Confidential until published

Trifluridine-tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies. A Single Technology Appraisal.

Produced by	School of Health and Related Research (ScHARR), The University of Sheffield
Authors	Matt Stevenson, Professor of Health Technology Assessment, ScHARR, University of Sheffield, Sheffield, UK
	Sue Harnan, Senior Research Fellow, ScHARR, University of Sheffield, Sheffield, UK
	Shijie Ren, Senior Research Fellow, ScHARR, University of Sheffield, Sheffield, UK
	Andrew Metry, Research Associate, ScHARR, University of Sheffield, Sheffield, UK
	Martin Orr, Research Associate, ScHARR, University of Sheffield, Sheffield, UK
	Fiona Campbell, Research Fellow, ScHARR, University of Sheffield, Sheffield, UK
	Mark Clowes, Information Specialist, ScHARR, University of Sheffield, Sheffield, UK
	Raj Sripadam, Consultant Clinical Oncologist, Clatterbridge Cancer Centre, NHS Foundation Trust, Bebington, Wirral CH63 4JY
Correspondence Author	Matt Stevenson, Professor of Health Technology Assessment, ScHARR, University of Sheffield, Sheffield, UK
Date completed	Date completed (21/08/2019)

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 12/93/72.

Declared competing interests of the authors

None of the authors has any conflicts of interest to declare.

Acknowledgements

We would like to thank Dr Jamal Hayat for clinical advice relating to this project. We would also like to thank Paul Tappenden, ScHARR, for providing comments on the draft report and Andrea Shippam, Programme Manager, ScHARR, for providing administrative support and in preparing and formatting the report.

Rider on responsibility for report

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

This report should be referenced as follows:

Stevenson M, Harnan S, Ren S, Metry A, Orr M, Campbell F, Clowes M, Sripadam R. Trifluridine-tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies.: A Single Technology Appraisal. School of Health and Related Research (ScHARR), 2019.

Contributions of authors

Matt Stevenson and Andrew Metry critiqued the health economic analysis submitted by the company. Sue Harnan and Fiona Campbell summarised and critiqued the clinical effectiveness data reported within the company's submission. Shijie Ren and Martin Orr critiqued the statistical aspects of the submission. Mark Clowes critiqued the company's search strategy. All authors were involved in drafting and commenting on the final report.

Copyright belongs to The University of Sheffield

Copyright is retained by Servier Laboratories for figures 9, 10, 11, 12, 13, 14 and 15.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

CONTENTS

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

List of tables

List of figures

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

List of Boxes

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

Abbreviations

Adverse events	AEs
AF	Acceleration Factor
AFT	Accelerated Failure Time
AIC	Akaike Information Criterion
ASCO	American Society of Clinical Oncology
ASCO GI	The American Society of Clinical Oncology Gastrointestinal Cancers Symposium
AWMSG	All Wales Medicines Strategy Group
BIC	Bayesian Information Criterion
BSA	Body Surface Area
BSC	Best Supportive Care
CEAC	Cost-Effectiveness Acceptability Curve
CS	Company Submission
CT	Computed Tomography
DSU	Decision Support Unit
eMIT	electronic Market Information Tool
EORTC QLQ-C30	European Organisation for Research and Treatment of Cancer Quality-of-Life Questionnaire Core 30
EQ-5D-3L	EuroQol 5 dimensions 3 level
ERG	Evidence Review Group
ESMO	European Society of Medical Oncology
GEE	Generalised Estimating Equation
GEJ	Gastro-Oesophageal Junction Cancer
HAS	Haute Autorité de Santé
HER2	Human Epidermal Growth Factor Receptor 2
HERC	Health Economics Research Centre
HR	Hazard Ratio
HRQoL	Health-Related Quality of Life
HTAD	Health Technology Assessment Database
ICER	Incremental Cost Effectiveness Ratio
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
KM	Kaplan-Meier
mGC	metastatic Gastric Cancer
MRU	Medical Resource Use
NHS EED	National Health Service Economic Evaluation Database

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

NICE	National Institute for Health and Care Excellence
NR	Not Reported
ORR	Overall Response Rate
OS	Overall Survival
PAS	Patient Access Scheme
PD	Progressed Disease
PF	Progression-Free
PFS	Progression-Free Survival
PH	Proportional Hazards
PSM	Partitioned Survival Model
PSSRU	Personal Social Services Research Unit
QALY	Quality-Adjusted Life Year
QIC	Quasi-likelihood under Independence Model Criterion
RCT	Randomised Controlled Trial
ROW	Rest Of World
SACT	Systematic Anti-Cancer Treatment
SLR	Systematic Literature Review
SLV	Statens Legemiddelverk
SMC	Scottish Medicines Consortium
STA	Single Technology Appraisal
TAGS	Trifluridine/tipiracil versus placebo in patients with heavily pre-treated mGC trial
TFT	Trifluridine-Tipiracil
TSD	Technical Support Document
TTD	Time To Treatment Discontinuation

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

1 SUMMARY

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

The company provided an appropriate description of metastatic gastric and gastro-oesophageal junction cancers, the current practice guidelines regarding lines of treatment and the potential positioning of trifluridine/tipiracil (TFT) (Lonsurf®) in the treatment pathway.

1.2 Summary of clinical effectiveness evidence submitted by the company

The company submitted a systematic literature review which the ERG believes identified all important studies.

The pivotal trial was TAGS, a Phase III, randomised, double-blind, placebo-controlled study in patients with heavily pre-treated metastatic gastric cancer, conducted at 110 sites in 18 countries, comparing 35mg/m2 TFT and BSC with placebo and BSC. The trial reported all key efficacy outcomes listed in the NICE scope. Overall survival (OS) was positively affected by TFT treatment with a hazard ratio (HR) of 0.69 (95% CI: 0.56–0.85) and a difference in median survival of 2.1 months between arms. Analyses adjusted for relevant prognostic factors gave a similar HRs. Progression free survival (PFS) was also positively affected by TFT treatment, with a HR 0.57 (95% CI: 0.47–0.70) and a 0.2 month difference between arms. Small benefits were reported for response rates and duration of response as may be expected given the stage of disease, however, the disease control rate was significantly improved. Health related quality of life was shown to be largely maintained with TFT treatment.

In subgroup analyses, for OS, patients with prior ramucirumab treatment had HR 0.76 (95% CI 0.53 – 1.09) and those without an HR 0.66 (95% CI 0.51 – 0.85). The HR in Japanese patients was 0.77 (95% CI 0.46 to 1.30) compared with 0.68; (95% CI: 0.54 to 0.85) in the rest of the world (ROW). The median survival in the placebo group was 5.9 in the Japanese population and 3.3 months in the ROW. The HR for European patients was 0.67 (95% CI 0.53-0.86), but no Kaplan-Meier plots were provided by the company.

The ERG requested an analysis of European patients without prior exposure to ramucirumab. The company urged caution in the interpretation of this analysis as the TAGS study was stratified on Japan versus the ROW, although the ERG comments that approximately 95% of the ROW group were European.

[redacted]

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

Key adverse events included nausea, anaemia, decreased appetite, vomiting, diarrhoea, fatigue, neutropenia and asthenia thrombocytopenia. Anaemia and neutropenia were two outcomes where the incidence appeared to be markedly greater in the TFT group compared with the placebo group (anaemia 45% vs 19%; neutropenia 53% vs 4%).

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

The ERG agreed that due to the low quality of the evidence available in the third-line setting, a network meta-analysis would have relied on strong assumptions and was not appropriate for this appraisal. Clinical advice to the ERG also indicated that chemotherapy was infrequently administered as a third-line treatment (approximately 10-15% of patients). The company did not synthesise a single-arm phase II study (EPOC1201) conducted in Japan with the pivotal TAGS study, and the ERG agreed it had low relevance to the decision problem.

Ramucirumab does not have a positive NICE recommendation, which means patients in England will largely be ramucirumab-naïve. There were mixed views from clinical advisors to the ERG and NICE about whether prior ramucirumab treatment would alter prognosis, but agreement that as ramucirumab and TFT work differently there should be no impact on treatment efficacy. In the absence of a strong indication that prior ramucirumab treatment alters prognosis, the ERG assumes there is no impact, though this is uncertain. The ERG therefore prefers an estimate of a HR or Acceleration Factor (AF) from the entire population rather than the non- ramucirumab patients only. However, the ERG notes that the non-ramucirumab population are less heavily pre-treated and their disease duration is shorter than the prior ramucirumab group.

Clinical advice to the ERG and NICE indicated that European patients have the highest generalisability to the decision problem, due to biological and/or treatment pathway differences between Europe and the USA and in particular, Japan. A subgroup analysis of European patients was reported, but no Kaplan-Meier curves were available. Exclusion of the Japanese patients from the full TAGS population leads to an under-representation of Asians compared with the English population (ERG-calculated 1% compared with 7.5% respectively), whilst their inclusion leads to over-representation (14.4%). The generalisability of Japanese patients to the more diverse Asian population in England is also unclear. The ERG concludes that analyses excluding Japanese and USA patients is preferred, although accepts that this breaks the stratification of the TAGS study.

In the requested analysis of European patients with no prior ramucirumab treatment, baseline imbalances in some prognostic characteristics were larger than in the full study population for which efficacy estimates were unadjusted.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

Other issues were also noted. There were more gastric patients than would be usual in England, though this was thought unlikely to affect estimates of efficacy. Incorrect dosing within the trial [redacted]. It was not clear whether the discontinuation rules applied in the TAGS study were mandatory, and whether these will be applied in clinical practice in the UK. Monitoring in clinical practice may occur more frequently than in the TAGS study (every 4-6 weeks rather than every 2 months) which may lead to earlier discontinuations and may impact on efficacy, adverse event rates and costs.

1.4 Summary of cost effectiveness submitted evidence by the company

The submitted economic model was clear, relatively simple and generally well programmed, with minor errors amended in the clarification process. The company submitted a partitioned survival model comprising three health states (progression free, post progression, and death). The weekly transitions between health states were inferred via extrapolated PFS and OS curves fitted to data from TAGS. Health-related quality of life (HRQoL) data were collected using the European Organisation for Research and Treatment of Cancer Quality-of-Life Questionnaire Core 30 (EORTC QLQ-C30) within the TAGS trial and then mapped to EuroQol five dimensions three-level (EQ-5D-3L) values using a published algorithm. The time horizon in the base case was 10 years, with discounting of both benefits and costs at 3.5% per annum. The company's base case results suggested that TFT was cost-effective compared with BSC at an incremental cost-effectiveness ratio (ICER) threshold of £50,000 per quality-adjusted life year (QALY) gained. The probabilistic ICER for TFT compared with BSC was £45,314 per QALY gained when treating ramucirumab-naïve patients. The ICER was sensitive to the selected parametric survival models fitted to data from combinations of geographical region and prior ramucirumab use.

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

A key difference between the approach undertaken by the company and that preferred by the ERG is related to selection of the patient population relevant to England. Based on clinical advice the ERG believes that: prior ramucirumab treatment is unlikely to affect the relative effectiveness of TFT; that there is no strong signal that prior ramucirumab treatment affects prognosis; and that the European subgroup is the most relevant to the decision problem, albeit noting the limitation in breaking stratification.

The ERG also prefers the use of independent curve fits rather than the use of dependent curve fits using a HR or an AF to account for the efficacy of TFT.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

The ERG noted the limitations in the study developing a mapping algorithm between the EORTC QLQ-C30 and the EQ5D-3L selected by the company to derive the model's base case utility values as the study involved 48 Greek patients with non-metastatic gastric cancer. The ERG requested the company to apply two alternative mapping algorithms (Versteegh *et al*. and Longworth *et al*.) to assess the impact on the ICER. The company did not consider these algorithms appropriate to inform the model citing differences in the patient populations used to derive the mapping algorithms, and instead provided an analysis calculating utility values from a mapping algorithm by Marriott *et al*.

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

1.6.1 Strengths

The clinical evidence was from a good quality phase III, randomised, double-blind, placebo-controlled study. All key efficacy and safety outcomes were reported. A good quality systematic literature review supports the submission.

The submitted mathematical model was of good quality. The company responded well to the clarification questions raised and provided a revised model and undertook the analyses requested.

1.6.2 Weaknesses and areas of uncertainty

A network meta-analysis was not possible, so the efficacy of TFT compared to other chemotherapy regimens used at third-line is unclear. Clinicians estimate between 10-15% of patients may receive chemotherapy at third-line in England.

It is unknown which subgroup's results, are of most relevance for the purpose of decision making. Subgroups of interest include those based on the prior use of ramucirumab treatment and geographical region. In addition, the TAGS study did not collect EQ-5D data and all mappings between the EORTC QLQ-C30 and the EQ5D-3L have limitations.

1.7 Summary of exploratory and sensitivity analyses undertaken by the ERG

As stated in Section 1.5, the ERG preferred alternative assumptions in the base case on multiple occasions to the company. The ERG explored relationships between both prior ramucirumab treatment and disease prognosis and prior ramucirumab treatment and the relative efficacy of TFT. The relationships were explored in both the whole population and the European cohort of the TAGS trial. The possible permutations resulted in eight scenarios, each of which was explored by fitting alternative parametric survival distributions. The ERG proffered a tentative base case but this could not be evaluated as the data were not available to the ERG.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

Based on the analyses provided by the company and the ERG's exploratory analyses the ERG believes that the cost per QALY gained of TFT compared with BSC is likely to be in excess of £50,000. Whilst, the ERG's tentatively preferred scenario could not be evaluated, many component factors such as: using independent curves; assuming that prior ramucirumab use does not affect prognosis; assuming that prior ramucirumab use does not affect the relative treatment effect of TFT; using a European population; and reducing utility values, all increase the ICER. The ERG notes that some of these factors, in isolation, increase the ICER to greater than £50,000 per QALY gained.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

2 BACKGROUND

2.1 Critique of company's description of underlying health problem

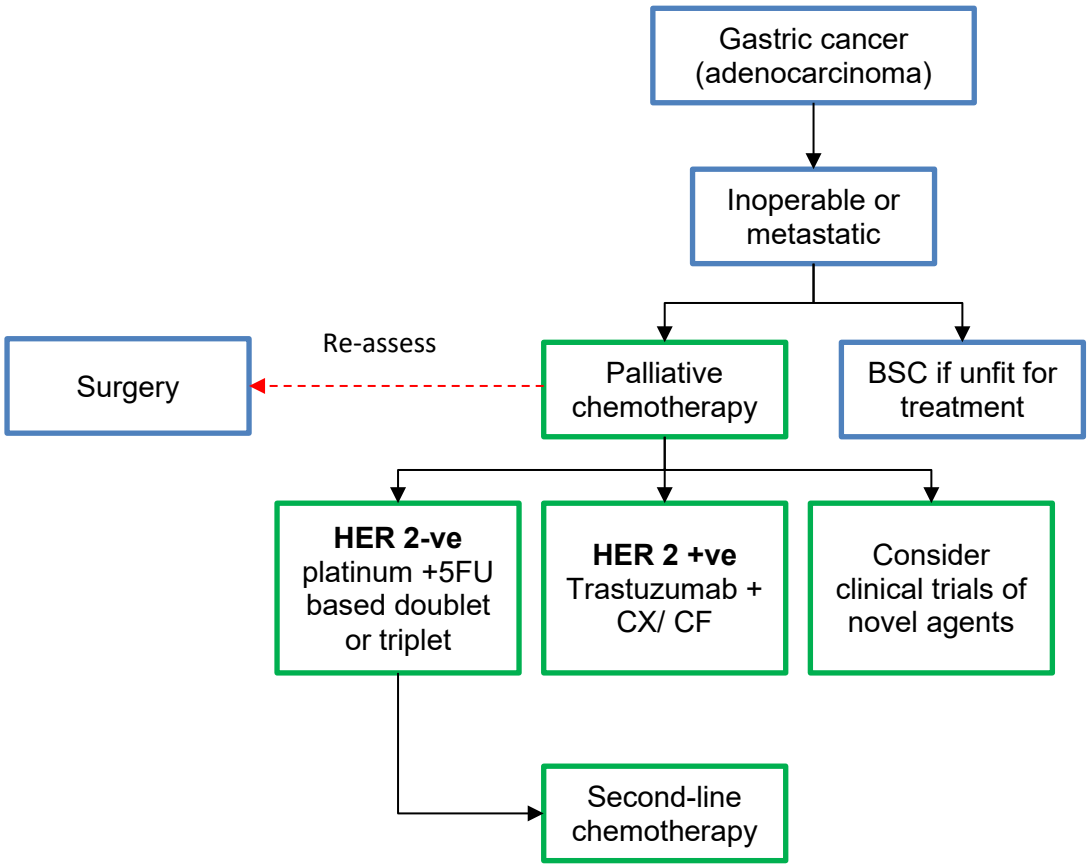
Gastro-oesophageal cancers are malignant tumours characterised by uncontrolled cell growth in the tissues of the stomach. Such cancers rapidly progress and have significant impacts on patients. A detailed description of the epidemiology, risk factors, prognoses, diagnosis and methods to stage the severity of gastro-oesophageal cancers has been provided within the company submission (CS).[1]

2.2 Critique of company's overview of current service provision

The CS detailed the typical treatment pathway for patients with metastatic gastric cancer, and provided a schematic of current European Society of Medical Oncology (ESMO) guidelines. This has been reproduced in Figure 1, although this diagram appears to indicate that patients who are human epidermal growth factor receptor 2 (HER2) positive would not receive a second line of treatment, which is incorrect. The company stated that the majority of patients (95%) within the UK would receive a doublet regimen rather than a triplet regimen. Second-line treatment chemotherapy is provided after disease progression or recurrence. Few patients are deemed fit enough to receive third-line treatment and few treatments provide a meaningful benefit over best supportive care (BSC). Prognosis is very poor with the average survival time for patients in the third-line setting being less than six months.[2-4] NICE does not currently recommend any third-line treatment for metastatic gastric or gastro-oesophageal junction cancer.

