain metastases from the HER2CLIMB trial. The HR for PFS in patients without brain metastases is provided in CS B.2.6.4 (0.57, 95% CI 0.41, 0.80) and the HR for OS is provided in CS B.2.7.2 (0.72, 95% CI 0.48, 1.08). A caveat to this analysis is that some of the comparator studies reported small proportions of patients with brain metastases at baseline, but they did not report outcomes separately by brain metastases status. A further caveat is that PFS in patients without brain metastases in HER2CLIMB is an exploratory endpoint with nominal statistical significance testing. Table 19 shows there is no change in statistical significance in this subgroup but the HRs are less favourable for the tucatinib combination versus comparators than those based on the whole trial population. This is to be expected as patients in HER2CLIMB with brain metastases had numerically (but not

statistically) better PFS HRs than patients without brain metastases (see section 3.6.3 of this report).

Table 19 ERG Bayesian HR NMA – Random effects scenarios (HRs from HER2CLIMB in patients without brain metastases)

Tucatinib combination versus	OS HR (95% CrI)	PFS HR (95% CrI)
Eribulin	0.60 (0.17, 2.36)	0.26 (0.11, 0.66)
Capecitabine	0.58 (0.20, 1.91)	0.36 (0.17, 0.73)
Vinorelbine	0.62 (0.15, 2.80)	0.25 (0.08, 0.75)

Note: the OS analysis includes the Pivot et al correction

Although the NCT02225470 trial comprises a mixed HER2+/- population, we did not deem it necessary to conduct a scenario analysis excluding this study. NCT02225470 is peripheral to the network in that it only serves to connect vinorelbine to the network via eribulin (Figure 4). Hence, its omission would only serve to remove the indirect comparison with vinorelbine and would not impact any of the other treatment effects.

4 COST EFFECTIVENESS

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

The company conducted a combined systematic literature search to identify published economic evaluations, utilities and resource use or cost data for locally advanced unresectable, or metastatic HER2+ breast cancer with progression after previous treatment. The search reported in CS section B.3.1 and Appendix G was conducted in November 2018, and updated in March 2021, as reported in an addendum submitted with the company's response to clarification questions.[18] The reporting of the search strategies and results was clear. Results are presented in CS section B.3.1 and CS Appendix G for economic evaluations; CS section B.3.4.3 and Appendix H for the review of utilities; and CS section B.3.5 and Appendix I for the review of resource use and cost data. The clarification addendum summarises results from the original and updated reviews.

Sixteen economic evaluations were included in the original review and an additional five in the update. None these evaluations address the current decision problem. One recent abstract reported a US cost-utility analysis comparing trastuzumab deruxtecan with tucatinib combination treatment (Vondeling et al. 2020), and another abstract reported a US budget

impact analysis comparing neratinib, lapatinib and tucatinib combination therapies (Anderson et al. 2020).[19 20]

The search identified four UK evaluations of other treatments for the population of interest, including: the NICE (TA458) and Scottish Medicines Consortium (SMC) appraisals of T-DM1; and the SMC cost comparison of intravenous and subcutaneous trastuzumab.[21-23]The company use TA458 to inform decisions about the economic evaluation (CS B.3.3.1 Table 18) and health state costs (CS B.3.5.2). It is not clear why the NICE appraisal of eribulin (TA423)[24] is not included in the company's review of economic evaluations. However, it is referred to as a source of information in relation to their economic evaluation (CS B.3.3.1 Table 18), utilities (CS B.3.4), and the cost of adverse events (B.3.5.3).

ERG conclusion

The company's search strategy and eligibility criteria for their review of cost-effectiveness studies are appropriate. The search did not identify any economic evaluations that directly address the decision problem.

4.2 ERG summary and critique of the company's economic evaluation

4.2.1 NICE reference case checklist

The ERG assessed the company's economic evaluation against NICE Reference Case requirements as shown in Table 20.

Table 20 NICE reference case

Issue	Reference case	ERG comment
Perspective on outcomes	All direct health effects, whether for patients or, when relevant, carers	Yes
Perspective on costs	NHS and PSS	Yes
Type of economic evaluation	Cost–utility analysis with fully incremental analysis	The company only reported pairwise comparisons, but their model includes a 'multiway analysis' function that facilitates full incremental analysis.

Issue	Reference case	ERG comment
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	Yes. The base case model has a time horizon of 20 years (from 54 to 74 years of age). The company state that they explored a time horizon of 30 years (CS B.3.2), but this is not reported, and the model is limited to a maximum of 20 years. This is not an important omission, given survival predictions for the population of interest (modelled 10-year survival is less than 1%).
Synthesis of evidence on health effects	Based on systematic review	Yes. The company use results from their systematic review and NMA to model survival outcomes (CS B.2.9 & B.3.3.4).
Measuring and valuing health effects	Health effects should be expressed in QALYs. The EQ-5D is the preferred measure of health-related quality of life in adults.	Yes. The model estimates QALYs. Utilities for the tucatinib combination is derived from HER2CLIMB EQ-5D-5L data. Utilities for comparators are based on values used in NICE TA423[24] (CS section B.3.4.10).
Source of data for measurement of health-related quality of life	Reported directly by patients and/or carers	Yes
Source of preference data for valuation of changes in health-related quality of life	Representative sample of the UK population	Yes. HER2CLIMB EQ-5D-5L data are valued using the van Hout crosswalk algorithm with the UK value set (CS section B.3.4.1 to B.3.4.2).[25] Comparator utilities from TA423 are derived from a mapping to UK EQ-5D-3L values and direct elicitation from a UK general population sample.[1 2]

Issue	Reference case	ERG comment
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Yes
Evidence on resource use and costs	Costs should relate to NHS and PSS resources and should be valued using the prices relevant to the NHS and PSS	Yes
Discounting	The same annual rate for both costs and health effects (currently 3.5%)	Yes
PSS, personal social services; QALYs, quality-adjusted life years; EQ-5D, standardised instrument for use as a measure of health outcome.		

4.2.2 Model structure

The company describe the structure and key features of their model in CS section B.3.3. Assumptions are summarised in CS Tables 17-19 and 33; and parameters in CS sections B.3.3 to 3.5, with an overview in CS Table 33. The model uses a partitioned survival structure, with a cycle length of 1 week and 20-year time horizon. No half-cycle correction was incorporated due to the short cycle length. Costs and QALYs are discounted at 3.5% per year. The model consists of three 'partitioned survival' health states, as illustrated in Figure 5:

- *Progression-free (PF):* the proportion of patients alive and progression-free (PFS)
- *Progressed disease (PD):* the proportion of patients alive (OS) minus the proportion of patients alive and progression-free (PFS)
- *Death:* calculated as one minus the proportion of patients alive (OS)

Patients enter the model in the progression-free state and can experience disease progression or death. While in the progressed disease state, patients receive subsequent lines of anticancer therapy and supportive care.

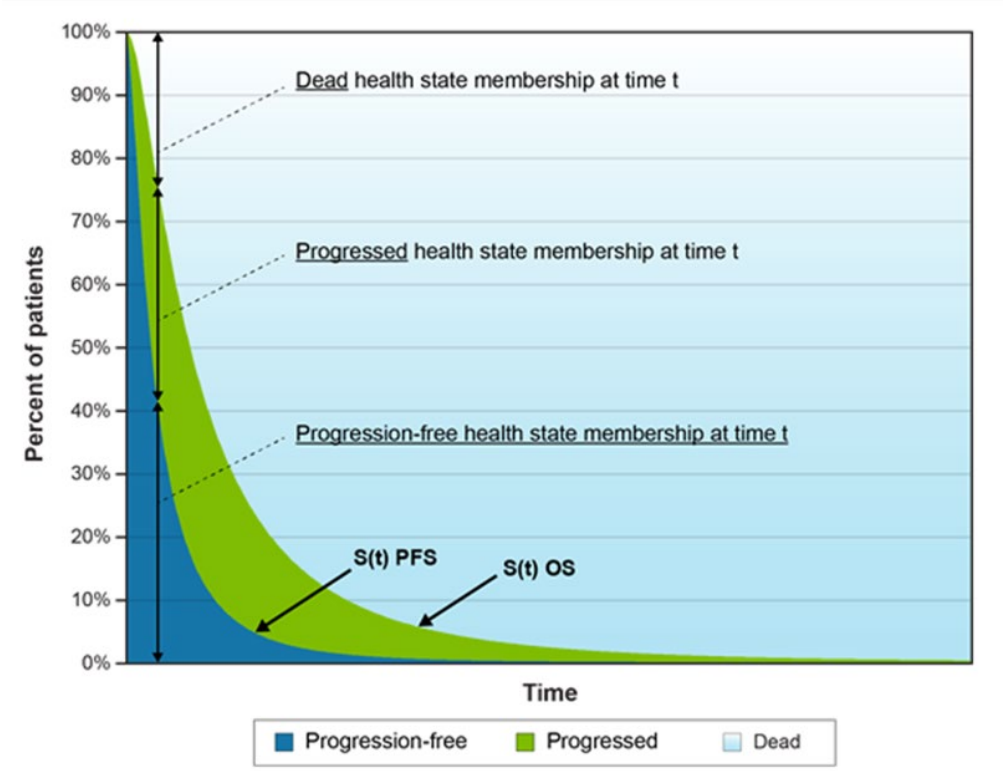


Figure 5 Partitioned survival model structure
Source: reproduced from CS Figure 22

ERG conclusion

The model structure is appropriate and accurately implemented. We agree with the partitioned survival approach.

4.2.3 Population

The company model a population of adults with HER2+ metastatic breast cancer who have received two or more prior anti-HER2 regimens (CS section B.3.2.1). Baseline characteristics of the modelled cohort are based on those of patients in HER2CLIMB: mean age 54 years; mean body surface area 1.8 m^2; mean body weight 69.5 kgs (CS Table 33).

The HER2CLIMB trial included patients with presence or a history brain of metastases: nearly 50% of the total study population (CS Table 6). The NICE scope requested subgroup analysis for patients with brain metastases if the evidence allows. The company did not attempt to model cost-effectiveness for this subgroup, on the basis that there is a lack of clinical evidence for the scope comparators in this subgroup. See 3.3.3 above for a discussion of this and other differences in patient characteristics between HER2CLIMB and other trials included in the NMA.

ERG conclusions

- The modelled population is consistent with the licensed indication for the tucatinib combination and that specified in the NICE scope.
- It is not clear whether the HER2CLIMB trial (which included a high proportion of patients with brain metastases) or other trials in the NMA (which largely excluded patients with brain metastases) provide a more realistic reflection of the target population in routine practice. This heterogeneity has implications for survival modelling because OS and PFS extrapolations anchored by curves fitted to the HER2CLIMB data differ from those based on fractional polynomial curves estimated from the NMA (see section 4.2.6).
- Modelling cost-effectiveness for the subgroup of patients with brain metastases is problematic, due to the lack of evidence for the comparators in this subgroup. However, omitting this subgroup analysis does not negate the need for estimates of relative treatment effectiveness across this heterogeneous evidence base.

4.2.4 Interventions and comparators

The company describe the intervention and comparators in their decision problem in CS section B.1.2. The submitted model includes the tucatinib combination and scope comparators (eribulin, capecitabine and vinorelbine). It also includes trastuzumab with capecitabine, which can be used as a reference arm to link survival extrapolations for the tucatinib combination with those for comparators in the network of clinical evidence. The company focus on the comparison with eribulin in their economic analysis (CS B.3.7 to B.3.9). They report scenarios with pairwise ICERs for the tucatinib combination compared with capecitabine and with vinorelbine (CS Tables 40 and 41), but do not explore the sensitivity of these results, or report a full incremental analysis.

We note that trastuzumab + capecitabine is not licensed or recommended by NICE for the population of interest and was not included in the NICE scope. Expert advice to the ERG is that clinical practice varies, and that some NHS centres do use trastuzumab with capecitabine at third line. It is not clear if this constitutes routine practice, or therefore whether it should be included as a comparator in the economic analysis.

ERG conclusion

The company only report pairwise cost-effectiveness estimates, rather than a full incremental analysis as recommended by NICE. In ERG analysis, we present full incremental results for the tucatinib combination against all of the scope comparators

(see sections 6.3 and 6.4). We also report incremental results including trastuzumab + capecitabine, as this is used in some NHS centres.

4.2.5 Perspective, time horizon and discounting

The company uses a 20-year time horizon and take the perspective of the NHS and PSS in England. Both costs and outcomes (life years and QALYs) are discounted at 3.5%, in line with the NICE guidance.

4.2.6 Survival analyses

The company uses data from the HER2CLIMB study as well as the NMA to produce long term extrapolations of PFS and OS (CS B.3.3.4). Three main sets of survival parameters are included in the model and are discussed below:

1. OS and PFS curves fitted to HER2CLIMB data
2. OS and PFS curves estimated from the fractional polynomial NMA
3. Relative treatment effects from the hazard ratio NMA

4.2.6.1 OS and PFS curves fitted to HER2CLIMB data

Survival curves fitted to HER2CLIMB data for the tucatinib combination and trastuzumab + capecitabine. The company report analysis with 23 survival models fitted to individual patient data from the trial, including:

- Conventional parametric distributions (exponential, Weibull, log-normal, log-logistic, gamma and generalised gamma)
- Flexible spline models (Weibull 1, 2 and 3 knot)
- Stratified versions of parametric and flexible spline models, equivalent to fitting separate models by treatment arm
- Hybrid models combining Kaplan-Meier estimates for an initial period with exponential, Weibull, log-normal or log-logistic extrapolations

Details of how these curves were fitted and the approach used to select the company's base case and scenario models are described in CS B.3.3.4 and Appendix L. See sections 4.2.6.4 and 4.2.6.5 below for ERG commentary on the trial-based OS and PFS extrapolations respectively.

CS Appendix L also reports survival curves fitted to HER2CLIMB data but constrained to give long-term projections that are consistent with an external data source (Kaufman et al. 2015).[3] However, the company chose not to include the survival models with external data in

the economic model, as they considered that the curves based on HER2CLIMB data alone had a good fit and gave plausible extrapolations (see CS Section B.3.3.4).

4.2.6.2 OS and PFS curves estimated from the fractional polynomial NMA

Survival curves are also available for the tucatinib combination, the comparators and other treatments in the network from fixed and random effects fractional polynomial NMA models. The economic model only includes parameter sets for the company's preferred fractional polynomial models: P1=0 (Weibull) for OS and P1=-2 and P2=-0.5 for PFS. The model uses 6,000 draws from the fractional polynomial posterior distributions for the probabilistic analysis and the means of these parameter sets for deterministic analysis.