It is worth noting that ramucirumab received marketing approval in 2015 for patients who had received previous chemotherapy although this was not recommended by NICE. However, within the multi-national pivotal study for trifluridine-tipiracil (TFT) (TAGS)[2] a proportion of patients had received ramucirumab, and the level to which these data are generalisable to England is unclear.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved



Key: BSC, Best supportive care; CF: cisplatin and 5-fluorouracil; CX: cisplatin and capecitabine; ECF: epirubicin, cisplatin and 5-fluorouracil; ECX: epirubicin, cisplatin and capecitabine; EOF: epirubicin, oxaliplatin and 5-fluorouracil; EOX: epirubicin, oxaliplatin and capecitabine; DCF: docetaxel, cisplatin and 5-fluorouracil; ESMO: European Society for Medical Oncology; FOLFIRI: folinic acid, fluorouracil and irinotecan; HER2 +ve/-ve, Human epidermal growth factor receptor 2 negative/positive.

Note: Doublet combinations of platinum and fluoropyrimidines are generally used, but triplet regimen options also include: ECF, ECX, EOF, EOX, DCF or FOLFIRI. Please note this is also the treatment pathway for inoperable advanced disease

Figure 1: The treatment pathway for metastatic gastric cancer provided by the company based on ESMO guidelines

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

2.3 Critique of company's definition of the decision problem

2.3.1 Population

The population within the decision problem matches that within the NICE scope[5] in considering patients with metastatic gastric cancer (mGC) or gastro-oesophageal junction cancer (GEJ) who have received two previous regimens of treatment. The TAGS study was multi-national which may mean that some sub-group data from the study may be more generalisable to England than the data from the entire study population. The company provided evidence on both a subgroup of the population that had no prior ramucirumab and on the full TAGS study population.

2.3.2 Intervention

The intervention matches that of the final NICE scope,[5] which is the use of TFT, a novel oral cytotoxic chemotherapy, in combination with BSC. TFT currently has a marketing authorisation for use in metastatic colorectal cancer, with an extension for mGC and GEJ expected by November 2019. The dose of TFT is dependent on body surface area (BSA) with a recommended starting dose of 35mg/m^2 administered orally twice daily on days 1-5 and 8-12 of each 28-day cycle. The minimum dose is 20 mg/m^2, with a maximum of 80mg.

2.3.3 Comparators

The comparators listed in the final NICE scope[5] are chemotherapy and BSC. The company have not included chemotherapy in the decision problem, stating that this deviation was based on "*current guidelines, a systematic literature review and expert opinion validating the lack of an evidence-based active chemotherapy option in the third-line setting.*" Clinical advice provided to the ERG supported the company's view that there is no established treatment for mGC or GEJ following two previous treatment regimens.

2.3.4 Outcomes

The outcomes in the CS are in line with those in the final scope issued by NICE.[5]

2.3.5 Other relevant factors

TFT has a patient access scheme (PAS) in place related to the treatment of metastatic colorectal cancer, which is a simple discount of [redacted]. This discount is also applicable to TFT for the use in mGC and GEJ. TFT is linearly priced with a pack of 20 15mg/6.14mg tablets costing £500 at list price and a pack of 60 20mg/8.19mg tablets costing £2000 at list price.[6]

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

3 CLINICAL EFFECTIVENESS

3.1 Critique of the methods of review(s)

The company presented a systematic review aiming to "*identify the available clinical efficacy, safety and tolerability evidence related to the third- or later-line treatment of advanced and/or metastatic GC (including gastroesophageal junction cancer [GOJ/GEJ]*)." (p10 of systematic literature review (SLR) report). An initial search was conducted until 27th June 2018 with an update searching to 28th February 2019 (CS Appendix D).

The CS suggests there are no standard comparators in the third-line setting, and that very few patients would receive anything other than palliative care at this point. As such, whilst the NICE scope includes chemotherapies as comparators, these are not contained in the company's economic analysis. The clinical advisors to the ERG agreed that they are not relevant comparators for the majority of patients. For the sake of transparency and completeness, the review of chemotherapy agents in a third-line setting is of some (but not crucial) relevance, and is discussed briefly in Section 3.1.53.1.5.

3.1.1 Searches

The evidence searches for the SLR report and subsequent update conducted by Servier Laboratories are reported in Appendices D1.1 and D1.2 of the CS, respectively. Both reviews used similar methods of identifying evidence, with only minor differences between the search strategies.

Searches covered all of the key databases recommended by NICE (MEDLINE including Medline-in-Process, Embase and Cochrane) plus two years' worth of relevant conference proceedings and HTA websites. For both the SLR report and the update, MEDLINE and Embase were searched simultaneously in a "multi-file" search (on Embase.com). This technique is not usually recommended since it limits the ability to optimise the strategy for each source.

Searches are generally well-designed and executed and include both Emtree subject headings and free text terms, with appropriate use of truncation and proximity search strings to increase the sensitivity.

The ERG noted some logical errors in Servier's update searches (a missing line in one search strategy, and incorrectly numbered lines in another) and queried these with the company (Clarification question, A6[7]). The company acknowledged the errors, blaming them on an attempt to re-format the search strategies for aesthetic reasons, but gave assurance that the actual searches were conducted correctly. (Clarification response, A6). On this basis, the ERG does not believe any relevant studies are likely to have been missed.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

3.1.2 Inclusion criteria

The inclusion criteria are reported in Table 6 of the CS. These appear appropriate to the ERG, although one of NICE's decision problem-defined outcomes (duration of response) was omitted. However, this does appear to have been data extracted and reported in the SLR report.

For the original review, study selection was conducted by two independent reviewers, which is a high quality methodology.[8] For the update, only one reviewer conducted study selection, leaving the update at some risk of bias and error, though the extent to which this operated is unclear.

3.1.3 Critique of data extraction

For the original review, the company used a pre-agreed data extraction form, although it was not clear if this was piloted and tailored to the specifics of the review. Data were extracted by one reviewer and checked by a second, which is likely to result in reliable data extraction. For the update, only one reviewer extracted data, although it is not clear if this was into a data extraction form, or directly into data tables for presentation in the report. Data extraction in the update is therefore at some risk of bias and error; however, the extent to which this operated is unclear. This may be important as the results from the pivotal trial (Shitara *et al.* 2018[2]) were obtained from the updated search. The ERG checked key data and these were generally correct, though a few minor inconsistencies between the CSR, Shitara *et al.* 2018[2] and the CS were noted.

The information provided by the company in Appendix D and the SLR report appears to be complete, relevant, and to provide an appropriate level of detail. However, the ERG notes some possible mistakes in the original review. For example, in Table 5 of the SLR report, column 2 states that studies in all patients were at third or later lines, but column 5 shows one study (Li et al. 2016[9]) included patients at second- and third- lines of therapy. This is further at odds with a statement that "*The review identified only four RCTs that were solely conducted in GC patients receiving treatment in third- or further-line therapy.*" (p19, SLR report).

3.1.4 Quality assessment

Quality assessment for all included studies was provided in the SLR report. It used the items listed in the NICE user guide for evidence submissions[10] and was conducted by one reviewer and checked by a second. The CS and Appendix D of the CS focussed only on the TAGS study. The quality assessment in the CS was similar to the SLR assessment, but added reasons for scores for some items. It was unclear if the reasons for scores were checked by a second reviewer, as they were not presented in the SLR report. Reasons for scores were missing for some items, but were provided in the clarification response.[7]

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

Study quality was also assessed using the Cochrane Risk of Bias 2 (RoB2) tool[11] (Appendix D of the CS), but items relating to risk of bias due to deviations from the intended interventions were not scored clearly, as all options remained in column three. Clarification of the scores were requested by the ERG. This assessment appears to have been conducted by one reviewer and not checked. The company provided scores and reasons in their clarification response, with the explanation that RoB2 automatically greyed out some items due to answers to earlier questions. The answers relating to protocol deviations have been incorporated into Section 4.2.2.3.

The quality assessment of the key TAGS trial, comparing the CS scores with the ERG's own scores, is provided in Table 1. The company scored the trial at low risk of bias for all items. The ERG had concerns about imbalances between treatment arms in potentially prognostic baseline characteristics, although these are addressed for overall survival (OS) by an analysis adjusting for key factors. The ERG also noted some small imbalances in treatment discontinuation, withdrawals and loss to follow-up, but that these are unlikely to affect results greatly.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

Table 1: Summary of risk of bias using the items listed in the NICE user guide for evidence submissions[10], as judged in the CS and by the ERG

Question	Company's score (Yes/No/Unclear) with reason (From CS[1])	ERG's score (Yes/No/Unclear) with reason (based on Shitara 2018[2])
Was randomisation carried out appropriately?	Yes Patients were enrolled by study investigators. Eligible patients were randomised (2:1) to trifluridine/tipiracil plus BSC or placebo plus BSC via a dynamic allocation method (biased coin) with an interactive-voice web-response system (IXRS). Almac (Craigavon, UK) operated the IXRS and created the algorithm that generated the individual patient allocation when the study site accessed the system. The company had no other role in the trial. Once a patient's eligibility was confirmed and the criteria for randomisation were met, study-site personnel logged on to the IXRS to allocate patients to treatment. The IXRS randomly assigned study medication (trifluridine/tipiracil or placebo) by	Yes As per the company's response in column 2

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

	assigning a kit number to that patient. Randomisation was stratified by region (Japan vs rest of world), ECOG performance status (0 vs 1), and previous treatment with ramucirumab (yes vs no).	
Was the concealment of treatment allocation adequate?	Yes Patients, investigators and study-site personnel, those assessing outcomes, and those analysing the data were masked to treatment assignment. Tablets of identical appearance were used to maintain masking. Only personnel from the contract research organisations involved in drug labelling and distribution (Fisher Clinical Services [Allentown, PA, USA] and Bell Medical Solutions [Tokyo, Japan]) and IXRS activities (Almac) were aware of treatment assignment.	Yes Almac (Craigavon, UK had no other role in the trial other than randomisation. Once a patient's eligibility was confirmed and the criteria for randomisation were met, study-site personnel logged on to the IXRS to allocate patients to treatment. The IXRS randomly assigned study medication (trifluridine/tipiracil or placebo) by assigning a kit number to that patient. (CS, p40) As per column 2.
Were the groups similar at the outset of the study in	Yes	Unclear - It is unclear if the small imbalances affected results in unadjusted analyses. Clinical advisors to the ERG stated that ECOG status, number of metastatic sites, HER2 status, and previous chemotherapy regimens (number and type) are prognostic of survival

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

terms of prognostic factors?	Baseline demographic and disease characteristics were generally balanced between the two treatment arms.	in the third-line setting, and that the impact of sex and ethnicity is uncertain because there is little data. Of these, there was some evidence of imbalance in some factors (TFT compared with placebo) (see list in Section 4.2.2.1) Clinical advisors to the ERG were not too concerned at these imbalances. The ERG notes that an analysis for OS, which adjusted for these factors, was presented, but an equivalent analysis for progression-free survival (PFS) and other outcomes was not provided.
Were the care providers, participants and outcome assessors blind to treatment allocation?	Yes Please see above regarding concealment of treatment allocation.	Yes As per the company's response in column 2
Were there any unexpected imbalances in drop-outs between groups?	No Reason from clarification response A14: *"The majority of patients who discontinued, did so due to progressive disease. In the trifluridine/tipiracil group, nine of the 11 deaths resulting in treatment discontinuation were attributed to disease progression (the cause of the*	Unclear - It is unclear whether small imbalances in drop outs are due to patients being "at risk" for longer. Patients in the TFT arm were slightly more likely to stop treatment due to withdrawal of consent (4.2% vs 3.6%); adverse events (9.8% vs 6.5%); physician decision (3.3% vs 1.8%). [redacted] but not reported for OS.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

	other two deaths was septic shock that was judged to be unrelated to treatment).”	
Is there any evidence to suggest that the authors measured more outcomes than they reported?	*No* Reason from clarification response A14: *“There are no outcomes that are not accounted for in the clinical study report.”*	No All outcomes are reported in the CSR.
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, an ITT analysis was performed. For OS, in the absence of death confirmation or for patients alive as of the OS cut-off date (30th April, 2018), the survival time was censored at the date of last study follow-up or the cut-off date, whichever was earlier. For PFS, patients who were alive with no disease progression as of the analysis cut-off date (31st March, 2018) were censored at the date of the last tumour assessment. Patients who received non-study cancer treatment before disease progression were censored at the date of the last evaluable tumour assessment before the non-study cancer treatment was initiated.

Adapted from Systematic reviews: CRD’s guidance for undertaking reviews in health care (University of York Centre for Reviews and Dissemination)

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

3.1.5 Evidence synthesis

There is some lack of clarity around the total number of included studies across the original and update reviews, but only four RCTs[2, 4, 9, 12] were in solely third-line (or later) patients, or reported the results for this subgroup separately (CS p29). The four RCTs related to TFT, apatinib and nivolumab. As apatinib and nivolumab are not licensed in England and cannot therefore be considered as comparators, only the study relating to TFT is relevant to the decision problem.

Additionally, the SLR report found two non-randomised trials and fourteen single arm studies, but the majority of these were not of treatments licensed in the England. Any attempt to compare these results to the TFT results would involve strong assumptions and are likely to produce highly uncertain results, with most studies having fewer than fifty participants. Due to the lack of data and the rarity of use of chemotherapy in the third-line setting, the ERG agrees that performing a network meta-analysis would not be useful for this appraisal.

A single arm study of 35mg/m^2 and 40mg/m^2 TFT from Japan (EPOC1201)[13] was not synthesised with the TAGS study.[2] The CS states that this is because some patients in EPOC1201 had only one prior round of chemotherapy, and some had the wrong dose of TFT (40mg/m^2 TFT).The company argues elsewhere in the CS that gastric cancer operates quite differently in Japanese patients "*Due to biological differences between gastric cancer in Asian versus non-Asian patients efficacy of these treatments* [apatinib and nivolumab] *in European patients is uncertain, as recognised by ESMO and JSMO, who developed a Pan-Asian adapted ESMO clinical practice guideline for the management of patients with mGC."* (p66 of the CS). Evidence from the TAGS trial may support this in that Japanese patients in the placebo group had a median OS of 5.9 months compared with 3.3 months in patients from the EU and US. Whilst these latter two issues could potentially have been overcome through subgroup analyses, the ERG agrees that due to biological and clinical practice differences, the study has low relevance to the decision problem and meta-analysis was not necessary.

As such, the only study presented in detail in the CS is the TAGS study of TFT and no synthesis was performed.

3.2 Critique of the TAGS study, its analysis and interpretation

The pivotal trial for TFT is the TAGS study (Shitara *et al.*[14]). TAGS was a Phase III, randomised, double-blind, placebo-controlled study in patients with heavily pre-treated mGC, conducted at 110 sites in 18 countries, comparing 35mg/m^2 TFT and BSC with placebo and BSC.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

As discussed in Section 3.1.5, a single-armed study of TFT (EPOC1201)[13] was not reported in detail in the CS which the ERG deems appropriate. However, adverse event data was of interest to the appraisal and is included in Section 3.2.2.8.

The ERG verified that no other important studies were missed with a focussed search in Pubmed and citation searching in Google Scholar of TAGS and EPOC1201 key publications.[2, 13]

3.2.1 Study design: The TAGS study[2]

A table detailing the design of the TAGS study was provided in the CS (Table 7, p34)[1] and an adapted version is reproduced here for reference (Table 2).

The TAGS study mostly matched the decision problem specified by NICE (see Table 2). Clinical advice to the ERG indicated that the study was broadly in accordance with English populations and practice as detailed below, in sections *Population, Intervention, Comparator, Outcomes.*

A critical appraisal of the TAGS study is provided in Section 3.1.4.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

Table 2: Key design features of TAGS trial,[2] adapted from Table 7 of the CS[1]

Study	NCT02500043 TFT versus placebo in patients with heavily pre-treated mGC (TAGS): a randomised, double-blind, placebo-controlled, phase III trial Shitara K *et al*. Lancet Oncol 2018.[2]				
Study design	Phase III, randomised, double-blind, placebo-controlled study at 110 sites in 17 countries* to evaluate the efficacy and safety of TFT versus placebo in patients with previously treated mGC. Countries included in the European dataset: Belarus, Belgium, Czech Republic, France, Germany, Ireland, Israel, Italy, Poland, Portugal, Romania, Russia, Spain, Turkey, the UK Countries included in the USA dataset: the USA Country included in the Japanese dataset: Japan				
Population	507 patients were enrolled and randomly assigned, 337 to the TFT group and 170 to the placebo group. All patients were aged 18 or older with histologically confirmed, non-resectable, metastatic gastric adenocarcinoma (including adenocarcinoma of the GEJ) who had undergone two* previous chemotherapy regimens (and had experienced radiological disease progression) that contained fluoropyrimidine, platinum agents, and taxanes or irinotecan.				
Intervention(s)	Oral TFT (35 mg/m² twice daily on days 1–5 and days 8–12 every 28 days) plus BSC				
Comparator(s)	Placebo plus BSC				
Indicate if trial supports application for marketing authorisation	Yes	✓	Indicate if trial used in the economic model	Yes	✓
	No			No	
Rationale for use/non-use in the model	Pivotal phase III RCT.				
Reported outcomes specified in the decision problem	OS PFS Disease control rate Objective response rate HRQoL Safety				
Other reported outcome	Time to ECOG PS ≥ 2				

* There was some confusion over whether 18 (stated in Table 7 of the CS)[1] or 17 (stated on p41 of the CS[1] [redacted])[14] countries were included in the study, and whether patients had to have undergone one (stated in Table 7 of the CS, p34) or two (stated in journal article[2] [redacted])[14] previous chemotherapy regimens. The ERG has assumed that [redacted] on both matters.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

Population

The inclusion criteria for the TAGS study were very detailed (see Table 8 of the CS, p36), including previous types and number of rounds of chemotherapy, time since progression, reason for cessation of previous therapy (toxicity or refractory disease), and various other criteria. The exclusion criteria were also very detailed, and excluded patients on the basis of co-morbidities, previous treatment with TFT, pregnancy and other criteria. Recruitment was stratified by region of the world (Japan versus the rest of the world); ECOG performance status (0 versus 1); and prior treatment with ramucirumab (yes versus no). The clinical advice provided to the ERG suggested that the criteria should result in a population broadly in line with patients in England and that the stratification factors were appropriate, though there are other prognostic factors that could have been considered (see Section 3.2.2.1).

Intervention

The intervention matches the proposed license for oral TFT (35 mg/m² twice daily on days 1–5 and days 8–12 every 28 days) plus best supportive care. It is to be given for as long as "benefit" is being gained. In their clarification response (A11)[7], the company defines discontinuation reasons as patient request, disease progression, clinical progression, adverse events, physician's decision or pregnancy. It was not entirely clear whether the discontinuation rules described in A11[7] were mandatory or optional, and their application in a clinical setting may differ from that in a trial setting, as the drug has not been licensed yet, and a draft SPC was not included in the CS, nor requested by the ERG in the clarification process.

The CSR details specific conditions for dose reductions and resumption of treatment in Section 9.4.6, and summaries are given in the CS on p41 and p116. Patients who did not achieve the minimal criteria for resumption had treatment discontinued. Patients with haematological toxicity could have doses withheld until neutrophil and platelet levels returned to an acceptable level.

Comparator

The comparator was placebo twice daily plus BSC on days 1–5 and days 8–12 of each 28-day treatment cycle.

Clinical advice provided to the ERG indicated that patients under NHS care in the UK are likely to have better community care than some in other European countries, where the standard of care is more variable. This was thought unlikely to affect survival, but might mean patients receiving best supportive care in the UK have a higher HRQoL than patients in other European countries in the trial.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

Outcomes

A summary of the outcomes included in the TAGS trial, their definition and statistical analysis is provided in Table 3. The protocol-defined outcomes for TAGS that were also listed in the NICE scope were: OS; PFS; (objective) response rate; HRQoL; and safety. The TAGS study also measured disease control rate (DCR), a composite of complete response (CR), partial response (PR) and stable disease (SD), which could be classed as "response rate"; and time to ECOG PS ≥ 2, which was not listed in the NICE scope. Patients had a computed tomography (CT) scan at baseline and then every 8 weeks until disease progression.

The NICE scope listed duration of response, which was not provided in the CS, but is [redacted]. Data were provided in the company's clarification response, though the ERG was not able to identify a definition of duration of response in the CSR, the CS or the clarification response.

Adverse event data was not included in the CS for (EPOC1201)[13], though it was reported in the SLR report and the EPOC1201 study publication.[13]

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

Table 3: Summary of outcomes measured in TAGS and their relevance to the NICE scope. Collated from text in the CS, the CSR and the clarification responses

Outcome	Definition & Statistical analysis
Outcomes listed in NICE scope	
Primary outcome	
Overall survival (OS)	Time from date of randomisation to death. In the absence of death confirmation or where patients still alive, data censored at date of last study follow-up or the cut-off date (30th April, 2018), whichever was earlier. **Statistical analysis:** OS and radiologically confirmed PFS were analysed in the ITT population with a one-sided stratified log-rank test, with the Hazard Ratio (HR) and two-sided 95% CIs based on a prespecified stratified (region, ECOG performance status, ramucirumab exposure) Cox model and associated Kaplan-Meier survival estimates.
Key secondary outcome	
Progression free survival (PFS)	Time from date of randomisation until investigator-assessed radiological progression or death. Patients alive with no progression at analysis cut-off date (31st March, 2018), patients who received non-study cancer drug before disease progression, and those who had clinical but not radiological progression were censored at last tumour assessment. **Statistical analysis:** As for OS. patients were censored for (1) discontinued follow-up, (2) follow-up ongoing at the time of analysis, (3) missed visit (>91 days since last response), and (4) initiated anti-tumour therapy (Additional clarification response A2[7]).
Other secondary outcomes	
Objective response rate (ORR)	Objective CR or PR based on investigator review of radiological images and following RECIST criteria (version 1.1, 2009). Best overall response was best recorded response after randomisation but before disease progression or initiation of non-study cancer treatment. Best response of SD needed to be maintained for 6 weeks after randomisation. Best response of CR or PR did not have a minimum time limit, as per RECIST 1.1.
Disease control rate (DCR)	Proportion of patients with objective evidence of CR, PR or SD. **Statistical analysis:** NR

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

Duration of response	Not defined in the CS or the CSR, or the clarification response. **Statistical analysis:** Not defined.
HRQoL	EORTC QLQ-C30 – measures 5 functional scales (physical, role, cognitive, emotional, and social), 3 symptom scales (fatigue, pain, and nausea and vomiting), a global health status scale, and a number of single items assessing additional symptoms commonly reported by cancer patients (dyspnoea, loss of appetite, insomnia, constipation and diarrhoea) and perceived financial impact of the disease. Gastric Specific Module (QLQ-ST022) - 22-item instrument used alongside the 30-item QLQ-C30 core questionnaire, resulting in a total of 52 items. **Statistical analysis:** Based on Osoba *et al*. 1998,[15] for both questionnaires a mean change from baseline of at least 10 points was considered to be clinically relevant for the patients.
Adverse events	Graded according to the US National Cancer Institute's Common Terminology Criteria for Adverse Events (version 4.03) and recorded from the first dose of study drug (that is, day 1, cycle 1) until 30 days after the last dose of study drug. Includes haematology, serum chemistry and urinalysis. **Statistical analysis:** NR
Outcomes not listed in the NICE scope	
Time to Deterioration of ECOG Performance Status	Time from randomisation to ECOG performance status score of two or higher. **Statistical analysis:** NR

OS, overall survival; PFS, progression-free survival; CR, complete response; PR, partial response; SD, stable disease; NR, not reported.

Planned analyses

There were three analysis sets:

- **Intention to treat (ITT) population** – the primary population for all efficacy outcomes, included all randomised patients, according to treatment assigned at randomisation.
- **As-treated (AT) population**– patients who took any dose of study treatment, analysed according to the treatment they received. Used for the safety analyses.
- **Tumour-response (TR) evaluable population** – an ITT analysis only including patients with measurable lesions.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

There were a number of subgroup and "supportive" analyses planned (CSR p56-58). Of most relevance were:

OS

- [redacted] The company did not provide the SAP, which was not identified by the ERG until after the clarification round.
- Multivariate analyses including the stratification factors and potential prognostic/predictive factors (age group (< 65, ≥ 65 years); race (White, Asian, other); gender; number of prior regimens (≤ 2, ≥ 3); prior therapy (taxane, irinotecan); previous gastrectomy; gastroesophageal junction involvement; presence of peritoneal metastases; presence of liver metastases; presence of lung metastases; number of metastatic sites (1-2, ≥ 3); [redacted] histology subtype (diffuse, intestinal), and human epidermal growth factor receptor (HER)2 status at baseline.)

PFS

- Various analyses including clinical progression, radiological progression, initiation of non-study drug as a PFS events and [redacted]

There was a lack of detail in the CS relating to the analysis plans for some outcomes (see Table 3), especially relating to how missing data was handled. Stratification factors were adjusted in the analysis of OS and PFS, but whilst other potential prognostic or predictive factors were adjusted in an additional analysis for OS, a similar analysis was not presented for PFS. For PFS patients were censored for (1) discontinued follow-up, (2) follow-up ongoing at the time of analysis, (3) missed visit (>91 days since last response), and (4) initiated anti-tumour therapy (Additional clarification response A2[7]). [redacted]

3.2.2 Study results: The TAGS study[2]

3.2.2.1 Baseline characteristics

The baseline characteristics of patients in the study are provided in Table 4 (reproduction of Table 9 from the CS). Clinical advice provided to the ERG suggests that most proportions were in line with the patient population in England. One exception was that clinicians would expect the ratio of patients with gastric compared with gastroesophageal cancer to be around 40:60 in the third-line setting, as gastroesophageal cancer is generally more common. However, in the study, the ratio was roughly 71:29, indicating more gastric cancer patients than might be expect in an English population. Clinical advice provided to the ERG indicated that the two cancers behave similarly; however, the ERG notes that the graph of OS by subgroup presented in Shitara *et al.*[2] shows that the point estimate of the hazard ratio

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

for those in the gastric cancer subgroup is more favourable (0.67 (95% CI 0.52 to 0.87) than for those with gastroesophageal cancer (0.75 (95% CI 0.50 to 1.11), although the confidence intervals overlap considerably. The impact of including more patients with gastric cancer on estimates of efficacy is unclear.

The trial includes around 62% of patients who have had three or more regimens of chemotherapy. This appears to be at odds with clinical advice provided to the ERG, which indicated that most patients in England do not get third-line therapy as by this stage, they are too ill and the burden of treatment outweighs the benefits. This view is echoed in the clinical statements submitted for this appraisal,[16, 17] where clinicians estimate that only 10-15% of patients have third-line therapy. However, the subgroup analysis in Figure 3 of Shitara *et al.*[2] indicates that previous lines of therapy (2, 3 or ≥4) did not impact much on the point estimates of the hazard ratios for each subgroup so this may not impact much on estimates of efficacy. It may, however, impact negatively on median survival compared to an English population at third-line therapy, as the prior ramucirumab population had more prior lines of treatment than the no prior ramucirumab population.

Table 4 Patient baseline characteristics. Reproduction, with correction of errors according to CSR, of Table 9 from the CS

	All regions		[redacted]	
	TFT (n=337)	**Placebo (n=170)**	[redacted]	[redacted]
Age (years)			[redacted]	[redacted]
Median (range*)	64.0 (24–89)	62.5 (32–82)		
<65	183 (54%)	96 (56%)		
≥65	154 (46%)	74 (44%)		
Sex			[redacted]	[redacted]
Male	252 (75%)	117 (69%)		
Female	85 (25%)	53 (31%)		
Ethnicity			[redacted]	[redacted]
White	244 (72%)	113 (66%)		
Asian	51 (15%)	29 (17%)		
Other	4 (1%)	4 (2%)		
Not available	38 (11%)	24 (14%)		
Region			[redacted]	[redacted]
USA	21 (6%)	5 (3%)		
Europe**	270 (80%)	138 (81%)		
Japan	46 (14%)	27 (16%)		
ECOG performance status			[redacted]	[redacted]
0	123 (36%)	68 (40%)		
1	214 (64%)	102 (60%)		

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

	All regions		[redacted]	
	TFT (n=337)	**Placebo (n=170)**	[redacted]	[redacted]
Primary site			[redacted]	[redacted]
Gastric	239 (71%)	121 (71%)		
GEJ	98 (29%)	47 (28%)		
Both	0	2 (1)		
Measurable disease	306 (91%)	150 (88%)	[redacted]	[redacted]
Histology			[redacted]	[redacted]
Diffused	53 (16%)	21 (12%)		
Intestinal	103 (31%)	52 (31%)		
Mixed	14 (4%)	8 (5%)		
Unknown	132 (39%)	69 (41%)		
Not available	35 (10%)	20 (12%)		
HER2 status			[redacted]	[redacted]
Positive	67 (20%)	27 (16%)		
Negative	207 (61%)	106 (62%)		
Not assessed	62 (18%)***	37 (22%)		
No. of metastatic sites			[redacted]	[redacted]
1–2	155 (46%)	72 (42%)		
≥3	182 (54%)	98 (58%)		
Peritoneal metastases	87 (26%)	53 (31%)	[redacted]	[redacted]
Previous gastrectomy	147 (44%)	74 (44%)	[redacted]	[redacted]
No. of prior regimens			[redacted]	[redacted]
2	126 (37%)	64 (38%)		
3	134 (40%)	60 (35%)		
≥4	77 (23%)	46 (27%)		
Prior systemic cancer therapeutic agents			[redacted]	[redacted]
Platinum	337 (100%)	170 (100%)		
Fluoropyrimidine	336 (>99%[a])	170 (100%)		
Taxane[†]	311 (92%)	148 (87%)		
Irinotecan[‡]	183 (54%)	98 (58%)		
Ramucirumab	114 (34%)	55 (32%)		
Anti-HER2 therapy	60 (18%)	24 (14%)		
Immunotherapy (anti–PD-1/PD-L1)	25 (7%)	7 (4%)		
Other	77 (23%)	41 (24%)		

ECOG PS: Eastern Cooperative Oncology Group performance status; HER2: human epidermal growth factor receptor 2; PD-1: programmed death-1; PD-L1: programmed death-ligand 1

Note: Data are n (%) unless noted otherwise. *Please note that Europe refers to Belarus, Belgium, Czech Republic, France, Germany, Ireland, Israel, Italy, Poland, Portugal, Romania, Russia, Spain, Turkey, and the UK; †One patient did not receive a fluoropyrimidine; ‡All patients received irinotecan or taxane or both.