4.2.6.3 Relative treatment effects from the hazard ratio NMA

The economic model also includes relative treatment effects from the Bayesian HR NMA, fixed and random effect models (see section 3.4 and section 3.5 above). With an assumption of proportional hazards, the HR estimates can be applied to trial-based or fractional polynomial survival curves for a reference treatment to obtain curves for the other comparators. The economic model uses 6,000 correlated sets of HR NMA estimates for probabilistic analysis, and the mean of these parameter sets for deterministic analysis. We note that these means are similar, but not identical to the HR NMA results reported in the CS (section 3.5).

The company's base case uses input parameters described above for OS and PFS. Relative effects from the NMA ([redacted]) are applied to fractional polynomial survival curves for a reference treatment. The company chose lapatinib plus capecitabine as the reference, as it is the most used treatment in the NMA (described in Document B Section CS B.2.9.1.7). Figure 6 below shows the modelled OS and PFS curves for the company's base case, alongside HER2CLIMB KM data.

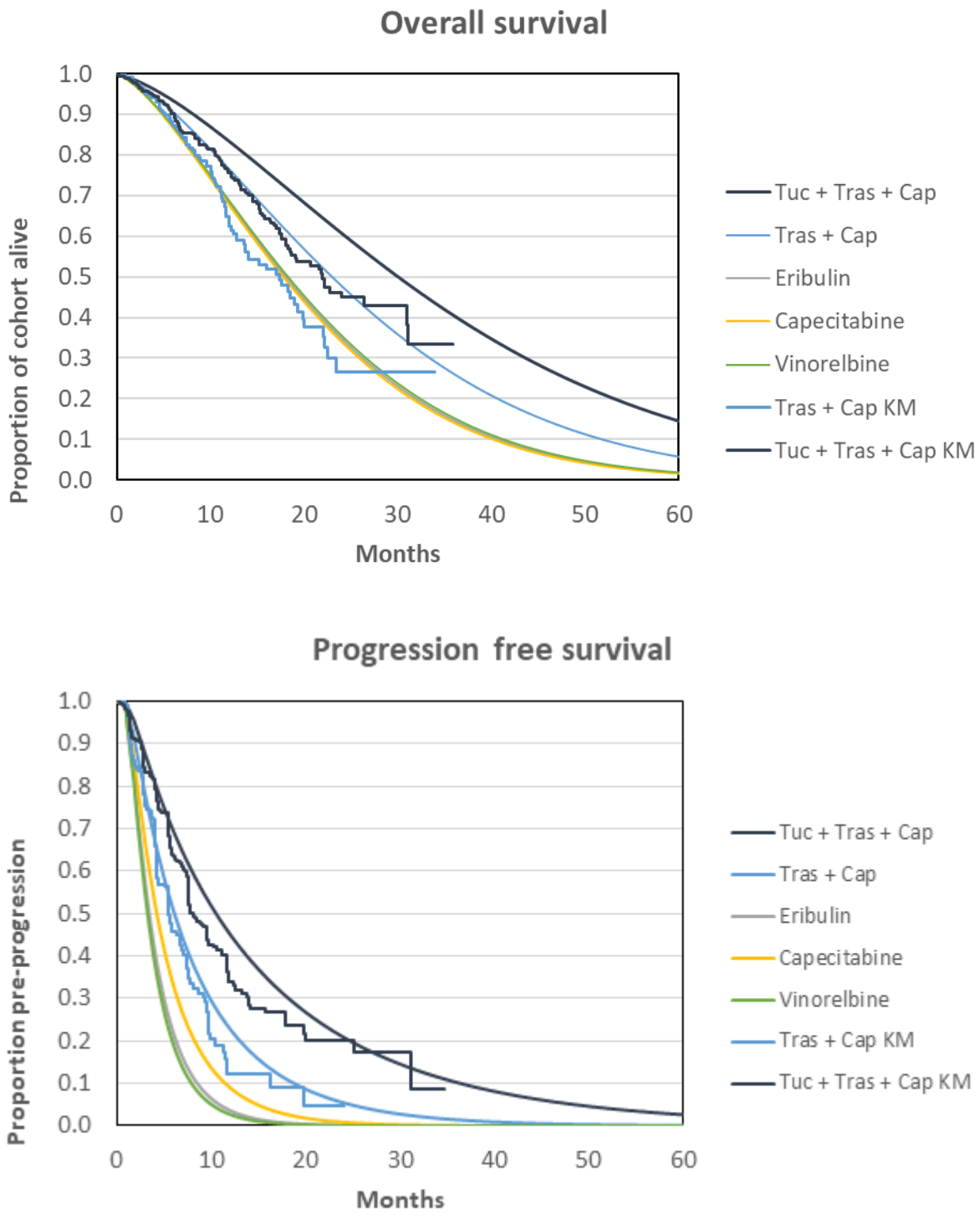


Figure 6 Survival curves from base case model with HER2CLIMB KM data

Source: Obtained from the company's model by the ERG

The company's base case extrapolations for the tucatinib combination have a poor fit to the HER2CLIMB trial data. In particular, modelled OS is much better than that observed in the trial. This is possibly due to differences between the HER2CLIMB population and that of

other trials in the evidence network (section 3.3.3). Nearly half of the patients in HER2CLIMB had brain metastases and one would expect poorer outcomes for this group. The difference in survival expectations between the data sources is potentially important for the economic model because even with a constant HR between treatments, the QALY gain will be greater if absolute survival is higher.

This leads to the question of which data source is more reflective of routine practice. The proportion of patients with brain metastases in HER2CLIMB may be higher than expected, but the zero or very low proportion in the other trials is certainly not representative. We therefore suggest that the survival curves fitted to HER2CLIMB data provide a more valid foundation for the economic analysis than the NMA fractional polynomials. However, it is important to acknowledge that a 'within-trial' model anchored by survival curves fitted to HER2CLIMB data may underestimate survival (and hence QALYs) if patients with a poor prognosis are over-represented in this trial.

ERG conclusions

- The company's base case survival estimates are substantially more favourable than those observed in the HER2CLIMB study. This may be due to the population in this trial, which included more patients with brain metastases than other trials in the NMA (which included few or no patients with brain metastases). This raises a question of which trials are more reflective of routine practice.
- The company's submission does not explore uncertainty over to the choice of reference survival curves for their base case (OS and PFS curves for lapatinib + capecitabine from the fractional polynomial NMA).The model includes an option to use 'within-trial' curves fitted to HER2CLIMB tucatinib and placebo combinations, which are adjusted for indirect comparators with relative effects from the NMA. However, this model was not used in company scenario analysis and was not functioning in the submitted model (see section 6.1).
- In addition, the company did not test the impact of alternative functional forms for PFS and OS in their model. Although a scenario with alternative survival models is reported (CS Table 39), QALY estimates from this scenario do not differ from those in the base case analysis (see section 5.2.2). This is not surprising as the model only includes one fractional polynomial function for OS and one for PFS.
- In ERG analysis, we explore the 'within trial' approach to survival modelling and test alternative scenarios for the PFS and OS survival functions fitted to HER2CLIMB data, and adjusted for comparators with HRs from the NMA (sections 6.3 and 6.4).

In the following sections, we summarise the company's approach to fitting OS and PFS survival functions to HER2CLIMB trial data and consider the choice of curves for ERG scenario analyses.

4.2.6.4 Overall survival for within-trial analysis

The company outline their approach to modelling OS with individual patient data from the HER2CLIMB trial in CS section B.3.3.4, with further details in CS Appendix L. The approach in the Appendix is thorough and consistent with methodological advice from the NICE Decision Support Unit.[26] It also makes use of techniques for flexible curve fitting and integration of longer-term external data to improve the plausibility of extrapolations.

The validity of the proportional hazards assumption between the HER2CLIMB arms is discussed in section 9.2 of CS Appendix L. Log-(log) survival and Schoenfeld residual plots are shown in CS Appendix L Figures 13 and 14, respectively. The log-(log) survival curves [redacted]. Plots of smoothed hazards (CS Appendix L Figure 15) show [redacted]. The rates were [redacted]. The smoothed hazard ratio plot (CS Appendix L Figure 16) [redacted].

The fit and plausibility of the 23 models fitted to HER2CLIMB OS data is discussed in CS Appendix L section 9.3. For convenience, we reproduce illustrations of the fitted models in Figure 7, Figure 8 and Figure 9 below (CS Appendix L Figures 17, 19 and 22 respectively) and a summary of the predicted means in Table 21 (CS Appendix L Table 9).

The company selected 13 models that they considered to have both a good fit to the trial data and plausible extrapolations (CS Document B Table 21). Out of these 13 models, the company chose the Weibull for use in the base case cost-effectiveness analysis. They justify this based on AIC/BIC statistics, visual inspection, external validation against the Kaufman et al. study (2015)[3] and the views of an external advisory board. Although the company state that they conducted a scenario analysis with [redacted] (CS B.3.3.4.2), it appears that this scenario actually used a [redacted] (CS Table 39).

Table 12 in CS Appendix L contains recommendations for the 'most likely', 'optimistic' and 'pessimistic' models based on the difference in mean survival between treatment arms. These were selected from the shortlist of models judged to have a good fit and plausible extrapolations. For OS fitted to HER2CLIMB data alone, this suggested:

- Most likely: Weibull, gamma (difference in mean survival 6.2, 6.4 months)
- Optimistic: Stratified Weibull, stratified gamma (mean difference 8.0, 8.8 months)
- Pessimistic: Gompertz (mean difference 4.8 months)

We note that the company's model does not constrain the mortality rate to be no less than that of people of the same age in the general population. However, given the poor survival of the modelled population, this is not expected to produce unrealistic predictions.

ERG conclusions

- The company's analysis of survival data from the HER2CLIMB trial follows methodological guidance and is thorough and well-reported. [redacted].
- A range of curves had a good visual fit to the trial data and produced plausible extrapolations (mean survival in the control arm of between [redacted] months)
- We agree with the company's choice of Weibull for their base case and consider that the stratified Weibull and Gompertz extrapolations provide an appropriate range of optimistic and pessimistic predictions of the mean difference in survival between the arms (8.8 and 4.8 months respectively). We use these distributions in ERG sensitivity analysis (see Table 31).

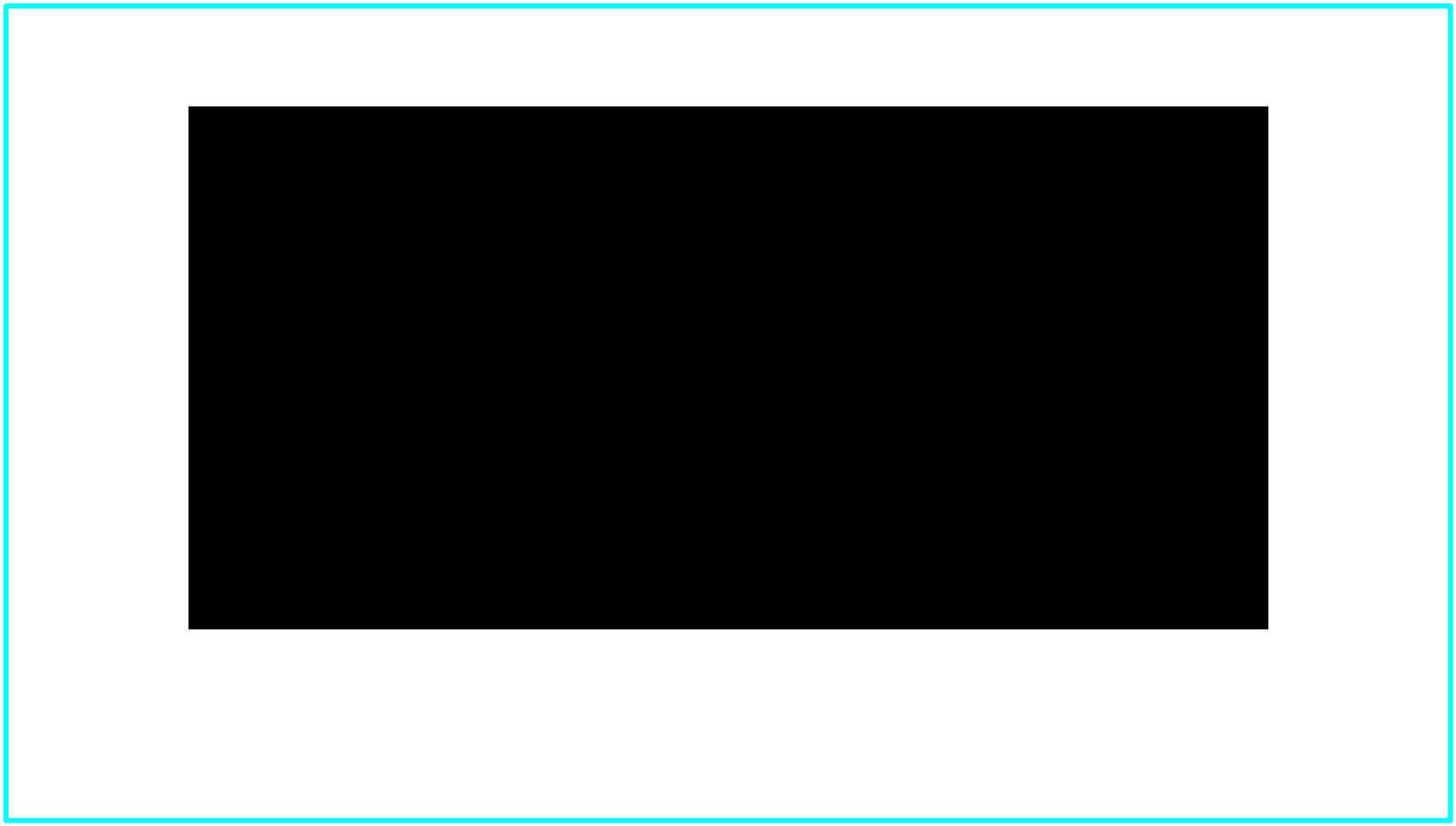


Figure 7 Standard parametric models fitted to the HER2CLIMB trial data: OS [redacted]