* this was given as IQR in the CS, but range in the CSR. The IQR in Shitara[2] suggests the CSR is correct.

** Servier could not identify these values at the time of clarification as derivation requires further interrogation of patient level data from the TAGS trial.

*** This was given as 62 (18%) in the CS, but 63 (19%) in Shitara.[2] Shitara has been preferred so the total is 337 patients.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

3.2.2.2 The potential impact of prior ramucirumab

Because ramucirumab does not have a positive NICE recommendation, the proportion of patients with prior exposure to ramucirumab is unlikely to reflect that of patients in England. This adds uncertainty to whether the full results from the TAGS study are generalisable to England, as firstly, the prognosis of patients who have received ramucirumab may be different from those that did not. Secondly, the relative efficacy of TFT may differ in patients who received ramucirumab and those that did not. These are discussed in turn.

The relative prognosis in patients with and without prior ramucirumab treatment is unknown. Clinical advice provided to the ERG was that prior ramucirumab use was unlikely to influence the natural history of the cancer. However, a view expressed in a clinical expert statement provided to NICE was that it may possibly mean that patients are in a "*better state*" although this observation was acknowledged to be anecdotal. Without a strong indication that prior ramucirumab treatment alters prognosis, the ERG prefers to assume that there is no impact associated with prior ramucirumab treatment but notes differences in the prior ramucirumab group and the no ramucirumab group in terms of prior lines of treatment and disease duration. Therefore, this assumption is uncertain.

In a further clinical expert statement provided to NICE it was commented, "*there is no reason why prior ramucirumab would alter the outcome for trifluridine-tipiracil. They work on completely different pathways and cross resistance would not be expected.*" As such, whilst the relative efficacy of TFT in patients with and without prior ramucirumab treatment is unknown, the ERG prefers an estimate of a HR or AF from the entire population rather than from only patients who had not received ramucirumab.

3.2.2.3 The potential impact of geographical region

Some recruitment took place in Japan, and the proportion of Japanese patients in TAGS was 14% in the TFT arm and 16% in the placebo arm respectively. Clinical advice provided to the ERG indicated that EU patients are likely to have the greatest generalisability to England, as disease prognosis and treatment practices are more similar within the EU than in Japan. This view is echoed in the CS (see Section 3.1.5). Being recruited in Japan, compared with recruitment in Europe or the USA was stratified for at baseline in the TAGS study, suggesting that the study investigators believed that being Japanese could affect the efficacy of TFT compared to that for EU or USA patients.

In its clarification response, the company asserted that Japanese patients should be included because England has an 8% Asian population (clarification response A22[7]), and that for their base-case (the no prior ramucirumab population) there were [redacted] Japanese patients. The most recently available census data for England and Wales (2011[18]) indicates that 7.5% of the population was Asian, and the majority were of Indian (approximately 3%) or Pakistani (approximately 2.5%) ethnicity. People of Japanese

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

ethnicity were not reported separately, but probably included in the category of "other Asian" (approximately 1.5%). The ERG notes that this makes the Asian population in TAGS around double that of the English population and in the company's base-case (no prior ramucirumab) around half that of the English population. In both the whole trial and the no prior ramucirumab group, there are likely to be a higher proportion of Japanese patients than is found in England and the generalisability of Japanese patients to the broader category of "Asians" is unclear. The exclusion of the patients recruited in Japan and the USA leaves a [redacted] (clarification response A3[7]), which is also an under-representation of Asians within the trial results compared to the English population. The ERG concludes that whilst exclusion of the Japanese patients from the whole trial, or use of the no-prior ramucirumab population leads to an under-representation of Asians compared with the English population, their inclusion leads to over-representation, and the generalisability of Japanese patients to the more diverse Asian population in England is unclear. The ERG concludes that analyses of European patients, or where not available, the ROW have highest relevance to the decision problem.

3.2.2.4 The balance of prognostic factors between arms

Clinical advice provided to the ERG indicated that ECOG status, number of metastatic sites, HER2 status and previous chemotherapy regimens (number and type) are prognostic of survival in the third-line setting. The impact of sex and ethnicity was thought to be uncertain due to a lack of data. The CS broadly agrees, stating that "*ECOG PS, age, number of previous chemotherapy regimens (two versus three), number of metastatic sites, and HER2 status were prognostic of improved OS*" (p56 of the CS). Of these, there was some evidence of imbalance in some factors (TFT compared with placebo): sex (75% and 69%); ethnicity (72% and 66% white); ECOG status 0 (36% and 40%); HER2 status (20% and 16% positive); number of metastatic sites (≥3 54% and 58%); number of previous regimens (≥4 23% and 27%).

The clinical advisors to the ERG noted the slight imbalances in potential prognostic factors but were not concerned. However, the ERG asked the company for clarification on how the results might affect efficacy estimates. The company responded (clarification response A16[7]) that their clinical advisors had not been concerned, and consequently no adjustments had been made in analyses. They added that the direction of effect of imbalances in prognostic factors was mixed and likely to counteract each other. In their OS analysis, however, the company has adjusted for [redacted]) and the adjusted result was similar to the ITT population primary analysis results (see Section 3.2.2.7).

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

[redacted]

3.2.2.5 Flow of patients through the trial

A flow diagram of patients through the trial was presented in the CS in the appendix relating to the systematic review (Appendix D) and in Shitara *et al*.[2] A version correcting identified errors was provided as Figure 5 in the clarification response and is reproduced here as Figure 2.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

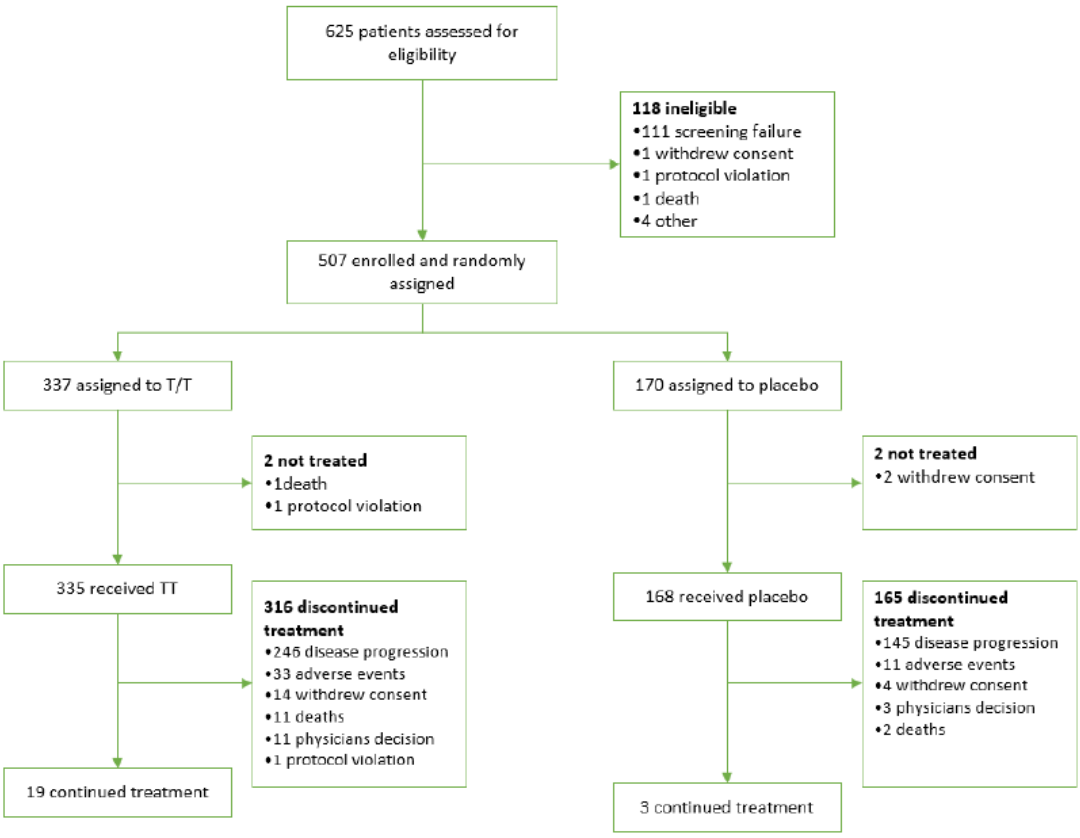


Figure 2 CONSORT flow diagram of patients in the TAGS study. Reproduction of Figure 5 of the clarification response[7]

The ERG calculated treatment discontinuation rates in Table 5. In total, the number of treatment discontinuations not due to disease progression was 21.4% in the TFT arm, and 12.9% in the placebo arm. The proportion of deaths were 3.6% in the TFT arm, but 1.2% in the placebo arm. The difference in deaths appears large, but for 9/12 patients in the TFT arm the cause of death was disease progression, meaning patients died before progression was recorded as an outcome. All other reasons are generally higher in the TFT arm, which may be due to patients being on treatment for longer. The only discontinuation reason that might affect OS is withdrawal of consent as all other patients would be followed up for survival, and this is largely balanced between arms (4.2% in TFT arm and 3.6% in placebo arm). For PFS, it is not clear to the ERG how withdrawal of consent, physician decision, protocol deviation and adverse events (AEs) were handled in analyses.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

Table 5 Reasons for treatment discontinuations in the TAGS study. Number of patients taken from Shitara *et al*.[2]

Reason for treatment discontinuation	TFT		Placebo	
	Number of patients (N=337)	% of patients in arm	Number of patients (N=170)	% of patients in arm
Death	12*	3.6	2	1.2
Protocol deviation	2 †	0.6	0	0
Disease progression	246	73.0	145	85.3
Adverse events	33	9.8	11	6.5
Withdrew consent	14	4.2	6‡	3.6
Physician decision	11	3.3	3	1.8
Still on treatment	19	5.6	3	1.8
Total	337	100	170	100

*One death occurred in the TFT arm before treatment started

†One protocol violation occurred in the TFT arm before treatment started.

‡Two withdrew consent from the placebo arm before treatment started.

3.2.2.6 Dosing delays, protocol deviations and non-study drugs

Dosing delays: 58% of patients receiving TFT and 22% of patients receiving placebo had dosing delays and reductions (Table 14 of CS); 11% and 1% had dose reductions; 13% and 17% had treatment discontinuation.

Protocol deviations: [redacted]

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

Non-study drugs: The ERG asked for clarification around how many patients received non-study drugs as this was not clear from the CS. [redacted] The ERG asked for clarification on how many patients discontinued treatment in order to receive a non-study drug; [redacted]

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

3.2.2.7 Efficacy of TFT

Key efficacy results for the TAGS study are presented in Table 6.

Table 6: The efficacy data from TAGS, with reference to NICE scope. Data taken from the CS[1], the CSR[7] and Shitara *et al*.[2]

	Whole trial population			**Ramucirumab–naïve subgroup (pre-specified analysis)**		
Outcome	**TFT N=337**	**Placebo N=170**	**Between group comparisons**	**TFT N=223**	**Placebo N=115**	**Between group comparisons**
Outcomes listed in NICE scope						
Primary outcome						
OS Median (months):	5.7, 95% CI: 4.8-6.2	3.6, 95% CI: 3.1-4.1	HR: 0.69 (95% CI: 0.56–0.85, one-sided p=0.0003, two-sided p=0.0006) Difference between medians: 2.1 months	6.0 (95% CI: 5.1-6.9)	3.3 (95% CI: 2.8-3.9)	HR: 0.66 (95% CI: 0.51–0.85)
Key secondary outcome						
PFS Median (months):	2.0 (95% CI1.9-2.3)	1.8 (95% CI 1.7-1.9)	HR: 0.57 (95% CI: 0.47–0.70, two-sided p<0.0001)	2.2 (95% CI:1.9-3.5)	1.8 (95% CI:1.8-1.9)	HR 0.5832 (95% CI: 0.4550- 0.7475)
Other secondary outcomes						
ORR* Rate	4.5%	2.1%	NR	NA	NA	NA
Disease control rate (DCR, composite of CR, PR and SD)	44.1%	14.5%	**p<0.0001**	NA	NA	NA
SD	39.7%	12.4%		NA	NA	NA
Duration of response	No summary statistics presented. See "**Response rates and duration of response"** below.			NA	NA	NA
HRQoL EORTC QLQ-c30 and QLQ-STO22	No comparative summary statistics presented in the clinical section. See clarification response A20 for full HRQoL summary statistics.			NA	NA	NA
Outcomes not listed in the NICE scope						
Time to Deterioration of ECOG Performance Status Median (months)	4.3 (95% CI: 3.7–4.7)	2.3 (95% CI: 2.0–2.8)	HR 0.69, 95% CI: 0.56–0.85, two-sided p=0.0005	NA	NA	NA

* Restricted to patients with measurable disease, i.e. 290/337 in the TFT group, 145/170 in the placebo group

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

Overall survival

In the ITT population, the hazard ratio for OS was 0.69; 95% CI: 0.56–0.85, p=0.0006, indicating patients lived statistically significantly longer in the TFT arm than in the placebo arm. Median OS was 5.7 months in the TFT arm and 3.6 months in the placebo arm; the difference in median survival was 2.1 months. At six months, 47% of TFT patients and 33% of placebo patients were alive. At one year, 21% and 13% respectively were alive. The Kaplan-Meier plot is provided as Figure 3a.

[redacted]

OS subgroup and supportive analyses

Prognostic factor subgroups: Subgroup analyses (including by stratification factor) were reported more fully in the CSR and are reproduced here as Figure 4. The study was not powered for these subgroup analyses, and no statistical comparisons were presented. Patients with measurable disease appeared to respond statistically significantly less well than those without measurable disease, on the basis of their confidence intervals not overlapping (0.74 (95% CI 0.59 to 0.93) compared with 0.21 (95% CI 0.09 to 0.52) respectively[2]). Where confidence intervals overlapped, the biggest differences between point estimates were seen for prior treatment with irinotecan (point estimate favours those without prior treatment) and prior treatment with taxane (point estimate favours TFT for those with prior treatment and favours placebo for those without prior treatment). Others with notable differences included age (<65 compared with ≥75), prior ramucirumab, "other" ethnicity, gastrectomy, tumour grade, peritoneal metastases, histology and HER2 status.

No prior ramucirumab: The pre-specified analysis of patients with no prior ramucirumab treatment indicated that the treatment effect was consistent with the main analysis with an HR 0.66 (95% CI: 0.51–0.85), and similar median survival (see Table 6).

Japanese patients compared with rest of the world: Patients in Japan had a median OS in the TFT (n=46) and placebo (n=27) groups of 6.3 months and 5.9 months, respectively, and a HR of 0.77 (95% CI 0.46–1.30). This compared with patients in rest of the world (ROW, comprising EU/US patients) who had a median OS of 5.4 months in the TFT (n=291) and 3.3 months in the placebo group (n=143), and a HR of 0.68 (95% CI 0.54–0.85).

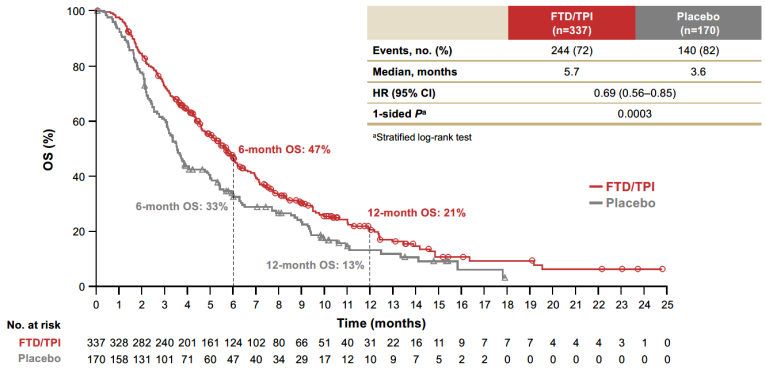
Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

EU patients with no prior ramucirumab: [redacted].

However, the ERG did not agree that subgrouping patients by prior ramucirumab treatment is necessarily appropriate (see Section 3.2.2.2). Subgroup data for patients in Europe regardless of prior ramucirumab treatment were presented in Shitara *et al.*[2] for OS, with an HR of 0.67; 95% CI: 0.53-0.86. Median survival was presented in the CSR (CSR[14], Figure 7) as [redacted]. No Kaplan-Meier plots were available to the ERG.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

a)



b)

Figure 3: **Overall survival Kaplan-Meier curves for a) the whole population and for b) patients in the EU with no prior ramucirumab treatment. Reproduction of Figure 7 of the CS and Figure 1 of the clarification response, respectively**

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

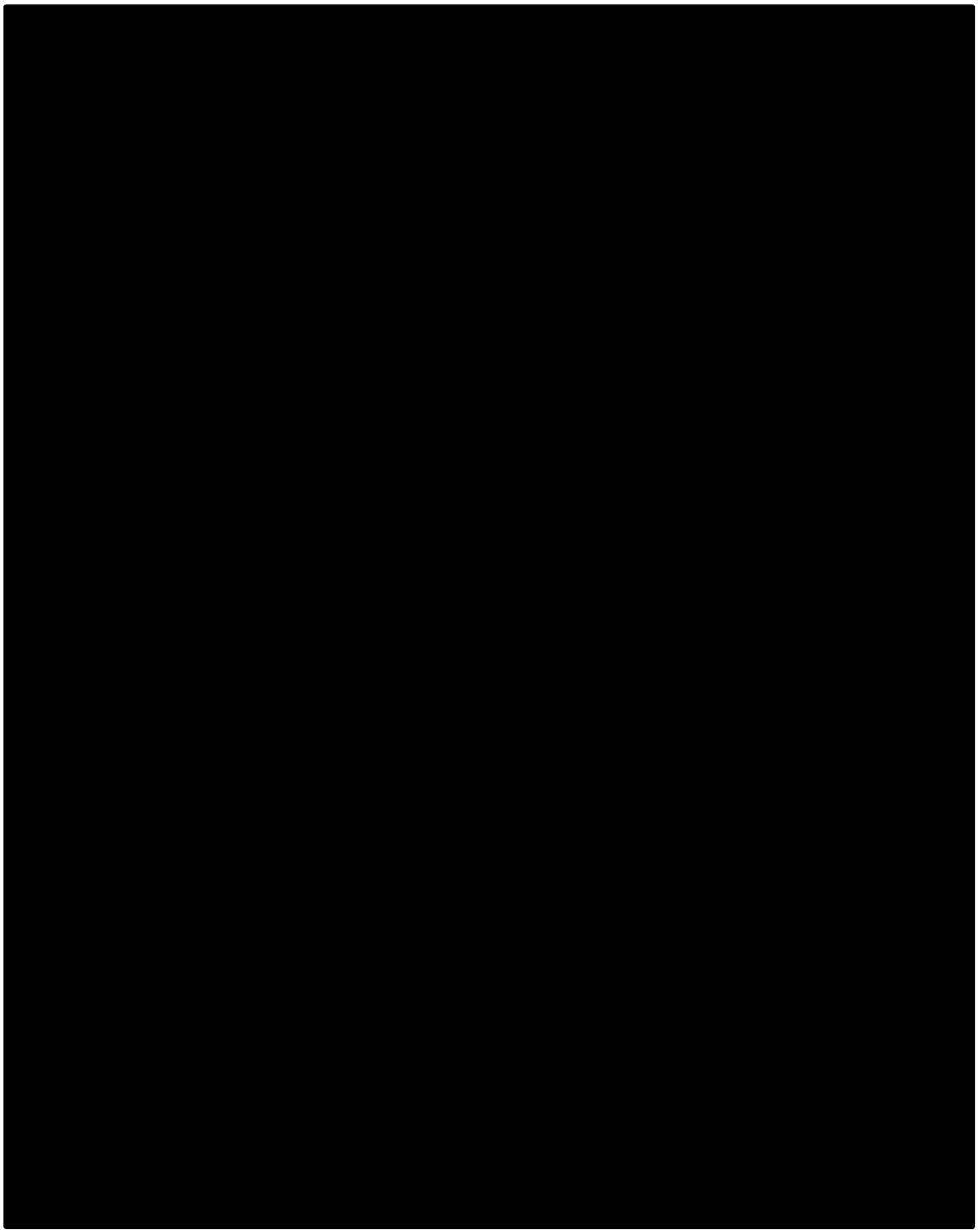


Figure 4: Reproduction of Figure 7 from the CSR, Hazard Ratio for Treatment Effect on Overall Survival by Selected Subgroups (Intent-to-Treat Population)

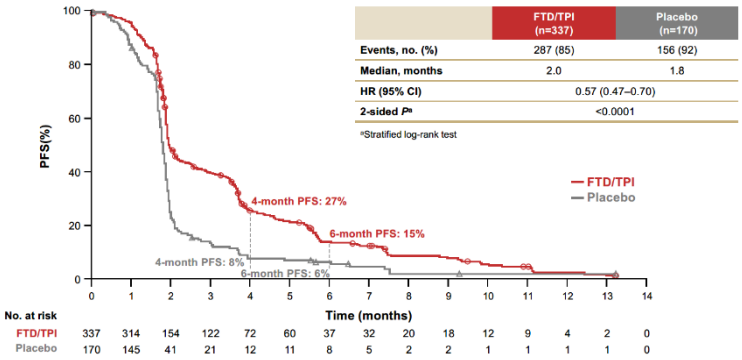
Progression-free survival

In the ITT population, the HR for PFS was 0.57, 95% CI: 0.47–0.70, p<0.0001, indicating patients progressed statistically significantly later in the TFT arm compared with the placebo arm. Median PFS

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

was 2.0 months in the TFT arm and 1.8 months in the placebo arm; the difference in median PFS was 0.2 months. At six months, 27% of TFT patients and 8% of placebo patients were progression free and alive. At one year, 15% and 6% respectively were progression free and alive. The Kaplan-Meier plot for the whole TAGS population is provided in Figure 5a with the value for European patients without prior ramucirumab use in shown in Figure 5b. Both plots have "steps" at two monthly intervals, presumably caused by radiological progression being observed at scheduled study assessment points. Clinical advice to the ERG suggested 4-6 weekly monitoring was usual in clinical practice in England. This may not always be a radiological assessment, but clinicians would make treatment decisions based on clinical or radiological progression, in accordance with the stopping rules described in Section 3.2.1. As such, patients may discontinue treatment earlier, which may affect efficacy, adverse events, and drug costs.

a)



b)

Figure 5: **Progression free survival Kaplan-Meier curves for a) the whole population in the TAGS study and for b) patients in the EU with no prior ramucirumab treatment. Reproduction of Figure 8 in the CS and Figure 2 in the clarification response, respectively**

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

PFS subgroup and supportive analyses

Prognostic factor subgroups: patients subgrouped according to potential prognostic factors were presented in the online supplement of Shitara *et al.*[2] Of most relevance to the appraisal was an analysis of European patients, regardless of ramucirumab treatment. The HR was 0.60; 95%CI: 0.48-0.75. The CS, Shitara *et al.* and the CSR report neither the median survival nor the Kaplan-Meier plots.

EU patients with no prior ramucirumab: [redacted]

[redacted]

Subgroups: Shitara *et al*[2] reports pre-specified subgroup analyses for PFS, but these were not reported in the CS, and are not reported here due to their low relevance to the health economic model (Chapter 4).

Response rates and duration of response

Response rate outcomes only included patients with measurable disease and ≥1 post-baseline assessment (the TR population; 287/337 (85%) patients receiving TFT, and 156/170 (92%) patients receiving placebo). Objective response rate (ORR, a composite of CR and PR) was 4.5% and 2.1% (*p*-value not reported) and were low in both arms, as would be expected in patients at third and later lines of chemotherapy. DCR (a composite of CR, PR and SD) was 44.1% and 14.5% ($p<0.0001$) in the TFT

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

and placebo arms respectively. DCR rates were largely due to SD (39.7% versus 12.4%, respectively). For the European, no prior ramucirumab population the DCR was [redacted] in the TFT group versus [redacted] in the placebo group.

Duration of response was only presented as a figure in the clarification response[7] and is presented here as Figure 6 a and b. [redacted]

No other subgroup or supporting analyses were presented in the CS. It is unclear what the results would be in patients without measurable disease.

[redacted]

Figure 6: Duration of response in a) the whole population in the TAGS study and for [redacted] Reproduction of Figure 3 & 4 in the clarification response

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

Time to ECOG Performance Status ≥ 2

The HR for Time to ECOG performance status ≥ 2 was 0.69 (95% CI: 0.56–0.85, p=0.0005) with a median of 4.3 months (95% CI: 3.7–4.7) and 2.3 months (95% CI: 2.0–2.8) in the TFT and placebo groups, respectively. This was not an outcome listed in the NICE scope. However, it indicates that patients maintained ECOG performance status for longer in the TFT arm compared with the placebo group.

Health-related quality of life

HRQoL was measured with the EORTC QLQ-C30 and the EORTC QLQ-STO22. The company did not report HRQoL using comparative summary statistics with *p*-values. Instead, a simple table of mean change from baseline was presented (with neither confidence intervals nor *p*-values reported, p56 of the CS). Additionally, a large table of summary values with standard deviations for each subscale was provided in the clarification response (question A20).[7] Using the pre-specified cut-off of a mean change from baseline of ≥10 points, the company stated that there were no clinically relevant differences (between groups) in the mean change from baseline. For some items there was a difference of ≥10 points in the mean change from baseline (within group), and these included less hair loss, role functioning, fatigue, pain and appetite loss. The company also highlighted a between-group clinically relevant difference in pain: 11.3 (Cycle 2) in favour of TFT; and role functioning: 10.0 (Cycle 3) in favour of placebo. The ERG comments that the cut point of 10 points was derived for the EORTC QLQ-C30, and of unknown significance to the EORTC QLQ-STO22[15] but concludes that the results suggest that HRQoL is not affected in a clinically relevant way, either positively or negatively, by treatment with TFT.

3.2.2.8 Safety of TFT: Adverse events

Only one study (TAGS[2]) reporting AEs was included in the CS review. These AEs are summarised in Table 12 (page 59) of the CS.[1] The safety analysis included all patients who took any dose of study treatment. All analyses using this population were based on the treatment received. Safety was assessed in the TAGS trial by investigators throughout the study and AEs were graded according to the US National Cancer Institutes' common Terminology Criteria for Adverse Events (version 4.03[19]) and recorded from the first dose of study drug until 30 days after the last dose of study drug. The overall incidence of AE events was 97.3% for the TFT group and 93.5% for the placebo treatment group. However, grade III or worse AEs occurred in 267 (80%) patients in the TFT group, but only in 97 (58%) patients in the placebo group.

The most common AE reported included: nausea; anaemia; decreased appetite; vomiting; diarrhoea; fatigue; neutropenia; asthenia thrombocytopenia. Anaemia and neutropenia were two outcomes where

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

the incidence appeared to be markedly greater in the TFT group compared with the placebo group (anaemia 45% vs 19%; neutropenia 53% vs 4%).

AEs resulting in death occurred in 13.4% (n=45 patients) in the TFT group and in 11.3% (n=19 patients) in the placebo group. The most frequently reported AE resulting in death in both treatment groups was general physical health deterioration.

In the economic model, treatment-emergent grade 3 or 4 AEs were included provided they occurred in at least 5% of patients in either treatment arm. In addition, febrile neutropenia (occurring in n=6 of TFT patients in the TAGs trial) and nausea (n=14 TFT, n=5 control) were included within the cost-effectiveness model owing to their high impact on patient HRQL and the cost of its treatment. The ERG notes potential discrepancies between the data used in the model, the data reported in the clinical section of the CS, and data reported in the CSR. The biggest discrepancy is for the category anaemia, which is reported as [redacted]) in the TFT and placebo arms respectively in the CSR Table 35, as 64 (ERG-calculated 19.1%) and 13 (ERG-calculated 7.7%) patients respectively in Shitara *et al.*[2] and the CS Table 12, but as 37 (11%) and 5 (3%) patients respectively in the modelling section (Table 23 of the CS). The ERG were not able to determine why there was an apparent discrepancy, but speculate this may be due to composite outcomes ("anaemia" in the modelling compared with "anaemia or decreased haemoglobin concentration" in Shitara *et al.*), and/or different categories entering the analysis ("treatment emergent" versus "any cause", or events in 2% versus 5% of patients).

A phase II trial (Bando *et al.* 2016[13]) in Japanese patients was excluded from the review. The ERG has however, included the AEs observed in this study for comparison (Table 7). The most commonly occurring more serious AEs (grade III/IV) in patients receiving the 35 mg/m^2 were neutropenia (69.0%), leukopenia (41.4% and anaemia (20.7%). One case of febrile neutropenia occurred, although no treatment related deaths were reported. For those patients in the 40 mg/m^2 group, neutropenia (83.3%) and leukopenia (66.7%) were slightly more frequent. The findings of this study are tabulated below alongside the incidence of grade III-V AE in the TAGS study.[2] Neutropenia, anaemia and leukopenia were the most common serious AEs in both studies. The incidence of leukopenia was greater in the study by Bando *et al.* (41.4% vs 9.3%).[13] More AE-related deaths occurred in the TAGS study when compared with the Bando *et al.*[13] study (13.4% vs 0%); however, AE-related deaths in the intervention and control groups were similar in the TAGS study.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

Table 7 Serious adverse events grade III or higher (Bando *et al.*[13], Shitara *et al.*[2] and CSR)

	EPOC1201		TAGS					
	Bando *et al.*[13]		Shitara *et al.*[2]		[redacted]		Model, CS Table 23	
	35 mg/m^2 (n=29) (%)	40 mg/m^2 (n=29) (%)	TFT (n=335) (%)	Placebo (n=168) (%)	[redacted] (%)	[redacted] (%)	TFT (n=335) (%)	BSC (n=168) (%)
Haematological								
Neutropenia	20 (69)	5 (83.3)	114 (34.0)*	0 (0)			77 (23)	√
Febrile neutropenia	1 (3.4)	0 (0)	4 (1.2)	0 (0)			6 (1.8)	√
Leukopenia	12 (41.4)	4 (66.7)	31 (9.3)	0 (0)			23 (6.9)	√
Anaemia	6 (20.7)	1 (16.7)	64 (19.1)	13 (7.7)			37 (11)	5 (3)
Asthenia			16 (4.8)	11 (6.5)			√	√
Thrombocytopenia			11 (3.3)	0 (0)			NR	NR
Hyponatraemia			4 (1.2)	7 (4.2)			NR	NR
Neutrophil count decreased			NR	NR			37 (11.0)	0 (0)
Non-haematological								
Anorexia/decreased appetite	3 (10.3)	1 (16.7)	29 (6.7)	11 (6.5)			√	√
Nausea	1 (3.4)	1 (16.7)	10 (3.0)	5 (3.0)			√	√
Vomiting	1(3.4)	1 (16.7)	12 (3.6)	3 (1.8)			NR	NR
Diarrhoea	0 (0)	0 (0)	9 (2.7)	3 (1.8)			NR	NR
Abdominal pain	0 (0)	0 (0)	14 (4.2)	15 (8.9)			√	√
Constipation	0 (0)	0 (0)	4 (1.2)	4 (2.4)			NR	NR
Deaths	0 (0)	0 (0)	45 (13.4)	19 (11.3)			NR	NR
At least one serious AE	NR	NR	143 (42.7)	70 (41.7)			NR	NR
Dysphagia	NR	NR	7 (2.1)	4 (2.4)			NR	NR
Gastrointestinal haemorrhage	NR	NR	4 (1.2)	1 (0.6)			NR	NR
Fatigue	NR	NR	23 (6.9)	10 (6.0)			NR	NR
Back pain	NR	NR	2 (0.6)	4 (2.4)			NR	NR
Blood (alkaline phosphatase	NR	NR	9 (2.7)	5 (3.0)				
Dyspnoea	NR	NR	6 (1.8)	6 (3.6)			NR	NR
Ascites	NR	NR	12 (3.6)	11 (6.5)			NR	NR
General health deterioration	NR	NR	22 (6.6)	15 (8.9)			NR	NR
y-[glutamylatransferase]	NR	NR	3 (0.9)	4 (3.0)			NR	NR

*This result includes both neutropenia and those with neutrophil count decreased. In the CSR these are reported separately.
√ represents where data is the same in all TAGs sources
AE, adverse event; NR, not reported; BSC, best supportive care.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

3.3 Critique of trials identified and included in the indirect comparison and/or multiple treatment comparison

No meta-analysis or indirect comparison was reported. The company states this is due to a paucity of relevant evidence in the third-line setting. The ERG agrees that an indirect comparison would not have been useful due to the quality and quantity of data available, and the infrequent use of chemotherapy in the third-line setting.