Source: Reproduced from CS Appendix L Figure 17

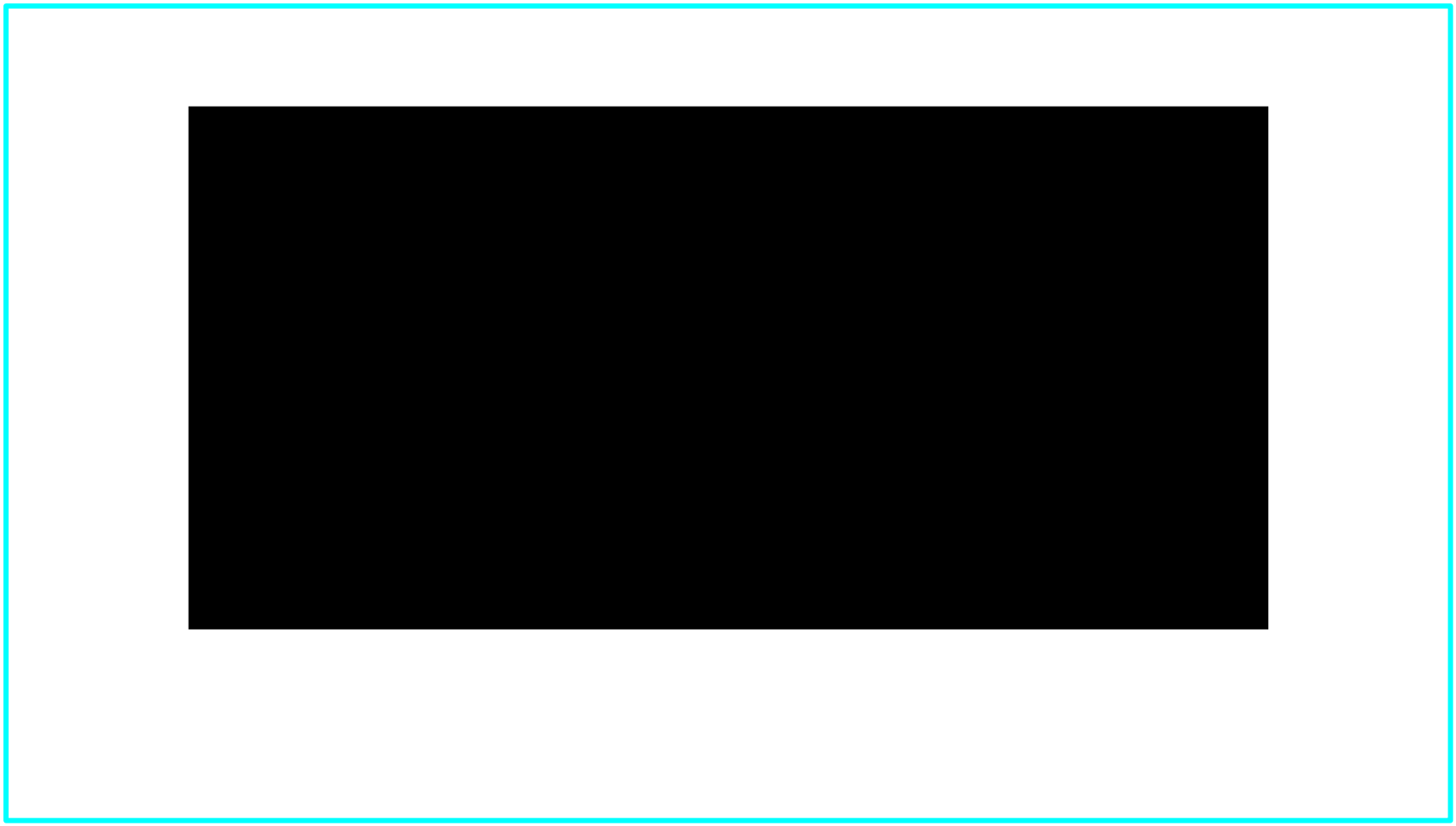


Figure 8 Flexible Spline-based models fitted to the HER2CLIMB trial data: OS [redacted]

Source: Reproduced from CS Appendix L Figure 18

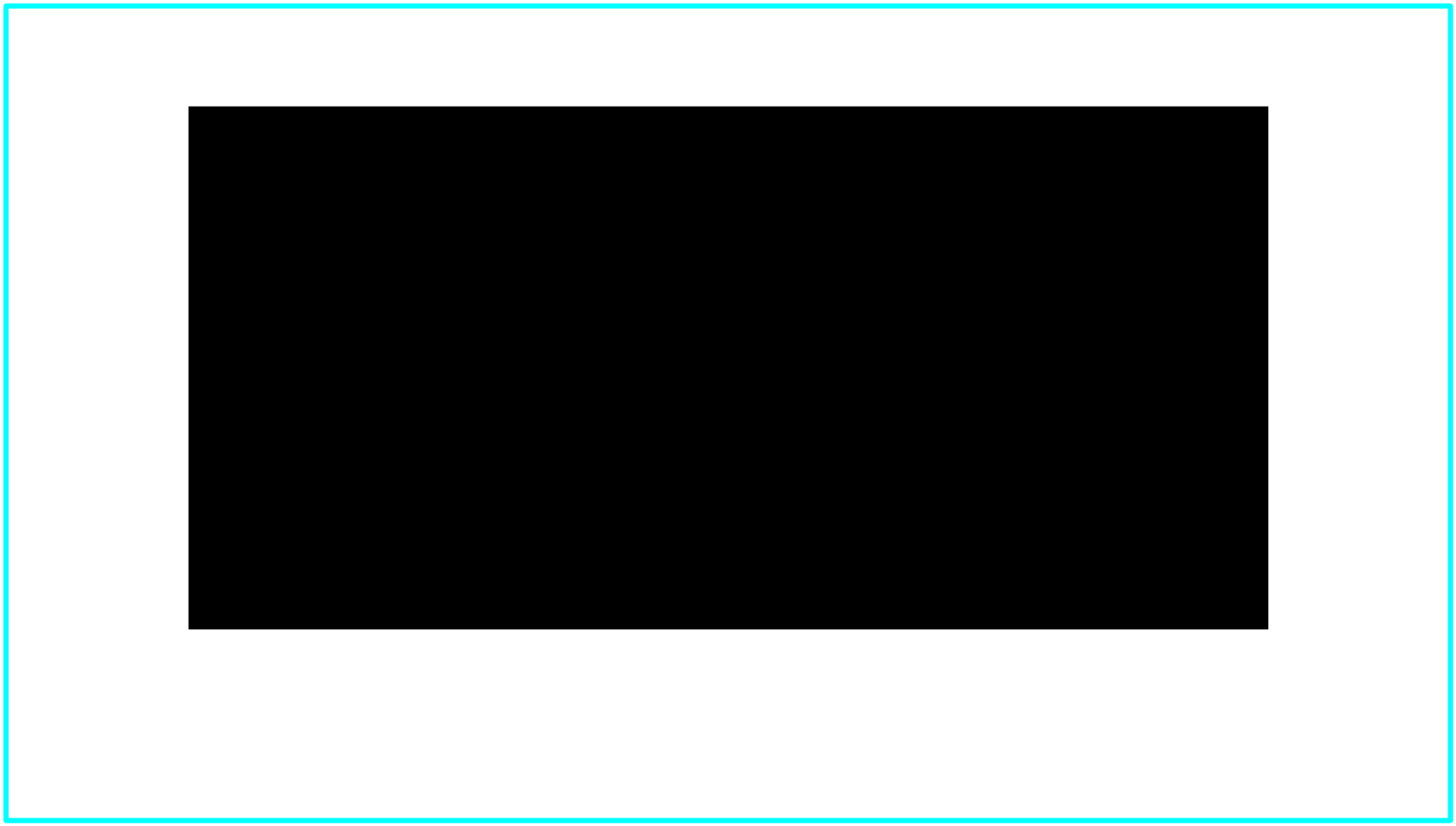


Figure 9 Hybrid survival models fitted to the HER2CLIMB trial data: OS [redacted]

Source: Reproduced from CS Appendix L Figure 22

Table 21 Predicted mean OS times in months for models fitted to the HER2CLIMB data

	Pbo+Tras+Cape			TUC+Tras+Cape			Difference, Months				
Model	Mean	Lower CrI	Upper CrI	Mean	Lower CrI	Upper CrI	Mean	Lower CrI	Upper CrI	Fit to RCT	Plausible
Exponential	[redacted]									Poor	Yes
Weibull	19.9	16.9	23.1	26.0	22.6	6.2	6.2	1.3	10.7	Good	Yes
Stratified Weibull	[redacted]					8.0				Good	Yes
Gompertz	20.7	16.8	29.4	27.4	22.3	4.8	4.8	1.4	8.9	Good	Yes
Stratified Gompertz	[redacted]					5.9				Good	Yes
Log-normal	[redacted]									Poor	No
Stratified log-normal	[redacted]									Poor	No
Log-logistic	[redacted]									Good	No
Stratified log-logistic	[redacted]									Good	No
Gamma	[redacted]					6.4				Good	Yes
Stratified gamma	[redacted]					8.8				Good	Yes
Generalized gamma	[redacted]					6.2				Good	Yes
Stratified generalized gamma	[redacted]					6.5				Good	Yes
Flexible Weibull (1 knot)	[redacted]					6.2				Good	Yes
Flexible Weibull (2 knots)	[redacted]					7.2				Good	Yes
Flexible Weibull (3 knots)	[redacted]									Good	No
Stratified flexible Weibull (1 knot)	[redacted]					7.1				Good	Yes
Stratified flexible Weibull (2 knots)	[redacted]					6.8				Good	Yes
Stratified flexible Weibull (3 knots)	[redacted]									Good	No
Hybrid exponential	[redacted]					7.7				Good	Yes
Hybrid Weibull	[redacted]									Good	No
Hybrid log-normal	[redacted]									Good	No
Hybrid log-logistic	[redacted]									Good	No

Source: Reproduced from CS Appendix L Table 9
Green shading indicates models that the company considered to provide both a good fit to RCT data and plausible extrapolations. Red indicates that the company considered that the model did not meet both of these criteria.

4.2.6.5 Progression-free survival for within-trial analysis

Evidence related to proportional hazards between the HER2CLIMB arms for PFS is shown in CS Appendix L section 10.2. The log-(log(survival)) plots [redacted] and the Schoenfeld test [redacted] (CS Appendix L Figures 40 and 41). Smoothed hazard rates (CS Appendix L Figure 42) indicate that [redacted]. The hazard ratio was [redacted] (CS Appendix L Figure 43).

The company fitted 23 PFS models to the HER2CLIMB data. See Figure 10, Figure 11 and Figure 12 (reproduced from CS Appendix L Figures 44, 46 and 49) and summary of the predicted mean OS times Table 22 (reproduced from CS Appendix L Table 15). The company concluded that seven models have a good fit to trial data and produced plausible extrapolations (CS B.3.3.4.1 and Table 20). These gave predictions of mean PFS between [redacted] months. They selected the flexible Weibull with 2 knots for the base case analysis, as it was in line with results from the models with the selected external data (Kaufman et al. 2015).[3] The company conducted a scenario analysis with the [redacted].

Table 12 in CS Appendix L concludes that the 'most likely', 'optimistic' and 'pessimistic' models from the shortlist with good fit and plausible extrapolations were:

- Most likely: [redacted]
- Optimistic: [redacted]
- Pessimistic: [redacted]

ERG conclusions

- [redacted].
- A range of curves had a good visual fit to the trial data and produced plausible extrapolations (mean PFS in the control arm from [redacted] months)
- For the ERG base case, we use the flexible Weibull 2 knots PFS curve. We include the stratified generalized gamma and stratified log-normal in scenario analyses (mean difference in PFS [redacted] months).

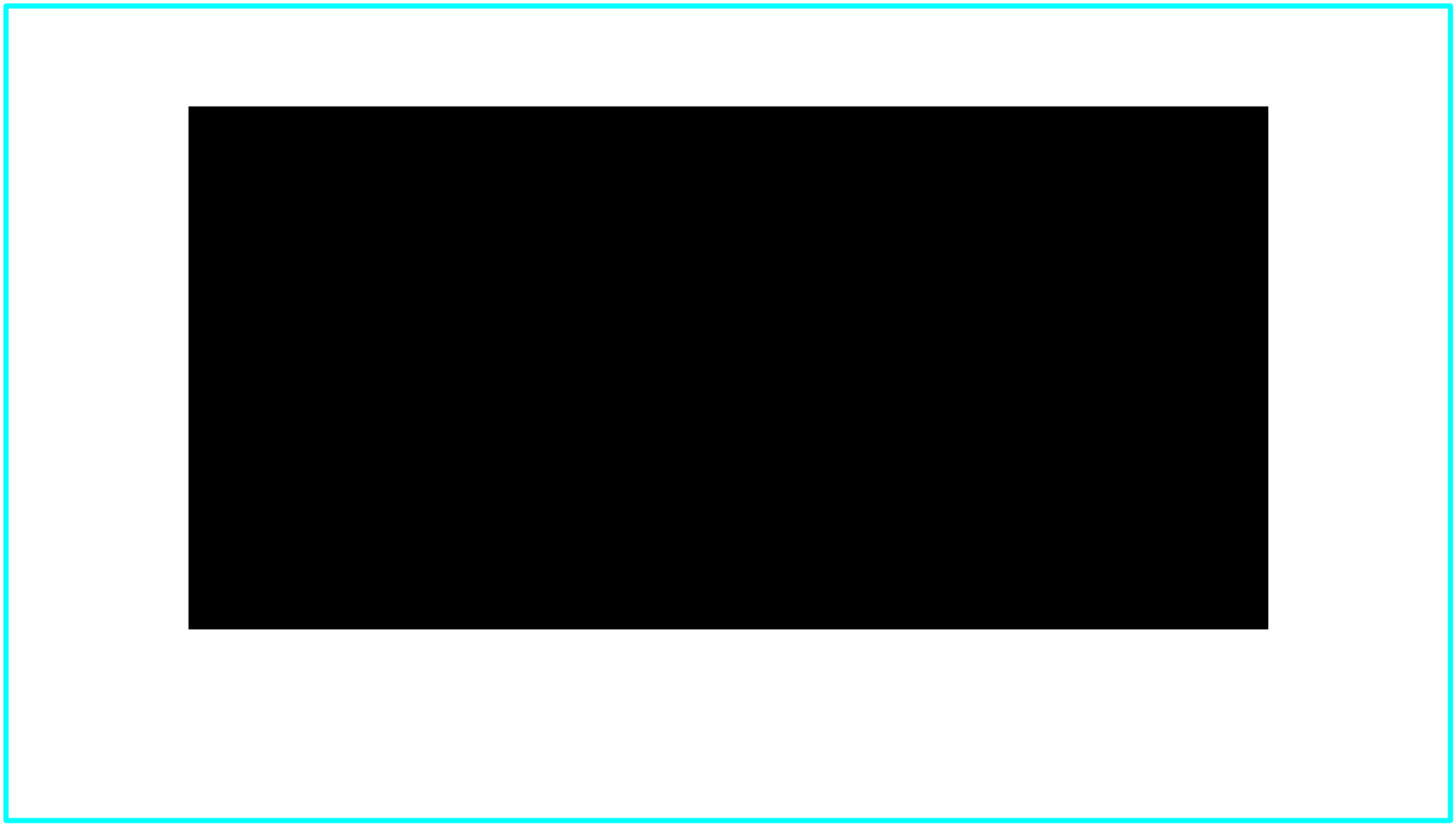


Figure 10 Standard parametric models fitted to the HER2CLIMB trial data: PFS [redacted]

Source: reproduced from CS Appendix L Figure 44

Figure 11 Flexible spline based fitted to the HER2CLIMB trial data: PFS [redacted]

Source: reproduced from CS Appendix L Figure 46

Figure 12 Hybrid survival models fitted to the HER2CLIMB trial data: PFS [redacted]

Source: reproduced from CS Appendix L Figure 49

Table 22 Predicted mean PFS times in months for models fitted to the HER2CLIMB data

	Pbo+Tras+Cape			TUC+Tras+Cape			Difference, Months				
Model	Mean	Lower CrI	Upper CrI	Mean	Lower CrI	Upper CrI	Mean	Lower CrI	Upper CrI	Fit to RCT	Plausible
Exponential	[redacted]									Poor	Yes
Weibull	[redacted]									Poor	Yes
Stratified Weibull	[redacted]									Poor	Yes
Gompertz	[redacted]									Poor	Yes
Stratified Gompertz	[redacted]									Poor	Yes
Log-normal	[redacted]									Poor	Yes
Stratified log-normal	[redacted]									Good	Yes
Log-logistic	[redacted]									Poor	Yes
Stratified log-logistic	[redacted]									Good	Yes
Gamma	[redacted]									Poor	Yes
Stratified gamma	[redacted]									Poor	Yes
Generalized gamma	[redacted]									Poor	Yes
Stratified generalized gamma	[redacted]									Good	Yes
Flexible Weibull (1 knot)	[redacted]									Good	Yes
Flexible Weibull (2 knots)	[redacted]									Good	Yes
Flexible Weibull (3 knots)	[redacted]									Good	No
Stratified flexible Weibull (1 knot)	[redacted]									Good	Yes
Stratified flexible Weibull (2 knots)	[redacted]									Good	No
Stratified flexible Weibull (3 knots)	[redacted]									Good	No
Hybrid exponential	[redacted]									Good	Yes
Hybrid Weibull	[redacted]									Good	No
Hybrid log-normal	[redacted]									Good	No
Hybrid log-logistic	[redacted]									Good	No

Source: Reproduced from CS Appendix L Table 15
Green shading indicates models that the company considered to provide both a good fit to RCT data and plausible extrapolations. Red indicates that the company considered that the model did not meet both of these criteria

4.2.6.6 Waning of treatment effects (HR tapering)

CS Appendix L also explores scenarios using external data and hazard ratio tapering, to reflect the waning of OS and PFS treatment effects after trial follow up. The NMA fractional polynomial data in the model includes estimated times to HR=1: approximately 4 years from the start of treatment for PFS and 6 years for OS. The model includes a function to gradually reduce HRs to 1 between maximum trial follow up and these timepoints. This was not used in the company submission. We report the effect of HR tapering in ERG scenario analysis (see section 6.4 below).

4.2.7 Treatment duration

The company used data from the HER2CLIMB trial to estimate time to treatment discontinuation (TTD) for the tucatinib combination and trastuzumab with capecitabine (CS B.3.3.5 and clarification response B4). A range of parametric and flexible spline models were fitted to the TTD data. All of the models gave a good visual fit to the trial data, except for the log-logistic and log-normal which overestimated treatment continuation towards the end of the trial. The Flexible Weibull with 2 knots was chosen for the base case analysis, as it has the best model fit statistics and aligns with the base case PFS model. We agree that it is reasonable to assume that TTD follows a similar shape to PFS because progression is the primary reason for treatment discontinuation.

TTD data were not available for eribulin, capecitabine and vinorelbine. For the base case, the company assumed constant hazards (exponential model), based on median treatment durations for clinical studies in the NMA (CS B.3.3.7). These sources are not reported in the CS or clarification response, but the model (sheet External_TTD) uses estimated mean treatment durations of:

- 5.61 months for eribulin (Kaufman et al. 2015 and Yuan et al. 2019)[3 14]
- 5.67 months for capecitabine (von Minckwitz et al. 2009 and Kaufman et al. 2015)[3 12]
- 3.98 months for vinorelbine (Yuan et al. 2003)[14]

The model also includes options to assume that TTD is equal to PFS or to limit treatment duration to the median times reported for clinical studies.[12 14 27] The company report results for a scenario with treatment duration restricted to mean exposure (CS Table 39). They do not report the results of assuming that TTD is equal to PFS, but we do in ERG scenario analysis (Table 36 below).

ERG conclusion

- We agree with the use of HER2CLIMB trial data to model treatment duration for the tucatinib combination and for trastuzumab with capecitabine. We also agree with the company's rationale for choosing the same survival function for TTD as for PFS (flexible Weibull with 2 knots). This has a good fit to the trial data and appears plausible (clarification response Figures B4a- B4f).
- The company's approach to modelling treatment duration for the external comparators is also reasonable. This assumes a constant rate of discontinuation, estimated from median treatment durations in clinical studies included in the NMA.
- We report results for a scenario with TTD assumed equal to PFS for all comparators in ERG analysis (see section 6.3 below). This is likely to overestimate treatment duration (and hence costs) for all treatments, as some patients are likely to stop treatment for reasons other than progression.
- The company's scenario in which treatment duration is limited to the median reported from clinical trials is likely to underestimate treatment costs. This will favour the tucatinib combination, which has the highest treatment cost and duration.

4.2.8 Adverse events

The company base case includes costs for treating grade 3 and 4 treatment-emergent adverse events (TEAEs) that occurred in at least 2% of patients for any of the treatment arms in the clinical studies (CS B.3.5.3). The references cited for TEAE incidences in the model are Murthy et al. (2020) for the HER2CLIMB treatments; von Minckwitz 2009 for capecitabine; and Yuan et al. for eribulin and vinorelbine.[4 12 14] Sources for the treatment costs for TEAEs are listed in CS Table 32.

The company assumes that the utility impacts of adverse events are captured in the health state utility weights (CS B.3.4.4). However, the model does include parameters for TEAE disutilities and durations, and QALY losses associated with TEAEs can be included in the model (CS Table 33). This has a negligible impact on cost effectiveness results.

We note that there are some discrepancies between TEAE incidence parameters in the model compared with CS Table 33 and cited sources: e.g. Murthy et al. cite an incidence of 2.5% for anaemia in the trastuzumab + capecitabine control arm but the model uses 0%;

and CS Table 33 cites a 0% incidence of vomiting in this control arm, compared with 3.6% in the model and Murthy at al.[4] These differences do not impact on cost-effectiveness results.

ERG conclusions

The company's approach to modelling adverse events is reasonable. The base case model includes estimated costs for treating adverse events and although the disutilities are not included, they can be added in a scenario analysis. We found some inconsistencies between adverse event incidences in the model and values reported in the CS and the cited sources. This does not impact on cost effectiveness results.

4.2.9 Health related quality of life

The company explain their approach to estimation of utilities in CS section B.2.4 and in their responses to clarification questions B1, B2 and B3. Table 23 below summarises the health state utilities used in the company's original base case economic model and their revised base case after clarification.

Table 23 Utility values in the company's original and revised base case analyses

Treatment	Health state (treatment cycle)	Original base case	Revised base case	Sources
Tucatinib combination	Progression free (1-2)	0.748	0.762	HER2CLIMB EQ-5D (CS Table 23 and clarification response B2)
	Progression free (3-4)	0.763		
	Progression free (5-6)	0.792		
	Progression free (7+)	0.807		
	Post progression	0.653	0.698	
Trastuzumab + capecitabine	Progression free (1-2)	0.770	0.762	
	Progression free (3-4)	0.765		
	Progression free (5-6)	0.741		
	Progression free (7+)	0.748		
	Post progression	0.698	0.698	
Eribulin	Progression free	0.783	0.706	TA423 (CS Table 24 and clarification response B3)
	Post progression	0.622	0.496	
Capecitabine	Progression free	0.691	0.701	
	Post progression	0.651	0.496	
Vinorelbine	Progression free	0.691	0.701	
	Post progression	0.651	0.496	

4.2.9.1 Health state utilities from HER2CLIMB

Utilities for the tucatinib combination intervention were obtained from HER2CLIMB EQ-5D-5L data, mapped to the EQ-5D-3L UK 'social tariff' value set using the crosswalk

procedure.[25] The company also report results for a scenario with an EQ-5D-5L value set, but we do not discuss this further as it does not follow the NICE recommended approach.[28]

The original base case uses simple means by HER2CLIMB study arm, as assessed at treatment cycle 3, 5, 7 and 9 to model pre-progression utility in cycles 1-2, 3-4, 5-6 and 7+ respectively (Clarification response Addendum B1). Post-progression utilities are based on data from the 30-day post-treatment assessment. As explained in CS B.2.6.9 and B.3.4.1, the EQ-5D-5L assessments were introduced as a protocol amendment and so only conducted for a subset of patients: 217/410 (53%) in the tucatinib combination arm and 112/202 (55%) in the placebo combination arm. The company argue that patient characteristics for this subset are reflective of those for the whole study population (see section 3.2.5.4 above).

In response to clarification question B2, the company revised their base case utilities for the tucatinib combination and trastuzumab + capecitabine based on a repeated measures analysis of the HER2CLIMB data with baseline utility as a covariate. Utility estimates differed between 'on treatment' and 'off treatment' assessments: 0.762 (95% CI: 0.744-0.781) and 0.698 (95% CI: 0.668- 0.728) respectively. However, other details of the analysis are not reported. In particular, there is no mention of tests for a difference in utility between the study arms or for a trend in utility over pre-progression treatment cycles. There is also no discussion of missing data or how this was handled, which is potentially important given the high proportion of missing data at later assessments. At treatment cycle 3, utility scores were available for 175 out of 213 patients with a baseline score (82%) in the tucatinib arm, and 89 out of 112 patients (79%) in the placebo arm. By treatment cycle 9, completion rates fell to 86/213 (40%) in the tucatinib arm and 38/ 112 (34%) in the placebo arm. And at 30 days post-treatment, scores were available for 72/213 (34%) in the tucatinib arm and 42/112 (38%) in the placebo arm.

4.2.9.2 Health state utilities for comparators

The company report results from their systematic literature review of utilities (CS B.3.4.3 and Appendix H; and Addendum for the updated search). They state that this did not identify any relevant 'primary' utility studies, but that a direct elicitation study by Lloyd et al. (2006)[1] was commonly used in economic evaluations, including the NICE appraisal of eribulin (TA423).[24] Lloyd et al. used a direct preference-based approach (standard gamble) to elicit utilities for different stages of breast cancer from a UK general population sample (n=100).

In TA423, the company (Eisai) estimated utilities by applying a mapping algorithm (Crott and Briggs 2010)[2] to health-related quality of life data (QLQ-C30) from a trial of eribulin compared with capecitabine (Study 301).[3] The ERG in TA423 (LRiG) accepted the company estimates for pre-progression utility (0.706 for eribulin and 0.701 for capecitabine) but argued that the estimate for progressed disease (0.679) was not plausible. The ERG preferred the Lloyd et al. estimate for progressed disease (0.496).[1] The NICE committee in TA423 concluded that the most plausible utility value for progressed disease was likely to be somewhere between the company and ERG estimates.

In the current submission, the company based their utility estimates for the comparators eribulin, capecitabine and vinorelbine on the values reported in TA423 (see Table 23 above). However, they used the wrong values for the Crott and Briggs estimates (taken from Table 50, rather than Table 57, in the Eisai company submission for TA423). This error was corrected in response to clarification question B3.

In their original submission, the company also argued that it would be appropriate to take an average of the TA423 company and ERG utility estimates, to reflect the committee's conclusion that the most plausible value lies between these estimates (CS B.3.4.8). However, the revised company base case only uses the TA423 ERG (Lloyd et al) estimate for progressed disease, rather than taking a mean of the Crott and Briggs and Lloyd et al. values.

4.2.9.3 Adverse event disutilities

The company assumes that utility loss due to treatment-related adverse events is captured in the utility weight for the tucatinib combination, as estimated from HER2CLIMB EQ-5D data (CS B.3.4.4). Similarly, they assume that utility loss due to adverse events are captured in the utility weights for eribulin and capecitabine from TA423. The model includes an option to include QALY loss associated with adverse events, although this was not applied in the company base case or scenario analyses. We found that this has a very negligible effect on the QALY results (see section 6.3).

ERG conclusions

- We agree with the use of pre-progression utilities from the HER2CLIMB for the trial (EQ-5D UK crosswalk values).[25] This is consistent with NICE preferred methods and relevant to the population and intervention in the decision problem.[29 30] In revised analysis of the HER2CLIMB utility data, the company use a repeated measures model with adjustment for baseline values (clarification

response B2): 0.765 for pre-progression and 0.698 for post-progression for the tucatinib combination. This is preferable to the approach in the original base case, though we have concerns about the lack of detail in reporting and potential for bias due to missing data, particularly at the post-treatment follow-up.