3.4 Critique of the indirect comparison and/or multiple treatment comparison

Not applicable.

3.5 Additional work on clinical effectiveness undertaken by the ERG

None, other than reported above.

3.6 Conclusions of the clinical effectiveness section

The ERG agrees that the CS has included all relevant trials, and that a network meta-analysis of the evidence relating to other third-line treatments was not feasible. The key trial (TAGS)[2] was a phase III double-blind, placebo-controlled, multicentre randomised controlled trial. It included all key outcomes identified in the NICE scope.

An OS advantage was observed (HR 0.69 (95% CI: 0.56–0.85, p <0.001)), with median survival of 5.7 months (95% CI: 4.8-6.2) in the TFT arm and 3.6 (95% CI: 3.1-4.1) in the placebo arm - a difference of 2.1 months. In subgroup analyses, for OS, patients with prior ramucirumab treatment had a HR of 0.76 (95% CI 0.53-1.09) and those without an HR 0.66 (95% CI 0.51-0.85). Patients from Japan had a HR of 0.77 (95% CI 0.46-1.30), and those from ROW had a HR of 0.68; (95% CI: 0.54-0.85). Patients from Europe had a HR of 0.67 (95% CI 0.53-0.86). The ERG requested an analysis of patients from Europe without prior ramucirumab treatment, for which the HR was not reported, [redacted]

A PFS advantage was observed (HR 0.57 (95% CI: 0.47–0.70, p<0.0001)), although the absolute benefit (difference in median PFS between arms) was only 0.2 months (2.0 months (95% 1.9-2.3) and 1.8 months (95% CI 1.7-1.9) in the TFT and placebo arms, respectively). The analysis was not adjusted for all prognostic factors.

Objective response rates were low (4.5% and 2.1% respectively) as would be expected in patients at third-line of treatment. Disease control rate was higher in the TFT arm than in the placebo arm (44.1%

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

and 14.5% respectively, p<0.0001), and was mostly driven by SD rather than CR or PR. HRQoL appeared largely maintained with TFT treatment.

Key adverse events were nausea, anaemia, decreased appetite, vomiting, diarrhoea, fatigue, neutropenia and asthenia thrombocytopenia. Anaemia and neutropenia were two outcomes where the incidence appeared to be markedly greater in the TFT group compared with the placebo group (anaemia 45% vs 19%, neutropenia 53% vs 4%).

The population recruited to TAGS was thought to be largely generalisable to the population in England, with some exceptions.

The study stratified patients according to prior exposure to ramucirumab and the company used the subgroup of patients without prior treatment with ramucirumab in the base case analyses within its model. The ERG has noted uncertainty in the clinical views about whether prior ramucirumab use affects the natural history of the disease, whilst two clinicians agreed that it is unlikely to affect the efficacy of TFT. Without a strong indication that prior ramucirumab treatment alters prognosis, the ERG prefers to assume that there is no impact but notes differences in the prior ramucirumab group and the no ramucirumab group in terms of prior lines of treatment and disease duration. Therefore, although this assumption is uncertain the ERG prefers an estimate of a HR or AF from the entire population rather than the non-the ramucirumab patients only

The inclusion of a larger proportion of Japanese patients in TAGS than are in the English population was potentially problematic as Japanese patients have a different natural history and treatment pathway than European patients. Whilst exclusion of the Japanese patients leads to an under-representation of Asians compared with the English population, their inclusion leads to over-representation, and the generalisability of Japanese patients to the more diverse Asian population in England is unclear. The ERG concludes that analyses of European patients, or where not available, the ROW have highest relevance to the decision problem.

There were also some minor imbalances in prognostic factors between arms at baseline. In the primary analysis of OS, adjustment for these factors did not affect the HR. [redacted]

There were more gastric patients (rather than gastroesophageal patients) than clinical advice to the ERG indicated would be expected, but clinical advice suggested that this is unlikely to affect estimates of efficacy.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

In terms of the intervention, it was not entirely clear whether the discontinuation rules were mandatory or optional, and their application in a clinical setting may differ from that in a trial setting.

Within the trial, [redacted]

Patients may be assessed for treatment continuation more frequently (4- to 6-weekly rather than every two months) in England than in the TAGS study, and this could lead to earlier discontinuations, with an unknown impact on efficacy, adverse events and costs.

For objective response rate and disease control rate, no adjustments were made and it was not clear how missing data were handled. Because there was a statistically significant difference in efficacy for patients with measurable disease compared with those without measurable disease, and this analysis only included those with measurable disease, the data may not be generalisable to the whole population.

[redacted]

In conclusion, TFT appears to confer an overall survival advantage with a HR of 0.69 in the whole TAGS study population. The midpoint for the HR is lower in subgroup analyses that may be more applicable to England: 0.68 for the ROW group; 0.67 for the European group and 0.66 for those who had not received prior ramucirumab treatment. Other outcomes indicate that the key benefit of TFT is the gain in OS, as there is only a 0.2 month absolute difference in median PFS, although there was a clear improvement after 2 months int eh whole TAGS study population, and no improvement in HRQoL compared with baseline values.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

4 COST EFFECTIVENESS

This section provides a structured critique of the economic evidence submitted by the company to support the cost effectiveness of TFT for mGC patients who have received two or more lines of treatment.

The two key components of the economic evidence presented in the CS are: (i) a systematic review of the relevant literature and (ii) a report of the company's *de novo* economic evaluation. The company also provided an electronic version of their economic model developed in Microsoft Excel. For brevity, in most cases treatment with TFT + BSC will be referred to as TFT, and placebo + BSC will be referred to as BSC.

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

4.1.1 Objective of cost effectiveness review

The company undertook an SLR to identify published evidence to support the company's cost effectiveness model. Details of the search strategies employed by the company are provided in Appendix G of the CS.

Searches were conducted in June 2018 and covered an appropriate range of databases (MEDLINE & Medline-in-process; Embase; Econlit and NHS EED and HTAD); relevant conference series; and international HTA websites (see Table 1 below for more detail). The search strategies are generally well-designed, although as with the clinical SLR, the ERG notes the use of a multi-file search to interrogate Medline and Embase simultaneously, with some associated loss of functionality. However, it is unlikely that the SLR has missed any relevant studies.

No citation is provided for the search filters used to identify economic evaluations, although in their response to the ERG's clarification letter (question A9) the company explained that the terms used were based on filters developed by the Scottish Intercollegiate Guidelines Network (SIGN). While SIGN filters are not formally validated, the ERG recognises that they are expert-designed and likely to retrieve most of the studies eligible for inclusion.

A cut-off date of 10 years was applied; this was justified by the company on the grounds that considerable change had been observed in this period in terms of technology evolution and quality of care, and that 2008 was also the date of the earliest study identified in the clinical effectiveness review (Clarification response, question A8).

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

Additionally, the company undertook searches as needed to populate its economic model (including utility studies and cost/resource use studies (reported in Appendices H and I, respectively)). The same sources have been used as for the economic SLR, and again the ERG is broadly satisfied with the searches as reported in the relevant sections.

Table 8: Data sources for the economic systematic review

Search strategy component	Sources	Date limits
Electronic database searches Key biomedical electronic literature databases recommended by HTA agencies	-MEDLINE® -MEDLINE® In-process -Embase® -The Cochrane Library including National Health Service Economic Evaluation Database (NHS EED) -EconLit® -Health Technology Assessment Database (HTAD)	2008-2018
Conference proceedings	-International Society for Pharmacoeconomics and Outcomes Research (ISPOR) -American Society of Clinical Oncology (ASCO) -The American Society of Clinical Oncology Gastrointestinal Cancers Symposium (ASCO GI).	2016-2018
Key international HTA websites	-National Institute for Health and Care Excellence (NICE) -Scottish Medicines Consortium (SMC) -All Wales Medicines Strategy Group (AWMSG) -Haute Autorité de Santé (HAS) -Statens legemiddelverk (SLV)	Not specified

4.1.2 The inclusion and exclusion criteria used in the study selection

The inclusion/exclusion criteria used by the company to facilitate study selection are presented in Table 9.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

Table 9: Inclusion/exclusion criteria for the economic review

Category	Inclusion criteria	Exclusion criteria
Population (P)	-Age: adults aged ≥18 years -Gender: any -Race: any -Disease: patients with either unresectable advanced/metastatic GC who are at stage IIIb and IV according to the American Joint Committee on Cancer guidelines, or GEJ cancer	• Paediatric patients • Patients with early stage GC • Oesophageal cancer • Localised GC
Intervention (I)	All pharmacological interventions	Non-pharmacological interventions
Comparator (C)	• Any pharmacological intervention • Placebo • Best supportive care	None
Outcome (O)	Studies were not excluded based on the reported outcome	None
Study design	-All economic evaluation studies based on models -Cost-effectiveness analysis -Cost-utility analysis -Cost-minimisation analysis -Cost-benefit analysis -Budget impact models -Cost-consequence analysis	-Letters, comments and editorials -Studies reporting clinical data only -Simple costing analysis studies
Line of therapy	Third- or further-line of therapy	First- or second-line of therapy
Search timeframe	2008 to 2018	Studies published prior to 2008
Language	No restrictions	None

GC, gastric cancer; GEJ, gastroesophageal junction

4.1.3 Findings of the cost effectiveness review

Four studies were identified that were of relevance to the decision problem; however, none of these included TFT as an option. All four models used a three-state partitioned survival model within their analyses.

4.1.4 Conclusions of the cost effectiveness review

As the company's searches did not identify any relevant studies of TFT, they developed a *de novo* health economic model.

4.2 Summary of the company's submitted economic evaluation

4.2.1 Population

The population included in the company's health economic analysis reflects adult patients with mGC including adenocarcinoma of the GEJ who have received at least two prior lines of treatment. The modelled patient characteristics reflect those of the full patient population within the TAGS trial[2] with

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

an average age of 62.5 years, and 27% of the population are assumed to be female. However, the selected BSA distribution for the base case used the European patient cohort of the TAGS trial (with an average of 1.77 m^2) as this was deemed more clinically appropriate for patients in England.

4.2.2 Interventions and comparators

In the TAGS trial, TFT was administered in combination with BSC, where BSC is provided to alleviate cancer-related symptoms and maximise the patient's health-related quality of life. TFT was taken orally at a dose of 35 mg/m^2 twice daily for 10 days (1-5 and 8-12) per 28-day treatment cycle. The comparator was placebo in combination with BSC.

The company stated that no treatments have been recommended by NICE for mGC patients who had received two or more prior lines of treatment, and that chemotherapy (as included in the final NICE scope[5]) is rarely used for such population. This was confirmed by a clinician advisory board held by the company, and therefore TFT + BSC was only compared with BSC within the company's economic analyses.

4.2.3 Perspective, time horizon and discounting

The base case model adopts an NHS and Personal Social Services perspective. The base case model uses a 10-year time horizon; shorter values were included in the company's scenario analyses. Both costs and QALYs were discounted at 3.5% per annum as recommended by NICE.[20]

4.2.4 Model structure

As part of its submission to NICE, the company developed a fully executable partitioned survival model (PSM) that comprised three mutually exclusive and exhaustive health states: (i) progression-free (PF); (ii) progressed disease (PD); and (iii) death. The model is similar to that of other treatments for advanced/metastatic cancer previously submitted to NICE as part of the STA process. The health states and possible transitions between these are shown in Figure 7, with the arrows for remaining in the same state added by the ERG. A weekly cycle length was used; according to the CS, this obviated the need for half-cycle correction.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

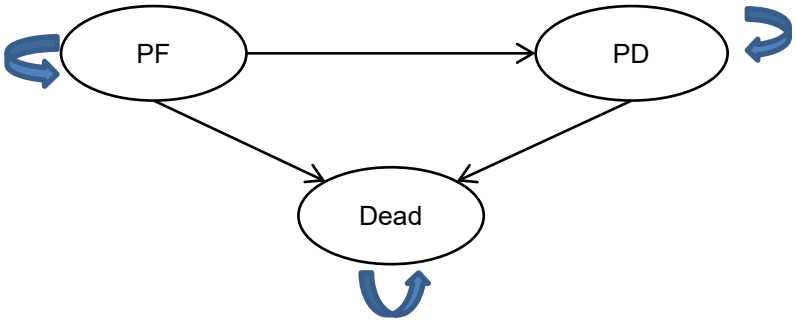


Figure 7: The company's model structure

All patients are assumed to enter the model in the PF health state and remain there until progression or death. As with a standard PSM, the transition probabilities between the health states are inferred via extrapolated PFS and OS curves fitted to the clinical trial data. For patients on TFT, parametric curves were fitted to time to treatment discontinuation data from the TAGS RCT in order to estimate the cost of TFT treatment. The company assumed that should the treatment discontinuation curve be higher than the progression-free survival (PFS) curve in the extrapolated period, then the discontinuation curve would be set equal to the PFS curve.

4.2.5 Evidence used to inform the company's model parameters

4.2.5.1 Treatment effectiveness and extrapolation in the base case

Data from the TAGS trial for patients with no prior ramucirumab treatment (n=222) were used for the extrapolation of PFS and OS in the TFT and BSC arms in the company's base case. At the time of data cut-off (30th April, 2018 for OS and 31st March, 2018 for all other clinical data), more than 90% patients in both arms had experienced the event of interest for both PFS and OS endpoints.

The company followed standard guidance for fitting and selecting survival models based on NICE Decision Support Unit (DSU) Technical Support Document (TSD) 14.[21] A full description of the survival extrapolation analyses undertaken by the company is presented in Section B.3.3.2 of the CS.[1]

The company investigated the use of a range of parametric survival models: exponential, Weibull, Gompertz, lognormal, log-logistic and generalised gamma. The company also explored two approaches to model the treatment effect: combined models (with a covariate for treatment assignment) and independent models (models were fitted independently to data for each treatment arm). Hence, 12 distinct extrapolations were available for use in each treatment arm for each of the PFS and OS dataset (6 from the combined modelling approach and 6 from the independent modelling approach). All survival analysis was performed using the "flexsurv" package in R.[22, 23]

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

4.2.5.1.1 Extrapolating OS

The company firstly assessed the appropriateness of using either a proportional hazards (PH) model or an accelerated failure time (AFT) model in the combined modelling approach for OS. The company concluded that exponential and Weibull PH models with a covariate for treatment assignment may be inappropriate and AFT models with a covariate for treatment assignment could be considered as appropriate. From assessment of the hazard plots, the company concluded that no specific models were ruled out, but the lognormal and log-logistic models may provide a better fit to the data than the exponential, Weibull and Gompertz models.