- The company use utility estimates for eribulin, capecitabine and vinorelbine from the NICE appraisal TA423.[24] The revised base case includes pre-progression utilities of 0.706 for eribulin and 0.701 for capecitabine and vinorelbine; and the post-progression utility of 0.496 from Lloyd et al.[1] However, the TA423 committee concluded that the most plausible post-progression utility lies somewhere between the Lloyd et al. estimate and an estimate of 0.679 (Crott and Briggs mapping of the Study 301 trial data).[2 3] For ERG scenario analysis, we use a mean of these values for post-progression utility (0.588).
- We also note that in TA423 the same post-progression utility was used across treatments. By comparison, the post-progression utility for the tucatinib combination in the company's revised base case (0.698) is much higher than that assumed for eribulin, capecitabine and vinorelbine (0.496). This difference is not based on comparative evidence and seems implausible. It is not clear why such a large difference should persist after progression and treatment discontinuation. For ERG analysis, we therefore use the same post-progression utility for the tucatinib combination and comparators.
- We further question the clinical plausibility of the difference in pre-progression utility for the tucatinib combination (0.762) and comparators (0.706 for eribulin and 0.701 for capecitabine and vinorelbine). This may well relate to differences in the trial populations (HER2CLIMB versus Study 301)[3 4] or valuation methods (crosswalk EQ-5D versus Crott and Briggs mapping),[2 25] rather than to differences in treatment-related quality of life. And clinical advice to the ERG is that adverse effects and quality of life will be similar across these treatments. We therefore test the effect of assuming the same pre-progression utility across treatments.
- The company assume that the impact of adverse effects on utility is captured in the trial-based utility estimates. This is reasonable and there is no evidence of a difference in the incidence or severity of adverse events between treatments. Furthermore, we note that the using the model option to include disutilities for adverse events has a minimal effect on results.
- The company's model does not include adjustment for the expected decline in utility with age. We have added this for ERG analysis, based on the relationship

between EQ-5D values and age estimated from Health Survey for England data (Ara and Brazier 2010).[31]

4.2.10 Resources and costs

The economic model includes costs for drug acquisition and administration for the tucatinib combination, comparators and subsequent treatments; follow-up and care; and treatment of adverse effects (CS Document B section B.3.5). The CS reported that a systematic literature review was conducted to identify costs and resource use (Appendices G and I, and clarification response Addendum).

4.2.10.1 Drug acquisition and administration costs

Drug acquisition costs for the tucatinib combination and comparators are summarised in CS Document B Tables 25 and 26; and drug administration costs are summarised in CS Document B Table 27. The dosing regimen for the tucatinib combination is based on the MHRA SmPC (CS Document B section B.3.5.1.1).

Total drug acquisition costs per 21-day treatment cycle are summarised in CS Table 28. These include adjustment for relative dose intensity, based on the HER2CLIMB trial for the tucatinib combination and trastuzumab + capecitabine and Yuan et al. (2019)[14] for eribulin and vinorelbine. The source of the relative dose intensity for capecitabine monotherapy (78.8%) is not clear. In CS Table this is attributed to the NALA study (Saura et al. 2020).[32] However, the online appendix (Table A3) in the Saura et al. paper quotes a relative dose intensity for capecitabine of 93% in the neratinib + capecitabine arm and 86% in the lapatinib + capecitabine arm (89% pooled across the arms). This causes a small increase in the estimated cost of capecitabine, but the impact on cost-effectiveness results is negligible.

The model includes wastage estimates for intravenous trastuzumab and T-DM1, but the company does not include these estimates in base case analysis or a scenario.

4.2.10.2 Subsequent treatment costs

Costs for post-progression anticancer treatments were estimated for patients entering the progressed state (CS Table 29). Drug acquisition costs were obtained from the BNF 2021[33] and eMIT 2021[34], and the dose and treatment duration for the drugs were based on related trials: HER2CLIMB, NALA, PHEREXA and EMILIA.[4 32 35 36]

The company use data from the HER2CLIMB trial to estimate use of subsequent treatments (Table 24 below). For the comparator drugs (that is, eribulin, capecitabine and vinorelbine) a weighted average for the HER2CLIMB treatment arms is used. These assumptions are not reflective of current clinical practice in England because pertuzumab, T-DM1, lapatinib and neratinib are not funded for fourth-line treatment of HER2+ metastatic breast cancer. We understand that after progression, about half of this patient group would receive trastuzumab with capecitabine and that others who receive further treatment would have chemotherapy alone (e.g. capecitabine or vinorelbine). We conduct an exploratory ERG scenario analysis based on these estimates, and assuming that 30% of patients would not receive further anticancer therapy (ERG scenario in Table 24).

Table 24 Proportion of patients receiving post-progression anticancer treatments

Subsequent treatment	Tucatinib Combination	Trastuzumab + capecitabine	Eribulin, vinorelbine, capecitabine	ERG scenario (all treatments)
Trastuzumab	[redacted]	[redacted]	[redacted]	
Pertuzumab	[redacted]	[redacted]	[redacted]	
T-DM1	[redacted]	[redacted]	[redacted]	
Lapatinib	[redacted]	[redacted]	[redacted]	
Neratinib	[redacted]	[redacted]	[redacted]	
Tras + cap				50.0%
Capecitabine				10.0%
Vinorelbine				10.0%
No treatment	[redacted]	[redacted]	[redacted]	30.0%

Source: adapted by ERG from CS Table 30.
T-DM1, trastuzumab emtansine; Tras + cap, trastuzumab with capecitabine

4.2.10.3 Health state costs

Resource use and costs for the pre- and post-progression health states and for end of life care are summarised in CS Document B Table 31. Assumptions about the frequency of consultations were taken from TA458[21], with unit costs from the NHS National Cost Collection 2018/19 and PSSRU 2020.[37 38] End of life care costs were based on TA458, updated for inflation.

4.2.10.4 Adverse events costs

Adverse events costs used in the economic model are summarised in CS Document B Table 32. The company included costs associated with Grade 3 and 4 treatment-emergent adverse events that occurred in at least 2% of patients in HER2CLIMB. They used previous NICE appraisals (TA423, TA496, TA621, TA579) to inform the cost estimates.

The company also included the cost of the supportive antidiarrheal medication-loperamide. Data from HER2CLIMB was used for the proportion of patients receiving loperamide in the tucatinib combination and trastuzumab + capecitabine arms. For the eribulin, capecitabine and vinorelbine comparators, the company assumed that the dose of loperamide and mean treatment duration were the same as for trastuzumab + capecitabine.

ERG conclusions

The company's estimates of resource use and costs are generally appropriate. In ERG analysis, we include drug wastage costs, and an alternative scenario for the cost of subsequent treatments (see section 6.3).

5 COST EFFECTIVENESS RESULTS

5.1 Company's cost effectiveness results

The company report their deterministic base case results in CS Document B Table 35. This and all other cost-effectiveness results in this report are conducted with a Patient Access Scheme (PAS) price discount for tucatinib and an assumed price discount for trastuzumab, with all other comparator and subsequent treatments at list price. We present results with all available PAS/CMU price discounts in a confidential addendum to this report. In their response to clarification question B2, the company provided results for a revised base case, which includes changes to the utility estimates from the analysis of HER2CLIMB EQ-5D data and values used in TA423 (see Table 25 below).

Table 25 Company's base case cost-effectiveness results, deterministic (PAS price for tucatinib and assumed discount for trastuzumab, all other drugs at list price)

	Total costs (£)	Total QALYs	Incremental costs (£)	Incremental QALYs	ICER (£/QALY)
Tucatinib combination	[redacted]	[redacted]	-	-	-
Eribulin	[redacted]	[redacted]	[redacted]	[redacted]	37,483

Source: Reproduced from company's response to clarification question B2 Table B2c

The company did not provide incremental analyses including all the other comparators. We report full incremental analysis for the company's base case in section 6.

5.2 Company's sensitivity analysis

5.2.1 Deterministic sensitivity analysis

The company report results from their one-way, deterministic sensitivity analysis in the tornado plot in CS Document B Figure 27. The variations for most input parameters were based on simple assumed percentages rather than empirical evidence. This applies to discount rates, mean age, body weight, body surface area, dose intensity, post progression rates, treatment costs, AEs and utilities. The company did not include any parameters for survival models in their deterministic sensitivity analyses. Their sensitivity analysis results indicated that relative dose intensity for the tucatinib combination, and health state utilities have the largest impact on the cost-effectiveness results.

The company did not update their one-way, deterministic sensitivity analysis alongside their updated cost-effectiveness results that was provided as response to their clarification response to question B2. We found similar results to those for the original base case.

5.2.2 Scenario analysis

The company reported nine scenario analyses (CS Document Table 39). They updated their results in their clarification response. We present results including the PAS price for tucatinib and an assumed discount for trastuzumab, and all other drugs at list price in Table 26 (reproduced from CS Document B Table 40 and company's response to clarification question B2 Table B2d).

Table 26 Scenario analyses explored in the model

No	Scenario	ICER (original)	ICER (updated)
Base case		£46,756	£37,483
1	[redacted]	[redacted]	[redacted]
2	Tucatinib combination utilities: EQ-5D-5L	[redacted]	[redacted]
3	[redacted]	[redacted]	[redacted]
4	[redacted]	[redacted]	[redacted]
5	[redacted]	[redacted]	[redacted]
6	Treatment duration: Restricted mean treatment exposure	[redacted]	[redacted]
7	Comparator: Vinorelbine	[redacted]	[redacted]
8	Comparator: Capecitabine	[redacted]	[redacted]
9	Blended ICER: [redacted]	[redacted]	[redacted]

With the PAS for tucatinib and assumed price reduction for trastuzumab (and list prices for other drugs), scenarios for the revised base case give ICERs ranging between [redacted]. The [redacted]. In both these scenarios, the [redacted].

[redacted].

We report additional ERG scenario analyses in section 6.3 below.

5.2.3 Probabilistic sensitivity analysis

The company conducted a probabilistic sensitivity analysis (PSA), with input parameter distributions as reported in CS Document B, Table 33. The company's probabilistic base case results for their original base case (PAS price for tucatinib, assumed discount for trastuzumab and list price for other drugs) are reported in CS Document B Table 37. The cost-effectiveness scatter plot and acceptability curve (PAS price for tucatinib and trastuzumab, list price for others) are shown in CS Document B Figures 23 and 25. The company did not update their probabilistic sensitivity analysis for their revised base case produced in response to clarification question B2. However, the ERG confirms that the probabilistic results are similar to the deterministic results: ICER for the tucatinib combination compared with eribulin £35,452 per QALY gained compared with £37,483.

ERG conclusions

- The company's deterministic and probabilistic sensitivity analyses do not provide an accurate reflection of parametric uncertainty because the variance assumed for many of the input parameters is not based on the available evidence.
- The company presented very limited scenario analyses and did not explore uncertainty related to different survival models fitted to OS and PFS.

5.3 Model validation

The company approach to validation is described in CS section B.3.11. This included assessment of clinical plausibility of PFS and OS extrapolations by an advisory board. They did not provide any information on model quality control, internal validity (i.e. comparing the model results with outputs from the HER2CLIMB trial) or external validity (i.e. comparison of the model results with external data.

The ERG conducted a series of quality checks of the company model. This included: checking that the input parameters in the model matched the values cited in the CS and in the original sources; and validating the results of the scenario and sensitivity analyses as reported by the company. We also conducted a series of 'white box' and 'black box' checks

to validate the model. We spotted a few inconsistencies between parameters in the model and values reported in the CS; these have been described in our critique above.

5.3.1 Internal validation

For internal validation, the ERG have provided a comparison of the modelled survival estimated with the observed data from the HER2CLIMB, produced in Figure 6 within section 4.2.6 of this document. The model OS estimates in the company's base case are not comparable with those in the HER2CLIMB trial, the model survival estimates are consistently higher than in the trial. We present a comparison of the survival estimates in Table 27 below.

Table 27 Comparison of survival predictions from the model and HER2CLIMB

Timepoint	Tucatinib combination		Trastuzumab + capecitabine	
	Model	HER2CLIMB	Model	HER2CLIMB
Overall Survival				
1-year	[redacted]	[redacted]	[redacted]	[redacted]
2-years	[redacted]	[redacted]	[redacted]	[redacted]
3-years	[redacted]	[redacted]	[redacted]	[redacted]
Progression Free Survival				
1-year	[redacted]	[redacted]	[redacted]	[redacted]
2-years	[redacted]	[redacted]	[redacted]	[redacted]
3-years	[redacted]	[redacted]	[redacted]	[redacted]

5.3.2 External validation

The company report a targeted search to identify studies that presented long-term survival data for metastatic breast cancer. Further details are in CS Appendix L. Of the 12 studies identified, only HER2CLIMB provided survival data for the tucatinib combination (612 patients for a duration of 3 years).[4] Two studies had follow up >10 years; 4 studies had follow up over 5 years but <10 years; and 6 studies were considered to provide a good match to the HER2CLIMB data. The study by Urruticoechea et al. (2017) included a trastuzumab + capecitabine control arm with five years of follow up.[35] This was a peripheral study in the company's NMA network that did not contribute to the indirect comparisons of interest for this appraisal. Figure 13 shows the modelled OS curve for tras + cap from the company's base case analysis, alongside their 'trial based' OS curve (fitted to HER2CLIMB trial data) and the KM data for the HER2CLIMB and Urruticoechea control arms. This shows the difference in OS estimates from these two trials, possibly due to differences in the patient populations (e.g. prevalence of brain metastases). It also shows the difference in OS extrapolations derived from the HER2CLIMB trial data alone, compared with the company's

base case modelling approach (fractional polynomial curve for the reference lapatinib + capecitabine, adjusted for with NMA HRs).

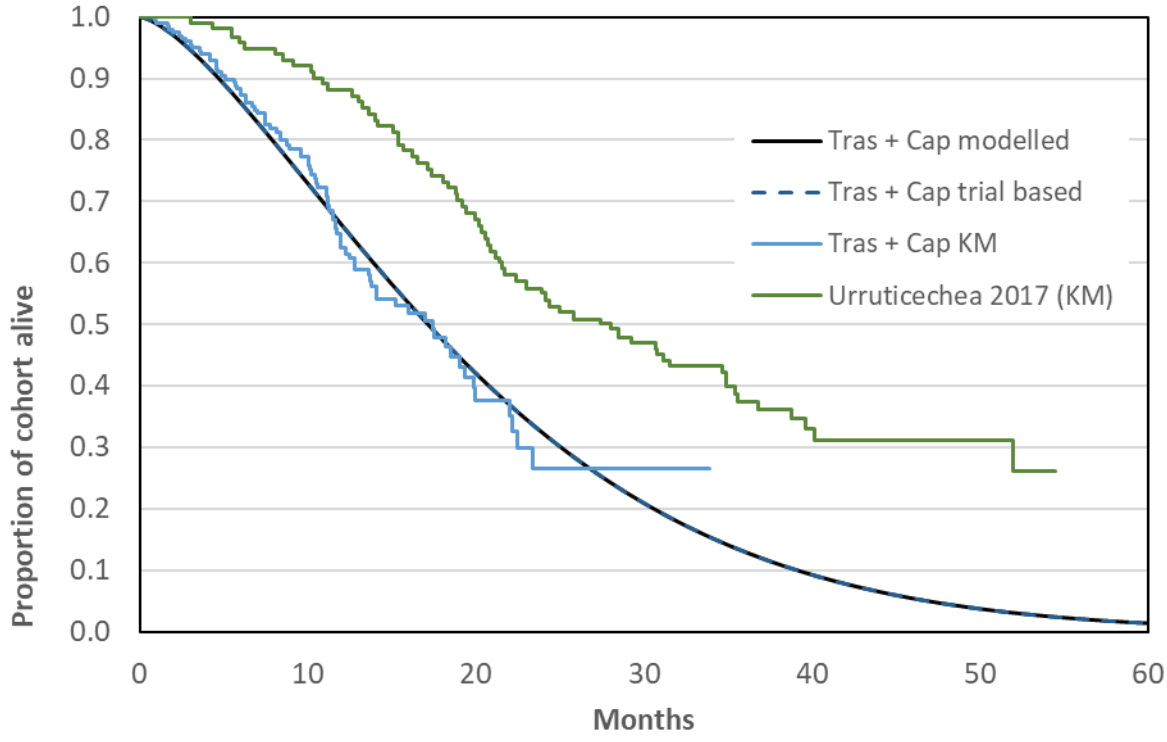


Figure 13 OS trastuzumab + capecitabine: modelled extrapolations and KM data

6 ERG'S ADDITIONAL ANALYSES

6.1 Corrections to the company's base case

The ERG did not identify any errors that affected the company's base case analysis. However, we did make some edits to the model to run the company's revised base case and to enable additional scenario analysis. These changes are described in Table 28.

Table 28 ERG changes to the company model

Issue	Change made	Location in model
Control sheet	Addition of new sheet to apply company and ERG scenario analysis a view summary of fully incremental results	'ERG' sheet. Note that ERG changes to the model are highlighted in green.
Utilities for HER2CLIMB arms	Added pooled estimates from RRMM analysis (clarification response Addendum B2)	'Default Data' sheet, rows 213 to 218. Controls on 'Utilities' and 'ERG' sheets.
Utilities for comparators	Added correct values for ERG and company analysis in TA423 as agreed in company response to clarification question B3	'Default Data' sheet, rows 219 to 229. Controls on 'Utilities' and 'ERG' sheets.
Utility age multipliers	Utility multiplier used to adjust utilities in Markov sheets as the cohort ages within the model. Adjustment based on Ara and Brazier 2010 formula for the general population.[31]	Coefficients for the Ara and Brazier formula added to 'Utilities' sheet. Controls on 'Utilities' and 'ERG' sheets. Edits to columns G, H, AD, AF, AI and AJ on 'Calc_Int' and 'Calc_Comp6' to 'Calc_Comp8' sheets.
Within-trial analysis	Extended direct trial survival estimates to end of 20-year time horizon (rows 1041 to 1062). This enables multiway comparison for within-trial analysis	CB1041 to MQ1062 on 'RCT Survival_PFS' and 'RCT Survival_OS' sheets
Gompertz extrapolations	#Num! error due to estimation of hazard from survival estimates below Excel smallest number	Error trap added to Survival Curves D7 and rows G and K
Taper for OS HR NMA	Corrected 'Time to HR=1' for the OS HR NMA - converted from years to months, as on for PFS HR NMA and FP sheets.	'Bayesian NMA HR Models_OS' BX3
Subsequent treatment	Included 'no subsequent treatment' as class in Dirichlet distribution for PSA. Does not change deterministic results	'Country-Specific Data' rows 128 to 165
Tucatinib discount	For clarity only, same method as for calculation of PAS discount for tucatinib for other drugs.	'Country-Specific Data' K47-L47

6.2 Company revised base case and scenarios

We show results for the company's base case and scenarios in Table 29 and Table 30 respectively. For the revised base case analysis, the company reports the pairwise ICER for the tucatinib combination compared with eribulin, £37,483 per QALY gained. In the full incremental analyses for scope comparators, eribulin is dominated and vinorelbine is subject to extended dominance, so the ICER for the tucatinib combination is [redacted] per QALY compared with capecitabine. If trastuzumab + capecitabine is also included, this is the correct incremental comparator for the tucatinib combination (ICER [redacted]).

Table 29 Company's revised base case, deterministic

Treatment	Cost [a]	QALYs	Pairwise ICERs Tuc + tras + cap vs. comparators	ICERs fully incremental	
				Excluding Tras + cap	Including Tras + cap
Capecitabine	[redacted]	[redacted]	[redacted]	-	-
Vinorelbine	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]	-	[redacted]
Eribulin	[redacted]	[redacted]	£37,483	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-	[redacted]	[redacted]

Source: Clarification response Table B2c (tucatinib combination and eribulin), other results produced by the ERG
[a] PAS discount for tucatinib and assumed discount for trastuzumab, other drugs at list price
Ext dom, extendedly dominated; ICER, incremental cost-effectiveness ratio; Tras + cap, trastuzumab with capecitabine; Tuc + tras + cap, tucatinib with trastuzumab and capecitabine.

We note that in the company's scenario with alternative survival curves [redacted], the QALYs do not change. This is because the company base case uses a curve from the FP NMA for a reference treatment (lapatinib + capecitabine), which is adjusted for other comparators using hazard ratios from the NMA. The model only actually includes one FP model for PFS and one for OS. Thus, it is not possible to do scenario analysis on the choice of survival curves for extrapolation in this version of the model. The change in cost for the tucatinib combination arm in this scenario is misleading. This is caused by the method for estimation of TTD. This is derived from fitted curves to HER2CLIMB data, with constraints that TTD cannot exceed PFS, and that PFS cannot exceed OS. Hence, in this scenario the trial-based survival models for PFS and OS change, which has an indirect effect on TTD.

Table 30 Company's scenario analyses on revised base case, deterministic

Treatment	Cost [a]	QALYs	Pairwise ICERs
Revised company base case			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	£37,483
Tuc + tras + cap	[redacted]	[redacted]	-
Survival curves ([redacted])			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-
NMA (PFS and OS [redacted])			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-
NMA (PFS and OS [redacted])			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-
NMA (PFS and OS [redacted])			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-
Treatment duration (restricted mean exposure)			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-

6.3 ERG scenarios applied to the company's base case

6.3.1 Within-trial analysis and survival extrapolations

Table 31 shows the effect of applying a 'within-trial' method for the survival extrapolations for the HER2CLIMB arms to the company's revised base case. This has the effect of reducing

QALYs across all treatments, and also reducing incremental QALYs and hence increasing the ICERs. Changes to the fitted OS survival model (stratified Weibull and Gompertz) have moderate impact on the ICERs.

Table 31 ERG additional scenarios for OS extrapolations, deterministic

Treatment	Cost [a]	QALYs	Pairwise ICERs
Revised company base case			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	£37,483
Tuc + tras + cap	[redacted]	[redacted]	-
Within-trial analysis – OS Weibull			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-
Within-trial analysis – OS stratified Weibull			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-
Within-trial analysis – OS Gompertz			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-

Table 32 shows results for the within-trial survival analyses with changes to extrapolations for PFS. Alternative survival models for PFS have a small impact on the ICER estimates.

Table 32 ERG additional scenarios for PFS extrapolations, deterministic

Treatment	Cost [a]	QALYs	Pairwise ICERs
Revised company base case			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	£37,483
Tuc + tras + cap	[redacted]	[redacted]	-
Within-trial analysis – PFS flexible Weibull 2 knots			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-
Within-trial analysis – PFS stratified generalised gamma			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-
Within-trial analysis - PFS stratified log-normal			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-

6.3.2 Indirect treatment effects

Changes to the NMA model used in the within-trial analysis are shown in Table 33. The ERG's random effects NMA with correction to the Pivot upper confidence limit (see section 3.6 above) reduces the differences in QALYs between the tucatinib combination and indirect comparators, hence increasing these pairwise ICERs.

Table 33 ERG additional scenarios for NMA analyses, deterministic

Treatment	Cost [a]	QALYs	Pairwise ICERs
Revised company base case			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	£37,483
Tuc + tras + cap	[redacted]	[redacted]	-
Within-trial analysis – HR NMA fixed effect (PFS and OS)			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-
Within-trial analysis – HR NMA random effects (PFS and OS)			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-
Within-trial analysis – HR NMA random effects with ERG Pivot correction			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-

6.3.3 Waning of treatment effects

The model includes a scenario to taper HR values from the end of maximum follow up, to a defined timepoint when the HR=1. The default time to HR=1 in the model for OS is 72 months from the start of treatment. This causes a moderate increase in the ICERs. We also tested the impact of reducing the time to HR=1 for OS to 48 months, which causes a further increase in the ICERs. For PFS, the default time to HR=1 is 48 months. This has very little impact on QALYs but reduces costs, hence ICERs are lower with PFS tapering.

Table 34 ERG additional scenarios for tapering of treatment effects, deterministic

Treatment	Cost [a]	QALYs	Pairwise ICERs
Revised company base case			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	£37,483
Tuc + tras + cap	[redacted]	[redacted]	-
Tapering of OS HRs from end of trial follow-up to HR=1 at 72 months			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-
Tapering of OS HRs from end of trial follow-up to HR=1 at 48 months			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-
Tapering of PFS HRs from end of trial follow-up to HR=1 at 48 months			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-

6.3.4 Utility scenarios

The model is sensitive to changes to assumptions about the health state utilities (Table 35). Reducing the difference in utility values between the HER2CLIMB arms and external comparators increases the ICERs for tucatinib compared with capecitabine, vinorelbine and eribulin. The effects of including utility loss due to adverse events and or adjusting for age have little impact on the ICERs.

Table 35 ERG additional scenarios for utilities, deterministic

Treatment	Cost [a]	QALYs	Pairwise ICERs
Revised company base case			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	£37,483
Tuc + tras + cap	[redacted]	[redacted]	-
Pre-progression utility 0.706 for tucatinib and Tras + Cap (TA423 eribulin value)			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-
Post-progression utility 0.588 for all treatments (mean of TA423 estimates)			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-
HER2CLIMB utilities for all treatments (0.762 pre-progression, 0.698 post-progression)			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-

Include AE disutilities			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-
Age adjustment for utilities			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-

6.3.5 Resource use and cost scenarios

Finally, we consider scenarios related to resource use and costs (Table 36). Assuming that treatment duration is equal to PFS increases costs for all treatments, but proportionately more for the tucatinib combination, hence increasing ICERs. The ERG scenario for subsequent treatment use does not have a big impact, except for trastuzumab + capecitabine (because the company's base case assumed higher use of some expensive anti-cancer drugs in this arm, based on HER2CLIMB data). Including drug wastage costs has little impact on overall costs or ICERs.

Table 36 ERG additional scenarios for resource use and costs, deterministic

Treatment	Cost [a]	QALYs	Pairwise ICERs Tuc + tras + cap vs. comparators
Revised company base case			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-

Treatment duration based on PFS (all treatments)			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-
Include costs for drug wastage			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-
ERG subsequent treatment scenario (50% tras, 20% cap/vin: [redacted] per person)			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-

6.4 ERG preferred analysis and scenarios

ERG preferred assumptions are:

- Within-trial analysis: OS and PFS fitted to HER2CLIMB trial data for the tucatinib combination and trastuzumab + capecitabine (see section 4.2.6 above).
- Relative effects for other comparators from the HR NMA with random effects and the ERG correction for the Pivot upper confidence limit (3.6).
- Health state utilities from HER2CLIMB EQ-5D analysis applied to all treatments (4.2.9.2).
- ERG scenario for the use of subsequent treatments (4.2.10.2)
- Adjustment of utilities for age
- Costs for drug wastage.

The cumulative effect of ERG preferred assumptions to the company's base case is shown in Table 37.

Table 37 Cumulative change from company base case to ERG base, deterministic

Treatment	Cost [a]	QALYs	Pairwise ICERs	Change to pairwise ICERs
Revised company base case				
Capecitabine	[redacted]	[redacted]	[redacted]	
Vinorelbine	[redacted]	[redacted]	[redacted]	
Tras + cap	[redacted]	[redacted]	[redacted]	
Eribulin	[redacted]	[redacted]	£37,483	
Tuc + tras + cap	[redacted]	[redacted]		
+ Within-trial analysis (PFS and OS, with HR NMA fixed effect)				
Capecitabine	[redacted]	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]		
+ HR NMA random effects with ERG Pivot correction				
Capecitabine	[redacted]	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]		
+ HER2CLIMB utilities (0.762 pre-progression, 0.698 post-progression)				
Capecitabine	[redacted]	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]		
+ Age-adjustment for utilities				
Capecitabine	[redacted]	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]		
+ ERG subsequent treatment scenario (50% tras, 20% cap/vin: [redacted] per person)				
Capecitabine	[redacted]	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]		
+ Include costs for drug wastage (ERG preferred analysis)				
Capecitabine	[redacted]	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]		

Full incremental results from the ERG preferred analysis are shown in Table 38. Alternative scenarios applied to the ERG base case are shown in Table 39.

Table 38 ERG preferred analysis, deterministic

Treatment	Cost [a]	QALYs	Pairwise ICERs Tuc + tras + cap vs. comparators	ICERs fully incremental Excluding Tras + cap	ICERs fully incremental Including Tras + cap
Capecitabine	[redacted]	[redacted]	[redacted]	-	-
Vinorelbine	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]	-	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-	[redacted]	[redacted]

Source: Clarification response Table B2c (tucatinib combination and eribulin), other results produced by the ERG
[a] PAS discount for tucatinib and assumed discount for trastuzumab, other drugs at list price
Ext dom, extendedly dominated; ICER, incremental cost-effectiveness ratio; Tras + cap, trastuzumab with capecitabine; Tuc + tras + cap, tucatinib with trastuzumab and capecitabine.