Based on the assessment of the visual fit (Figures 21 and 22 of the CS[1]), the statistical goodness-of-fit using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) scores (Table 21 of the CS[1]), and long-term plausibility, the company selected the combined lognormal model as the base case for OS. Figure 8 presents the fitted lognormal model for both treatment arms. The company commented that the long-term extrapolation of OS using the combined lognormal was aligned with clinical expectation, with 5-year OS being 0.71% and 0.23% in the TFT and BSC arms, respectively, and 10-year OS being 0.08% and 0.02% in the TFT and BSC arms, respectively.

The company explored alternative survival models within its scenario analysis. The model assumes that the probabilities of death are always higher, or equal, to those in the general population at the corresponding age.[24]

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

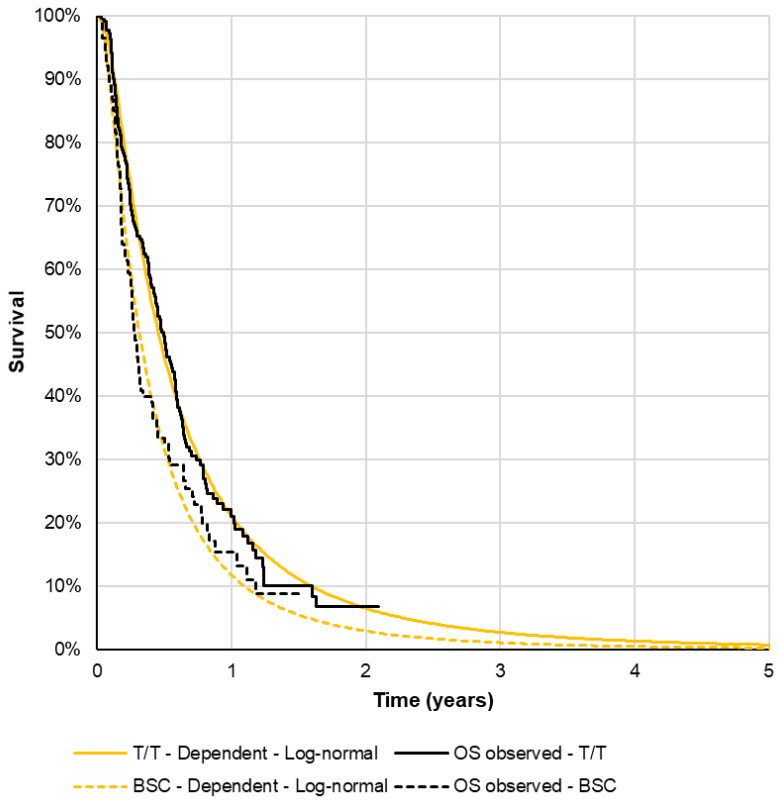


Figure 8: The company's base case OS extrapolation

4.2.5.1.2 Extrapolating PFS

From the assessment of the appropriateness of using either a PH model or AFT model in the combined modelling approach, it was concluded by the company that both PH models and AFT models with a covariate for treatment assignment may be inappropriate for PFS. From assessment of the hazard plots, the company concluded that the exponential, Weibull, Gompertz, lognormal and log-logistic models may be inappropriate as the hazard of a PFS event did not follow the trend associated with these parametric models.

The company selected the generalised gamma model fitted independently to both arms as the base case for PFS due to a good visual fit (Figures 31-32 of the CS[1]), and also that the long-term PFS extrapolation was aligned with clinical expectation (0.00% for both arms at 5-years). The company argued that although the generalised gamma model was not associated with the lowest AIC or BIC, and was considerably higher than some alternative models, it was the only model that allows for a more flexible hazard shape. Figure 9 presents the fitted generalised gamma curves for both treatment arms. The company explored alternative parametric survival models in scenario analyses.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

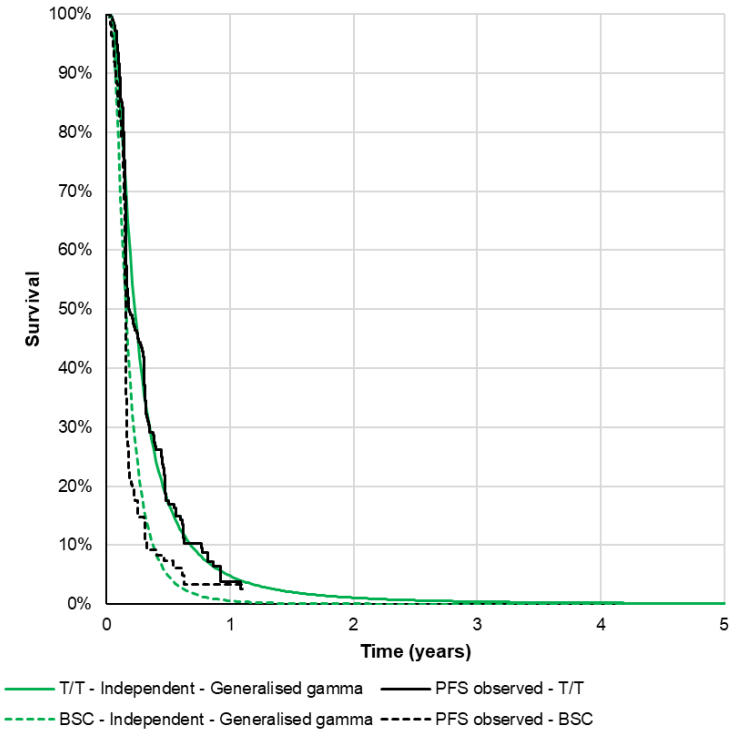


Figure 9: The company's base case PFS extrapolation

The company also explored the use of a hybrid Kaplan-Meier (KM) approach in scenario analyses because of the presence of "*kinks*" in the PFS curves due to the 8-weekly progression assessment visits. In these scenario analyses, PFS was modelled directly from the KM data up to a specified cut-point and following this time point, the failure rates associated with the parametric model within the base case was used. The company's scenario analyses tested different cut-points ranging from 12 weeks (selected as the minimum plausible cut-point) to 33 weeks (selected as the maximum plausible cut-point that represented nearly every observed PFS event).

During the clarification process, the ERG asked the company to provide the extrapolation for PFS using more flexible models such as Royston and Parmar[25] natural spline models because of the complex shape of the observed hazard (clarification response A28).[7] The company provided the extrapolation results using spline models with treatment as a covariate using 1, 3, 5 and 10 internal knots, and concluded that the hazard- and odds-based splines provided better fit than the normal-based splines and the fit improves with the increase in the number of knots (clarification response A28). The company argued that the spline models did not provide a substantial improved fit when compared with the generalised gamma model (the model chosen in the base case), and are expected to "over-fit" the data with 10 knots spline models.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

The ERG also asked the company to estimate the ICER using the European population who have not had ramucirumab treatment (clarification question B4). The company did not describe the extrapolation analyses performed for this subpopulation group in the clarification response, but the extrapolation analyses results were provided within the submitted economic model. It was assumed that the types of statistical distribution chosen in the base case (i.e. dependent lognormal for OS and independent generalised Gamma for PFS) were appropriate in these analyses.

4.2.5.2 Treatment safety

AEs were included in the model to account for the potential cost and HRQoL burden of experiencing events whilst on treatment. Treatment-emergent grade III or IV AEs were included in the model where at least 5% of patients experienced them in one or more treatment arm within the TAGS trial. The only exceptions were the inclusion of febrile neutropenia and nausea. Febrile neutropenia was included due to its significant impact on HRQoL and costs; nausea was included based on expert opinion sought by the company. The incidence rates used to inform the economic model are presented in Table 23 of the CS. The company applied the impact of adverse events on costs and quality of life as one-off events for one cycle at the start of the model. The values are discussed in Section 4.2.5.4.

4.2.5.3 Duration of treatment

In the TAGS trial, the company reported that patients discontinued treatment on TFT mainly due to disease progression (73%) or suffering adverse events (10%). Treatment duration data were collected from individual patients and time to treatment discontinuation (TTD) KM curves were constructed for patients with no prior ramucirumab.

The six independent parametric survival models were fitted to the TTD data, and the generalised gamma model was selected for inclusion in the base case due to its good visual fit (Figure 36 of the CS[1]) and because it had lower AIC and BIC values than the other models (Table 36 of the CS[1]).

Figure 10 presents the TTD KM curve and the fitted generalised gamma model for patients on TFT with no prior ramucirumab experience used in the company's model base case. In order to preserve the structural correlation between progression and treatment discontinuation, the company capped the TTD curves according to the selected PFS curve.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

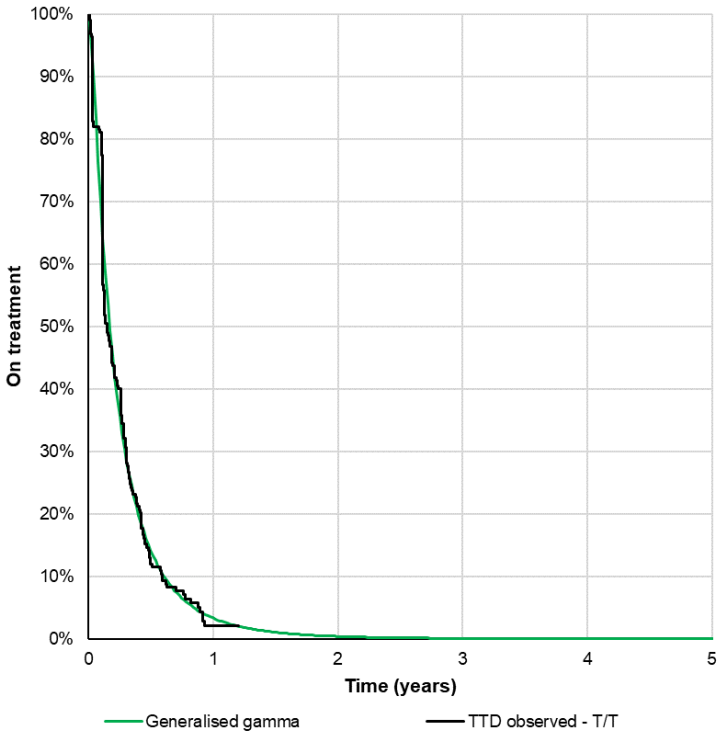


Figure 10: The company's base case TTD extrapolation

4.2.5.4 Health related quality of life

The SLR carried out by the company identified five unique HRQoL studies relevant to the technology appraisal; however, none of these considered relevant to the UK population as all were either Japanese or Chinese studies. In addition, four of the studies sourced the utility values from other studies or trials whose patient populations were considered irrelevant by the company to the decision problem.

HRQoL data were collected using the EORTC QLQ-C30 within the TAGS trial at three different time points, with compliance rates varying between 73% and 100%. A mapping algorithm was applied to these data to estimate the corresponding EQ-5D-3L values. The company used a published mapping algorithm (Kontodimopoulos *et al.*[26]) in its base case as it was the only published algorithm developed in a gastric cancer population according to the latest version of the University Of Oxford Health Economics Research (HERC) mapping database.[27] In an additional clarification question, the ERG questioned why this relatively small study (n=48) was used in preference to other mappings of EORTC-QLQ30 to EQ-5D-3L. This is discussed in detail within Section 5.3.4.2.

The company fitted generalised estimating equation (GEE) regressions using the "geepack" package[28] in R to the mapped utility values to account for the repeated observations for individual patients with four models considered: Model 1 with progression as a covariate; Model 2 with progression and treatment as covariates; Model 3 with progression and no prior ramucirumab as covariates; and Model

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

4 with progression, treatment and no prior ramucirumab as covariates. The goodness-of-fit statistic, Quasi-likelihood under Independence Model Criterion (QIC), was used to compare between the four models. The company stated that the inclusion of covariates for treatment and prior ramucirumab experience did not improve model fit and these were found to be statistically insignificant predictors of utility. Therefore, the company selected the simplest model (Model 1) whereby utility varies with progression status, and tested the other models in its scenario analyses.

During the clarification process, the ERG asked the company to clarify whether including known confounders, such as age and gender, in the GEE regression analysis would have an impact on the results (clarification question B6). In response, the company performed a regression analysis including age, gender and progression status as covariates.

[redacted]

In addition, utility values from previous relevant technology appraisals (TA208 and TA378) were considered for the scenario analysis. These were derived from EQ-5D data collected directly from the appraisals' main clinical trials. Table 10 summarises the six sets of utility values used in the company's economic model.

Table 10: The different sets of utility values included in the company's economic model

	Model 1 (*base case*)	Model 2	Model 3	Model 4	NICE TA378	NICE TA208
Line of treatment	3L+	3L+	3L+	3L+	2L+	1L
Covariates included	Progression	Progression and treatment	Progression and ramucirumab experience	Progression, treatment and ramucirumab experience		
PF utility	0.764	0.786	0.760	0.782	0.737	0.729
PP utility	0.652	0.672	0.647	0.667	0.587	0.577
TFT associated utility		-0.029		-0.029		

PF, progression-free; PP, post-progression

In addition to utility values associated with the two health states, the company's base case analysis applied utility decrements due to AEs. The proportion of patients experiencing a given AE was taken from the safety data of the TAGS trial, with the associated utility decrements being sourced from published literature. For any given AE, the associated QALY loss was calculated by multiplying the

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

utility decrement by the duration over which the AE impact was expected to last. Table 29 of the CS presents the disutility values and duration of AEs included in the company's base case analysis. The frequency of each event is provided in Section 5.2.5.5.5 in this report. AEs attributed to a QALY loss of 0.005 and 0.002 associated with TFT and BSC respectively in the company's base case, which were deducted in the first cycle of the model.

4.2.5.5 Resources and costs

The costs and resource use included in the base case model comprised: drug acquisition costs; drug administration costs; medical resource use (MRU) associated with TFT or BSC; off-treatment and post-progression related costs; AE costs; and end of life care costs. These are discussed in the following sections.

4.2.5.5.1 Drug acquisition costs

TFT tablets are available in two concentrations; 15mg trifluridine/6.14mg tipiracil and 20mg trifluridine/8.19mg tipiracil (referred to as "15mg" and "20mg" respectively) and in two package sizes of 20 and 60. The four formulations had the same cost of £33.33 ([redacted] including the Patient Access Scheme (PAS) discount) per 1mg trifluridine/0.41mg tipiracil.

TFT is administered at a dose of 35mg/m^2 of BSA twice daily on 10 days each 28-day treatment cycle. This represents the licensed dose as well as the dosing followed within the TAGS trial. Table 32 of the CS presents the BSA bands with the required tablets per dose. The company used the BSA distribution of the European population of the TAGS trial within its base case analysis with an average BSA of 1.77m^2. A lognormal distribution fitted to the BSA distribution was combined with a "method of moments" approach to give a weighted average cost of £2,184.01 for TFT per 28-day treatment cycle ([redacted] with the PAS discount). In answering clarification question B9, the company amended a calculation error resulting in an average cost of £2,017.47 ([redacted] with PAS) per 28-day treatment cycle,

Within the TAGS study, three levels of dose reduction were reported (from 35mg/m^2 to 30mg/m^2, 30mg/m^2 to 25mg/m^2, and 25mg/m^2 to 20mg/m^2). The dosing associated with each BSA band is detailed in Table 33 of the CS. Table 34 of the CS presents the number of patients by dosing level for 14 cycles of TFT treatment. Data from the TAGS trial regarding dose delays were also applied in the company's base case model where [redacted] of patients were assumed to start treatment in the second model cycle.

BSC was assumed to have no associated costs within the company's base case analysis; however, post-progression drug costs were considered in the model as detailed in Section 4.2.5.5.4.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

4.2.5.5.2 Drug administration costs

Owing to its oral administration route, the company assumed that no medical resources are needed for TFT administration. Clinical advice provided to the company stated that some clinicians might send patients to a chemotherapy nurse before taking their first treatment cycle. Therefore, the company applied an administration cost of £22.5 (equivalent to 30 minutes of Band 6 nurse time) for the first treatment cycle within its model base case.