Table 39 ERG preferred analysis and scenarios, deterministic

Treatment	Cost [a]	QALYs	Pairwise ICERs
ERG preferred analysis			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-
OS stratified Weibull			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-
OS Gompertz			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-

NMA HR fixed effect (no Pivot correction)			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-
Utilities from TA423 (pre-progression 0.706/701; post-progression 0.588)			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-
Treatment duration equal to PFS (all treatments)			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-
Treatment duration restricted mean treatment exposure			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-
Subsequent treatment (trial-based)			
Capecitabine	[redacted]	[redacted]	[redacted]
Vinorelbine	[redacted]	[redacted]	[redacted]
Tras + cap	[redacted]	[redacted]	[redacted]
Eribulin	[redacted]	[redacted]	[redacted]
Tuc + tras + cap	[redacted]	[redacted]	-

Figure 14 shows the survival curves from this model, alongside HER2CLIMB KM plots. This shows that the within-trial analysis in the ERG preferred model gives a better fit to the results from the pivotal trial than the company's NMA based approach (Figure 6 above).

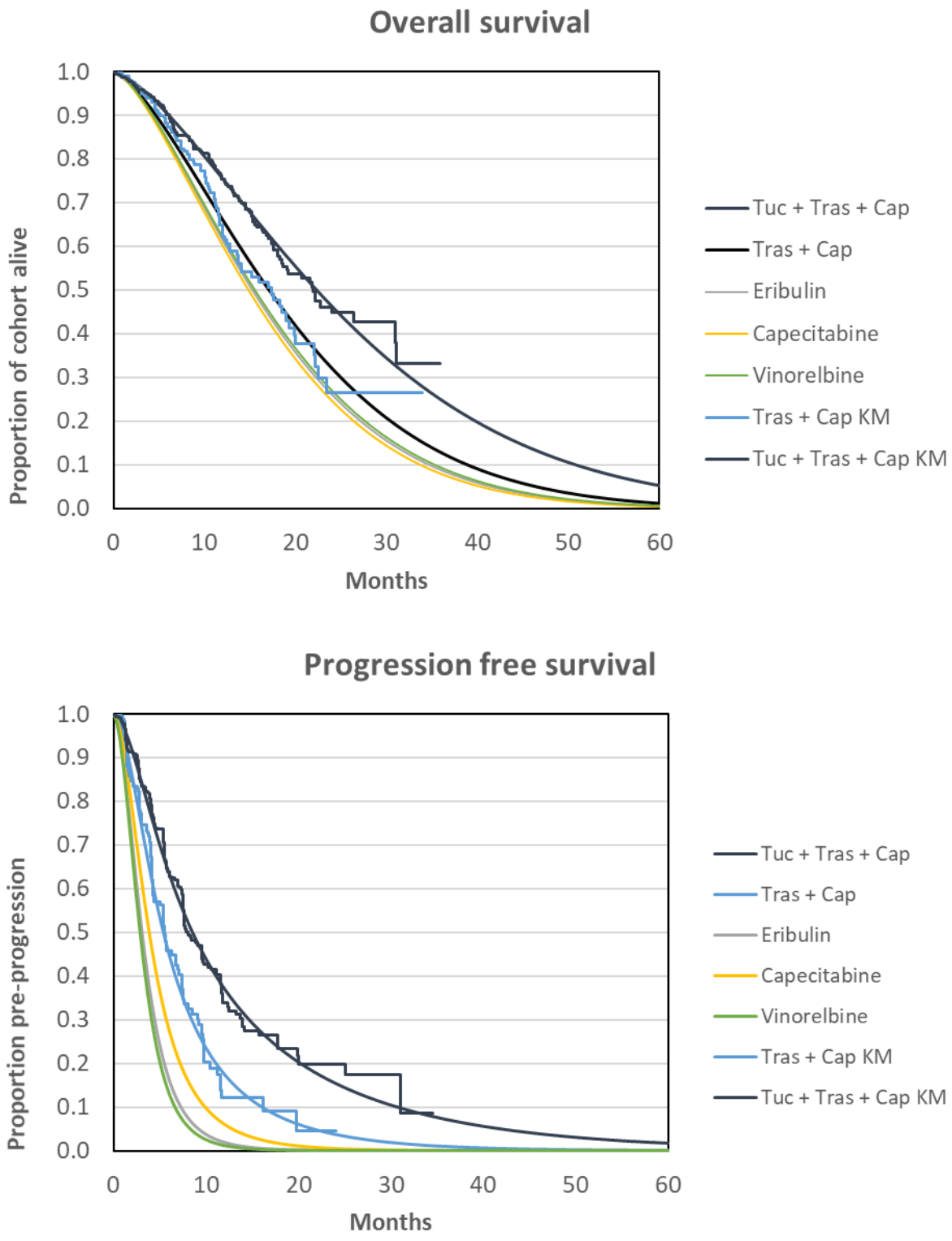


Figure 14 Survival curves from ERG preferred model, with KM data from HER2CLIMB

Source: Obtained from the company's model by the ERG

7 END OF LIFE

The company consider that the tucatinib combination meets NICE end of life criteria for patients with HER2-positive locally advanced or metastatic breast cancer after 2 or more anti-HER2 therapies (CS Table 15). They state that clinical experts in England agreed with their argument that the life expectancy at third line treatment is less than 24 months and the gain in life extension with the tucatinib combination is expected to be greater than 3 months. Furthermore, they state that their argument aligns with previous NICE appraisals for second line and third-line treatment in metastatic setting.

In Table 40, we summarise and critique the company's evidence in support of their case for end of life criteria applying to patients with HER2-positive locally advanced or metastatic breast cancer after 2 or more anti-HER2 therapies.

Table 40 Summary and critique of the CS case for meeting end of life criteria

Criterion	Data available	ERG comment
The treatment is indicated for patients with a short life expectancy, normally less than 24 months	The median OS achieved with the single-agent chemotherapy currently available in the third-line setting (eribulin) is less than 16 months	Median OS with eribulin ranged from 13.1 to 15.9 months in three clinical trials including patients with HER2+ and HER2-negative (HER2-) metastatic breast cancer. We agree with the company's assertion.
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 median OS difference between the tucatinib combination and the placebo combination in HER2CLIMB exceeds 3 months (21.9 versus 17.4 months)	For the tucatinib combination, the mean undiscounted life years based on the company's (revised) model is 2.91 years and that on the ERG's modelled base case is 2.19 years. Tucatinib combination extended life by greater than 3 months in both the ERG and the company's (revised) base case models.

In Table 41 below, we present a comparison of the undiscounted life years of the treatments in comparison for the company's (revised) base case and the ERG base case. We note that tucatinib combination extended life by greater than 3 months compared to the comparators in both the cases.

Table 41 Comparison of the undiscounted life years

Treatment	Undiscounted life years			
	Company's revised base case	Difference (Tucatinib vs comparator)	ERG base case	Difference (Tucatinib vs comparator)
Capecitabine	1.72	1.19	1.45	0.74
Vinorelbine	1.77	1.14	1.51	0.68
Tras + cap	2.22	0.69	1.68	0.51
Eribulin	1.74	1.17	1.49	0.70
Tuc + tras + cap	2.91		2.19	

ERG conclusion

We agree with the company that tucatinib combination meets both the end of life criteria.

8 References

1. Lloyd A, Nafees B, Narewska J, et al. Health state utilities for metastatic breast cancer. Br J Cancer 2006;95(6):683-90. doi: 10.1038/sj.bjc.6603326 [published Online First: 2006/09/13]

2. Crott R, Briggs A. Mapping the QLQ-C30 quality of life cancer questionnaire to EQ-5D patient preferences. Eur J Health Econ 2010;11(4):427-34. doi: 10.1007/s10198-010-0233-7 [published Online First: 2010/05/18]

3. Kaufman PA, Awada A, Twelves C, et al. Phase III open-label randomized study of eribulin mesylate versus capecitabine in patients with locally advanced or metastatic breast cancer previously treated with an anthracycline and a taxane. J Clin Oncol 2015;33(6):594-601. doi: 10.1200/JCO.2013.52.4892 [published Online First: 2015/01/22]

4. Murthy RK, Loi S, Okines A, et al. Tucatinib, Trastuzumab, and Capecitabine for HER2-Positive Metastatic Breast Cancer. N Engl J Med 2020;382(7):597-609. doi: 10.1056/NEJMoa1914609 [published Online First: 2019/12/12]

5. Kuksis M, Gao Y, Tran W, et al. The incidence of brain metastases among patients with metastatic breast cancer: a systematic review and meta-analysis. Neuro-Oncology 2020;23(6):894-904. doi: 10.1093/neuonc/noaa285

6. Paracha N, Reyes A, Dieras V, et al. Evaluating the clinical effectiveness and safety of various HER2-targeted regimens after prior taxane/trastuzumab in patients with previously treated, unresectable, or metastatic HER2-positive breast cancer: a systematic review and network meta-analysis. Breast Cancer Res Treat 2020;180(3):597-609. doi: 10.1007/s10549-020-05577-7 [published Online First: 2020/02/27]

7. ClinicalTrials.gov. A Study of Tucatinib vs. Placebo in Combination With Capecitabine & Trastuzumab in Patients With Advanced HER2+ Breast Cancer (HER2CLIMB).

8. Murthy RK, Loi S, Okines A, et al. Tucatinib, Trastuzumab, and Capecitabine for HER2-Positive Metastatic Breast Cancer – supplementary appendix. N Engl J Med 2020;382(7):597-609. doi: 10.1056/NEJMoa1914609 [published Online First: 2019/12/12]

9. Murthy R, Borges VF, Conlin A, et al. Tucatinib with capecitabine and trastuzumab in advanced HER2-positive metastatic breast cancer with and without brain metastases: a non-randomised, open-label, phase 1b study. Lancet Oncol 2018;19(7):880-88. doi: 10.1016/S1470-2045(18)30256-0 [published Online First: 2018/05/29]

10. Takano T, Tsurutani J, Takahashi M, et al. A randomized phase II trial of trastuzumab plus capecitabine versus lapatinib plus capecitabine in patients with HER2-positive metastatic breast cancer previously treated with trastuzumab and taxanes: WJOG6110B/ELTOP. Breast 2018;40:67-75. doi: 10.1016/j.breast.2018.04.010 [published Online First: 2018/04/27]

11. Pivot X, Manikhas A, Zurawski B, et al. CEREBEL (EGF111438): A Phase III, Randomized, Open-Label Study of Lapatinib Plus Capecitabine Versus Trastuzumab Plus Capecitabine in Patients With Human Epidermal Growth Factor Receptor 2-Positive Metastatic Breast Cancer. J Clin Oncol 2015;33(14):1564-73. doi: 10.1200/JCO.2014.57.1794 [published Online First: 2015/01/22]

12. von Minckwitz G, du Bois A, Schmidt M, et al. Trastuzumab beyond progression in human epidermal growth factor receptor 2-positive advanced breast cancer: a german breast group 26/breast international group 03-05 study. J Clin Oncol 2009;27(12):1999-2006. doi: 10.1200/JCO.2008.19.6618 [published Online First: 2009/03/18]

13. Cameron D, Casey M, Press M, et al. A phase III randomized comparison of lapatinib plus capecitabine versus capecitabine alone in women with advanced breast cancer that has progressed on trastuzumab: updated efficacy and biomarker analyses. Breast Cancer Res Treat 2008;112(3):533-43. doi: 10.1007/s10549-007-9885-0 [published Online First: 2008/01/12]

14. Yuan P, Hu X, Sun T, et al. Eribulin mesilate versus vinorelbine in women with locally recurrent or metastatic breast cancer: A randomised clinical trial. Eur J Cancer 2019;112:57-65. doi: 10.1016/j.ejca.2019.02.002 [published Online First: 2019/04/01]

15. Turner RM, Jackson D, Wei Y, et al. Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. Stat Med 2015;34(6):984-98. doi: 10.1002/sim.6381 [published Online First: 2014/12/06]

16. Turner RM, Davey J, Clarke MJ, et al. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. Int J Epidemiol 2012;41(3):818-27. doi: 10.1093/ije/dys041 [published Online First: 2012/03/31]

17. Therneau TM GP. Modeling survival data: extending the Cox model. . New York: Springer 2000.

18. RTI Health Solutions. Update of economic systematic literature review to support tucatinib in metastatic breast cancer, 2021.

19. Anderson SA, Carrico J, Brodtkorb T-H, et al. Abstract PS10-45: Budget impact of introducing neratinib for third-line treatment of HER2+ metastatic breast cancer in the United States: AACR, 2021.

20. Vondeling G, Palaka E, Livings C, et al. PCN69 Cost-Effectiveness Analysis of NOVEL Therapies for HER2+ Metastatic Breast Cancer:[FAM-] Trastuzumab Deruxtecan Versus Tucatinib Trastuzumab and Capecitabine Combination. Value in Health 2020;23:S433-S34.

21. NICE. Trastuzumab emtansine for treating HER2-positive advanced breast cancer after trastuzumab and a taxane – technology appraisal guidance [TA458] 2017 [updated 20 November. Available from: www.nice.org.uk/guidance/TA458 accessed March 2021.

22. Scottish Medicines Consortium (SMC). Following a full submission: trastuzumab 600mg/5mL solution for injection (Herceptin®) is accepted for restricted use within NHS Scotland [928/13] 2013 [updated 6 December. Available from: www.scottishmedicines.org.uk/medicines-advice/trastuzumab-herceptin-fullsubmission-92813/ accessed April 2021.

23. Scottish Medicines Consortium (SMC). Following a resubmission assessed under the orphan equivalent and end of life process: trastuzumab emtansine (Kadcyla®) is accepted for use within NHS Scotland [990/14] 2017 [updated 10 March. Available from: www.scottishmedicines.org.uk/medicines-advice/trastuzumab-emtansine-kadcyla-resubmission-99014/ accessed April 2021.

24. NICE. Eribulin for treating locally advanced or metastatic breast cancer after 2 or more chemotherapy regimens – technology appraisal guidance [TA423] 2016 [updated 21 December. Available from: www.nice.org.uk/guidance/TA423 accessed April 2021.

25. van Hout B, Janssen MF, Feng YS, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health 2012;15(5):708-15. doi: 10.1016/j.jval.2012.02.008 [published Online First: 2012/08/08]

26. Latimer N. NICE DSU Technical Support Document 14: Undertaking survival analysis for

economic evaluations alongside clinical trials - extrapolation with patient-level data. 2011 2013 [updated March. Available from: http://nicedsu.org.uk/technical-support-documents/survival-analysis-tsd/ accessed April 2021.

27. Seagen Inc. HER2CLIMB Clinical Study Report, 2019.

28. NICE. Position statement on use of the EQ-5D-5L value set for England, 2019.

29. NICE. Guide to the methods of technology appraisal – processes and methods 2013 [Available from: www.nice.org.uk/process/pmg9 accessed April 2021.

30. NICE. Position statement on use of the EQ-5D-5L value set for England (updated October 2019) 2019 [Available from: www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l accessed April 2021.

31. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. Value Health 2010;13(5):509-18. doi: 10.1111/j.1524-4733.2010.00700.x [published Online First: 2010/03/17]

32. Saura C, Oliveira M, Feng YH, et al. Neratinib Plus Capecitabine Versus Lapatinib Plus Capecitabine in HER2-Positive Metastatic Breast Cancer Previously Treated With >/= 2 HER2-Directed Regimens: Phase III NALA Trial. J Clin Oncol 2020;38(27):3138-49. doi: 10.1200/JCO.20.00147 [published Online First: 2020/07/18]

33. BNF. British National Formulary, 2021.

34. eMIT. Drugs and pharmaceutical electronic market information tool (eMIT) – DHA225 2021 [updated 11 March. Available from: www.gov.uk/government/publications/drugs-and-pharmaceutical-electronic-market-information-emit accessed April 2021.

35. Urruticoechea A, Rizwanullah M, Im SA, et al. Randomized Phase III Trial of Trastuzumab Plus Capecitabine With or Without Pertuzumab in Patients With Human Epidermal Growth Factor Receptor 2-Positive Metastatic Breast Cancer Who Experienced Disease Progression During or After Trastuzumab-Based Therapy. J Clin Oncol 2017;35(26):3030-38. doi: 10.1200/JCO.2016.70.6267 [published Online First: 2017/04/25]

36. Verma S, Miles D, Gianni L, et al. Trastuzumab emtansine for HER2-positive advanced breast cancer. N Engl J Med 2012;367(19):1783-91. doi: 10.1056/NEJMoa1209124 [published Online First: 2012/10/02]

37. NHS. 2018/19 National Cost Collection data (NHS Improvement) 2019 [Available from: www.england.nhs.uk/national-cost-collection/#ncc1819 accessed April 2021.

38. PSSRU. Unit costs of health and social care 2019. Personal Social Services Research Unit. 2020

9 Appendices

9.1 Company and ERG risk of bias assessment of HER2CLIMB

Assessment criteria	Company judgement		ERG judgement
Was randomisation carried out appropriately?	Yes	Yes – patients were randomised in a 2:1 ratio using a dynamic hierarchical randomisation scheme to receive tucatinib or placebo in combination with capecitabine and trastuzumab	Agree: low risk of bias
Was the concealment of treatment allocation adequate?	Yes	Yes – adequate blind allocation was achieved with the applied randomisation scheme	Agree: low risk of bias
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes	Yes – baseline patient characteristics were balanced between the treatment arms	Agree: low risk of bias We note a slight imbalance in the proportion of white participants and those with liver metastases, both of which are slightly higher in the placebo arm. The implications of this are unclear.
Were the care providers, participants and outcome assessors blind to treatment allocation?	Yes	Yes – the first part of the study was carried out blindly for the investigator, study centre personnel, clinical research organisation staff and sponsor personnel (except for prespecified Safety personnel)	Agree: low risk of bias[a]
Were there any unexpected imbalances in dropouts between groups?	No	No – balanced, low rates of dropouts were observed in both treatment arms: 23/404 (5.7%) patients discontinued tucatinib and 6/197 (3.0%) patients discontinued placebo	Agree: low risk of bias[b] We note a higher proportion discontinued placebo (86.3%) than tucatinib (70.8%), more commonly due to progressive disease in the placebo arm (68% vs 50%). Discontinuations due to adverse events were higher with tucatinib (5.7% vs 3.0%)
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No	No – all predefined endpoints were reported	Agree: low risk of bias We note DOR and CBR determined by the investigator as well as BICR are listed in the protocol as secondary endpoints, only BICR is reported in the CS.

			Investigator results are reported on clinical trials register.
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes	Yes – the primary endpoint was assessed in the first 480 enrolled patients (primary endpoint population), and patients without outcomes for PFS and OS were censored and those with missing data considered non-responders for ORR and CBR outcomes	Agree: low risk of bias EQ-5D was completed by a subset of the population.

Source CS Table 7 and CS Appendix D Table 22. [a] 'the first part of the study' refers to the phase prior to the open label extension (clarification response C1). [b] the rates in the company response are rates of adverse events leading to discontinuation (clarification response C2).

9.2 Summary of HER2CLIMB trial outcomes and statistical procedures

Trial outcomes

Outcome type	**Outcome measures** (CS Table 4)	**Outcome definitions**	**ERG comments**
Primary endpoint	PFS	Disease response and progression were evaluated in accordance with RECIST criteria version 1.1 by Blinded Independent Central Review (BICR)	Assessed in the 'primary endpoint population' (see discussion Section 3.2.1.1). Defined in the Clinical Study Report (CSR) as the time from randomisation to documented disease progression or death from any cause. Details of the BICR were not reported.
Secondary endpoints (pre-specified alpha-controlled see Section 3.2.4)	-PFS in people with brain metastases at baseline -OS -Confirmed objective response rate (ORR)		-Assessed in a subgroup of the total population -In the total population OS defined in the CSR (time from randomisation to death from any cause) -Confirmed ORR defined as the best overall response in those with measurable disease at baseline. No ERG concerns
Other secondary endpoints	-PFS -Duration of response (DOR) and clinical	-By investigator assessment -By BICR	-Assessed in the total population -In the CSR and trial protocol DOR and CBR determined by investigator assessment were also secondary endpoints.

	benefit rate (CBR)		DOR defined as the time from the first objective response to documented disease progression or death from any cause. CBR defined as those achieving stable disease (SD) or non-complete response (CR)/non-progressive disease (PD) for at least 6 months or a best overall response of confirmed CR or confirmed partial response (PR). Additional secondary / exploratory outcomes reported were time to brain progression by BICR (CS B.2.6.8)
Patient reported outcomes	HRQoL by EQ-5D-5L (CS 2.6.9)	Following a protocol amendment, subgroup of total sample	The CS reports baseline and endpoint data for the EQ-5D descriptive system and the EQ-5D VAS. The trial protocol states that the treatment and placebo group index value changes will be summarised and that responses on the descriptive system will be converted to EQ-5D Index scores using a valuation set as recommended by EuroQol. However, no index scores were provided in the CS but were provided in response to clarification question A2. As described in CS 2.6.9 the inclusion of HRQoL as an outcome was made at protocol amendment seven and consequently only a subset of the total population (tucatinib n=217; placebo n=112) had data.
Safety endpoints	-Adverse events -Clinical laboratory assessments; vital signs and other relevant safety variables -Frequency of dose modifications of tucatinib, capecitabine and trastuzumab	-Modifications could include dose holding, dose reductions and discontinuations	Matches CSR and protocol No ERG concerns

Summary of trial statistical procedures

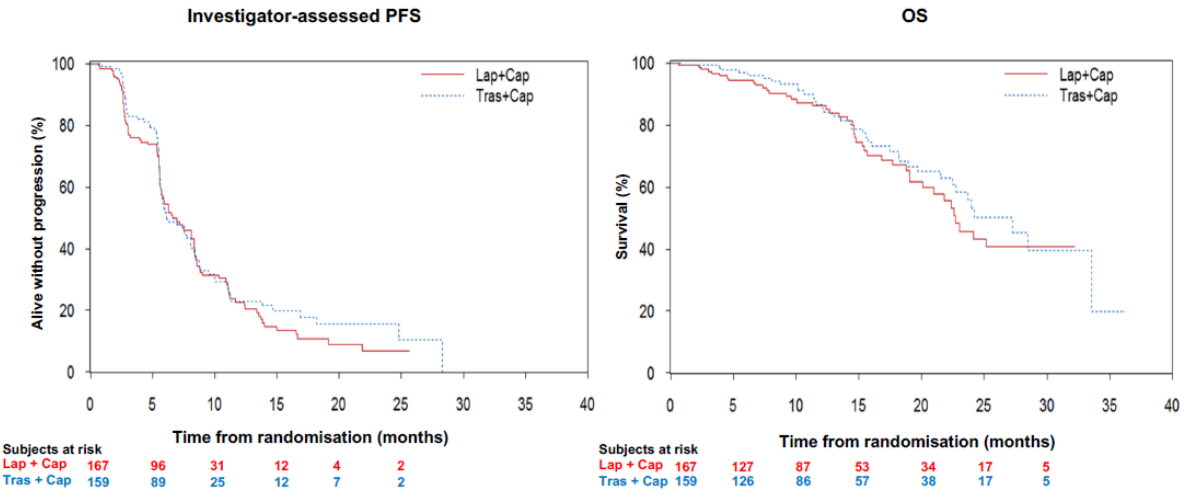
	ERG comments
Sample size calculation	Reported in CS Section B. 2.3.1 and B.2.4, with further detail in the trial protocol / statistical analysis plan and the CSR. The ERG has no concerns over the sample size calculation; the trial was large and appeared adequately powered for the reported outcomes.

Statistical approach for each outcome	Detail as reported in CS Section B.2.4 unless otherwise stated. - PFS in the primary endpoint evaluation set used a stratified, log-rank test controlling for the randomisation stratification factors after 275 PFS events (aim was for 288 events). Unstratified log-rank tests and the stratified and unstratified Wilcoxon tests were also reported in the trial publication to be undertaken as supportive measures. - With the primary endpoint PFS analysis being statistically significant, the alpha controlled key secondary outcomes of OS (total population) and PFS (brain metastases subgroup) were parallel tested (changed from hierarchical testing at protocol amendment 8 to allow for the importance of OS) at significance levels (alpha) initially set at 0.02 and 0.03, respectively, in an interim analysis. To control for multiplicity of outcomes and analyses (interim and final), the risk of a type I error was controlled using the group sequential Holm variable procedure (if only one of the two key secondary outcomes were statistical significant, the unused alpha could be passed to the other outcome [from the trial protocol / SAP]) with the Lan–DeMets alpha-spending function with an O'Brien–Fleming boundary (where the total number and timing of the interim analyses does not need to be specified in advance and how much of the alpha is 'spent' each time an analysis is undertaken is defined). In the interim analysis of PFS brain metastases and OS the 2-sided alpha's were 0.008 and 0.0074 respectively (CS page 32) and no further analysis is planned. The trial arms were compared using the same statistical approaches described above (taken from the trial publication). - With OS and PFS brain metastases being statistically significant, ORR was tested at a two-sided alpha level of 0.05 using a stratified Cochran–Mantel–Haenszel test. - PFS and OS curves were estimated with Kaplan-Meier methodology and stratified Cox proportional-hazards models to estimate hazard ratios (HRs) and 95% CI were undertaken. CS Figures 16 and 17 (Section B.2.9.1.6) provide log hazard plots for PFS and OS respectively as evidence that the proportional hazards assumption holds. -A re-randomisation procedure (with 10,000 alternative subject randomisations) was used to generate p-values for the primary endpoint and key secondary endpoint analyses to reflect the dynamic hierarchical allocation scheme (Table 5, ref SAP) - The additional secondary and exploratory outcomes were not subject to type 1 error control
Handling of missing data for each outcome	Participants without disease progression or death outcomes as appropriate for PFS and OS were censored at the time of the last assessment or the date of randomisation if there was no post baseline information (described in the trial protocol / SAP). Participants with missing data for ORR and CBR were considered non-responders / not having clinical benefit (described in the trial protocol / SAP). The CS do not provide any details of how missing EQ-5D data were handled.
Sensitivity analysis for statistical analyses	In the trial SAP (CS ref 15 Murthy 2020) Section 5.2.1.2 discusses potential sensitivity analyses which may be undertaken for the primary PFS endpoint, including in the case of a non-proportional hazard or stratification errors, and for missing disease response assessments and new anti-cancer therapy before disease progression or death. These were not discussed in the CS. The CSR reports that the latter two sensitivity analyses were undertaken

	and that results were [redacted] The proportional hazards assumption held and therefore sensitivity analyses were not required.
Prespecified subgroups	Pre-planned subgroups were reported in CS Table 4 (Age (≥65 or <65 years); Race (white or non-white); Hormone receptor status (HmR+ or HmR-); Baseline brain metastases (yes or no); ECOG performance-status score (0 or 1); Geographic region (US and Canada or rest of world)). The CS does not describe the statistical analyses for these subgroups but the trial protocol / SAP reports that the subgroup analyses used conventional stratified log rank statistical methods and stratified Cox proportional hazards regression models.

9.3 Corrected hazard ratio confidence interval for Pivot et al (2015)[11]

PFS and OS in patients with prior trastuzumab treatment (ITT)



	Lap + Cap (N=167)	Tras + Cap (N=159)
Median PFS, months	6.6	6.1
Hazard ratio (95% CI)	1.13 (0.85, 1.50)	

	Lap + Cap (N=167)	Tras + Cap (N=159)
Median OS, months	22.7	27.3
Hazard ratio (95% CI)	1.18 (0.76, 1.83)	

19

Personal communication, Professor Xavier Pivot, 16/06/21

9.4 ERG critique of the fractional polynomial NMA

As mentioned in section 3.4 of this report, the company conducted an NMA using fractional polynomial methodology, as an alternative to the HR NMA, to account for potential violation in proportional hazards. The ERG does not consider there is sufficient evidence to reject the proportional hazards assumption, and we therefore consider the HR NMA is appropriate for this appraisal. For this reason we do not discuss the fractional polynomial NMA in detail in this report. For completeness we provide a brief appraisal of the fractional polynomial NMA below.

The ERG considers the fractional polynomial analysis well conducted in terms of:

- [redacted] (CS Appendix D, page 94)
- [redacted] In general, the fractional polynomial NMA results were consistent with those of the HR NMA; where inconsistencies were noted, these were "not relevant to the NICE decision problem" (clarification response A14).
- As the company point out, there is no current methodology for the use of informative priors with the fractional polynomial model, hence a random effects model would likely overestimate uncertainty.
  [redacted].

However, the ERG also had a number of concerns:

- The choice of the preferred fractional polynomial models is somewhat opaque and the ERG cannot confirm the most suitable model were selected for PFS and OS in each case.
- [redacted]

[redacted]

Time-varying hazard ratio plots (as requested in clarification question A10) were only presented for the company's chosen best fit models, as detailed above. If these plots were provided for other fractional polynomial models the ERG could have investigated whether the shape of the hazards over time were clinically plausible.

- Furthermore, only the base case fractional polynomial models were available for use in the economic model so the impact of other fractional polynomial models, some of which may present a similar fit, is uncertain.