4.2.5.5.3 Medical resource use associated with treatment assignment

MRU data were estimated by the company based on consultation with clinical experts. These resources included oncologist consultations, computed tomography (CT) scans, and laboratory tests (full blood count, liver function test, and renal function test) as presented in Table 39 of the CS. The company stated that the MRU estimates in its base case analysis were different from those reported in the ramucirumab appraisal (TA378[29]) for reasons of following the TAGS trial protocol and avoiding potential double counting issues with AE costs or end of life care costs. For completeness, the company conducted a scenario where MRU resources were assumed the same as reported in TA378 as shown in Table 40 of the CS.

The company sourced MRU unit costs mainly from NHS Reference Costs 2017/18[30], the Commercial Medicines Unit (CMU) electronic Marketing Information Tool (eMIT), and inflated values using Personal Social Services Research Unit (PSSRU[31]) indices as appropriate.

4.2.5.5.4 Off-treatment and post progression related costs

In its base case, the company assumed that routine MRU costs for patients were based on treatment status (i.e. receiving TFT or not receiving TFT), as opposed to progression status. These included an oncologist consultation every 3 cycles of treatment. In the TA378 scenario, the company, as in the case with on-treatment MRU, also included costs of pain control, distress management, blood transfusion, and radiotherapy. These costs are detailed in Appendix 1; in the company's base case the costs were £211 per 28 days for patients receiving TFT and £54 per 28 days for people not receiving TFT.

Following progression in the TAGS trial, patients could undergo surgery, radiotherapy or continue onto further rounds of systematic anti-cancer treatment (SACT) which was assumed to involve a 3-cycle course of docetaxel. These costs were applied during the first model cycle following progression. Post-progression MRU estimates were extracted from the TAGS study with unit costs sourced from NHS Reference Costs 2017/18[30] and eMIT. In the company's base case, the average total costs incurred upon progression were £1,327 for patients who had received TFT and £1,532 for those who had received

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

BSC. The higher costs for BSC were due to the greater observed level of post-progression treatment in the BSC arm of the TAGS study.

4.2.5.5.5 AE costs

The rationale for the AEs included in the model is provided in Section 4.2.5.2. The costs associated with each were primarily sourced from NHS Reference Costs 2017/18.[30] Table 11 presents the frequency of AEs observed within the TAGS study and the costs associated with their management. This resulted in an average total cost of £306 and £87 to resolve AEs associated with TFT and BSC, respectively. These cost estimates were applied as a fixed sum within the model's first cycle.

Table 11: AE costs

Adverse event	Occurrence rates		Assumed cost to resolve	Source
	TFT	BSC		
Neutropenia	46%	0%	£164.55	Assumption of FBC cost + outpatient medical oncologist consultation (based on clinical experts' opinion)
Anaemia	26%	10%		
Decreased neutrophil count	26%	0%		
Leukopenia	7%	0%		
Abdominal pain	5%	11%	£319.68	NHS Reference Costs (2017/18[30]): Weighted average of day case abdominal pain with and without interventions (FD05A and FD05B)
Ascites	6%	8%		
Decreased appetite	10%	7%	£75.98	PSSRU 2018: Unit cost of a dietician appointment[31]
Fatigue	7%	6%	£0.00	Assumption of zero cost was based on clinical experts' opinion
Ascites	6%	8%		
Febrile neutropenia	2%	0%	£4,619.81	Wehler *et al.* (2017[32]) and inflated using PSSRU inflation indices[31]
Nausea	4%	3%	£163.58	NHS Reference Costs (2017/18[30]): Outpatient attendance – General Medicine

BSC, best supportive care; FBC, full blood count; TFT, trifluridine/tipiracil

For those patients in both the TFT and BSC arm who receive SACT upon progression, AE costs were set equal to those of people initially receiving TFT.

4.2.5.5.6 End of life care costs

In the company's base case, end of life care costs (health and social care costs) for colorectal cancer patients reported within Round *et al.* (2015[33]) were inflated. These were applied for all patients upon entry to the "Dead" health state. Alternative costs for cancer patients with lung, breast, and prostate cancers were also reported in the same publication. Based on clinical advice, the company decided that

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

the end of life care costs incurred by colorectal cancer patients were relevant to this appraisal; the costs of other cancer types were used in scenario analysis.

4.2.6 Model validation and face validity check

The company validated its economic model using two approaches. The first was holding a clinical advisory board attended by twelve UK practicing oncologists specialising in GC who validated the key aspects and assumptions of the model. The second approach was an internal quality control check of the company's model by a third party.

4.2.7 Cost effectiveness results

Following the clarification process the company submitted a revised version of the model that included updated estimates of the cost-effectiveness of TFT. All the results presented in this section and in Section 4.2.8 use the revised model and all results use the established price for TFT after consideration of the PAS. Table 12 shows the results of the company's base case analysis for both the deterministic and probabilistic versions of the model. The probabilistic sensitivity analyses (PSA) results are based on 10,000 iterations run by the ERG. Based on the probabilistic version of the model, TFT plus BSC is expected to generate 0.153 additional QALYs at an additional cost of £6,923, compared with placebo + BSC. The corresponding ICER is £45,314 per QALY gained. The deterministic version of the company's model produces a similar ICER of £45,164 per QALY gained.

Figure 11 shows the cost-effectiveness acceptability curve (CEAC) produced by the ERG when running the company's base case. Figure 12 plots the PSA results on the cost-effectiveness plane. Figure 13 presents the resultant survival curves for the first five years of the company's model.

Table 12: The Company's base case results

Treatment	Total QALYs	Total Costs	ICER (£ per QALY gained)
Deterministic			
Placebo + BSC	0.349	[redacted]	
TFT + BSC	0.502	[redacted]	£45,164
PSA (run by the Evidence Review Group)			
Placebo + BSC	0.351	[redacted]	
TFT + BSC	0.504	[redacted]	£45,314

BSC, best supportive care; ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year; TFT, trifluridine/tipiracil

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

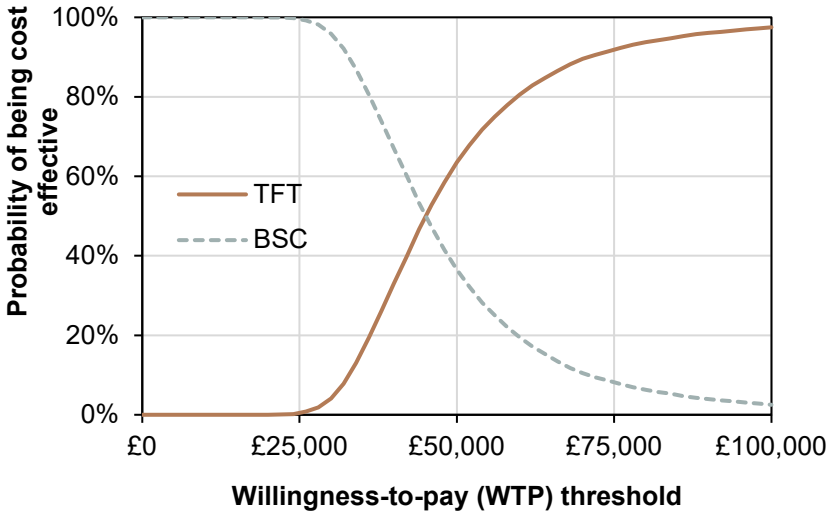


Figure 11: Company's base case cost-effectiveness acceptability curve

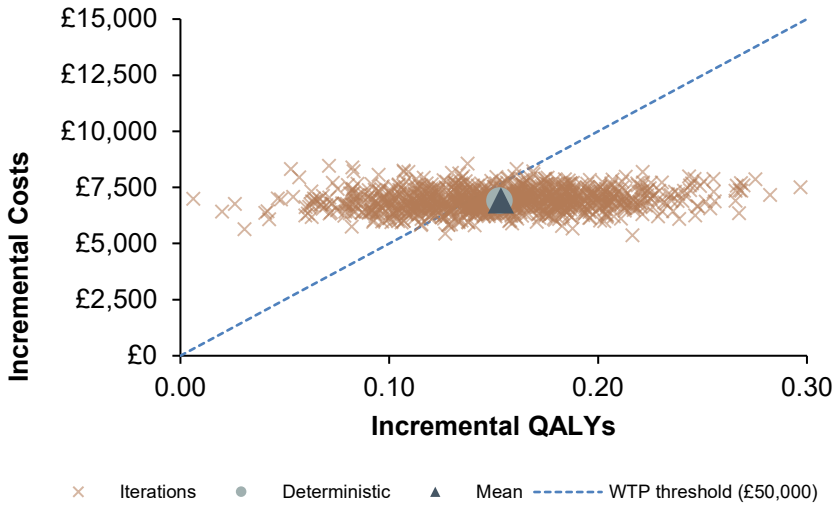


Figure 12: Company's base case cost-effectiveness plane

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

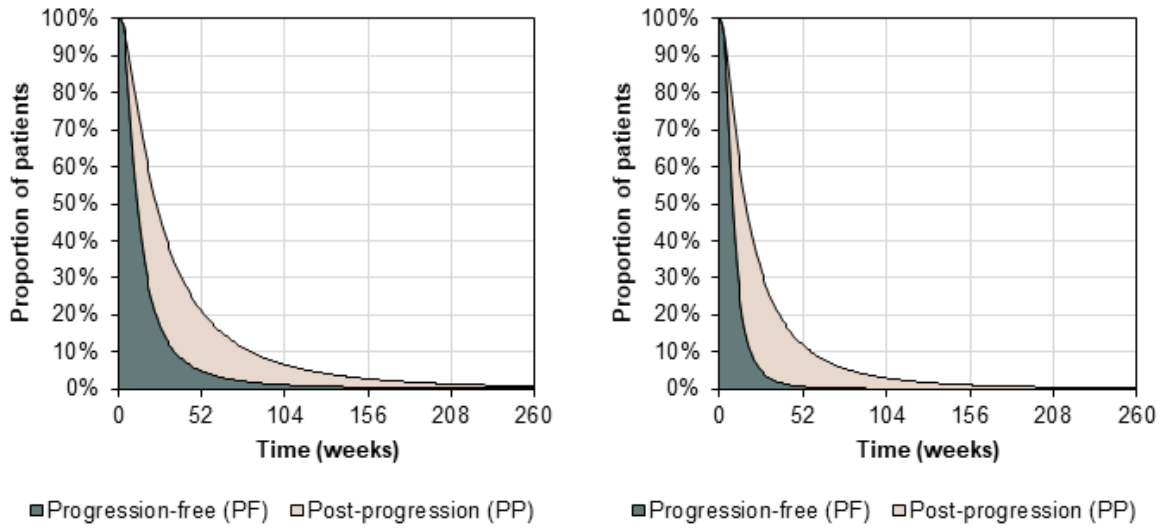


Figure 13: Company's base case survival curves (model traces) – BSC (left) and TFT (right)

4.2.7.1 Tornado diagrams

The company's tornado diagram, which shows the ten most influential parameters in terms of impact on ICER, is presented in Figure 43 of the CS. Within the tornado diagram, all parameters were varied between the upper and lower bounds of the 95% confidence intervals. The company stated that this analysis could not include parameters related to survival and utility as "*they are correlated (and so varying these in isolation would not be appropriate). Instead, the uncertainty associated with estimates of survival and utilities is discussed within the context of scenario analysis*." The ERG noted that BSA was not incorporated in the company's original one-way sensitivity analysis. In its response to clarification question B8,[7] the company estimated the standard deviation around the two parameters (mu and theta) of the fitted lognormal distribution. The resultant uncertainty was included in both one-way and probabilistic sensitivity analyses. This had a small impact on the ICER value.

The most influential parameters in this analysis were related to MRU frequencies and costs. None of the ICERs on the tornado plot exceeded £50,000 per QALY gained.

4.2.8 Sensitivity analyses

The company conducted sensitivity analyses, which included: (1) a range of scenario analyses, which included the effects of alternative survival extrapolations and data on the results; and (2) exploring the use of KM curves at different cut-points followed by extrapolation for the PFS data.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

4.2.8.1 Scenario and subgroup analyses

The company undertook several scenario analyses, which are presented in Tables 13 of the company's response to the clarification questions.[7] Generally, most scenarios produced ICERs that were similar to the company's base case ICER. Reducing the model time horizon resulted in non-linear increase in ICER values, and time horizons of two years or less resulted in ICERs, which were higher than £50,000 per QALY gained.

Eight out of the 12 tested survival parametric modelling scenarios of OS data presented in Table 13 of the company's clarification response[7] were associated with ICERs higher than £50,000 per QALY gained. These were: fitting independent exponential, independent generalised gamma, independent Gompertz, independent lognormal, independent Weibull, dependent exponential, dependent Gompertz, and the dependent Weibull. The company claimed that these curves provided a relatively poor fit to the KM curves.

All alternative PFS and TTD parametric models tested by the company resulted in ICERs, which were lower than £50,000 per QALY gained. The use of different spline-based models fitted to the PFS data produced ICERs, which were between £43,500 and £47,000 per QALY gained.

Within subgroup analyses, the company considered the entire TAGS population (with and without ramucirumab experience) as the data source for efficacy. Using the same survival models from the base case, this scenario produced an ICER, which was slightly higher than £50,000 per QALY gained. The ERG comments that the company did not undertake an assessment of which was the most appropriate curves to use in this population.

The ERG requested, in its clarification questions, a scenario analysis using only the European population with no prior treatment with ramucirumab. The company estimated new parameter values for the same curves fitted to the OS, PFS and TTD in the base case, which produced an ICER which was slightly above £49,000 per QALY gained. The company did not consider this scenario in its base case claiming that caution should be used when interpreting these results as this subgroup was not stratified for in the TAGS study.

4.2.8.2 Use of KM curves followed by parametric curve extrapolation for the PFS data

Due to the presence of "*kinks*" in the KM PFS data, the company conducted an additional scenario analysis to explore the impact of using the KM PFS curves for both treatment arms followed by the generalised gamma curves used within the base case analysis. The time cut-point between both sets of PFS curves were varied between 16 and 33 weeks in weekly increments. All cut-points produced ICERs that were between £44,000 and £45,200 per QALY gained.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

4.3 Critique of company's submitted economic evaluation by the ERG

4.3.1 Methods for reviewing the company's economic evaluation and health economic model

The ERG adopted a number of approaches to explore, interrogate and critically appraise the company's submitted economic evaluation and the underlying health economic model upon which this was based. These included:

- Scrutiny of the company's model and discussion of issues identified amongst the members of the ERG.
- Examination of the correspondence between the description of the model reported within the CS and the company's executable model.
- Re-running the DSA and PSA presented within the CS.
- Where possible, checking the parameter values used in the company's model against their original data sources.
- The use of expert clinical input to judge the credibility of the company's economic evaluation and the assumptions underpinning the model.

4.3.2 Adherence of the company's model to the NICE reference case

As shown in Table 13, the company's economic evaluation is generally in line with the NICE reference case.[30]

Table 13: Adherence of the company's model to the NICE reference case[30]

Element	Reference case	ERG comments
Type of economic evaluation	Cost-utility analysis with fully incremental analysis	The CS met the NICE reference case.[30]
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	The CS met the NICE reference case.[30] A 10-year time horizon was adopted. By this point, almost 100% of simulated patients were dead.
Synthesis of evidence on health effects	Based on trial outcome data and systematic review	The CS met the NICE reference case.[30] Health outcomes are modelled using the data collected in the TAGS study. The base case used a subgroup with no prior ramucirumab experience.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

Measuring and valuing health effects	Health effects should be expressed in QALYs. The EQ-5D is the preferred measure of HRQoL in adults.	The CS met the NICE reference case.[30]
Source of data for measurement of health-related quality of life	Reported directly by patients and/or carers	The CS met the NICE reference case.[30]
Source of preference data for valuation of changes in HRQoL	Representative sample of the UK population	EORTC-QLQ-C30 data collected in the TAGS study were mapped to EQ-5D-3L values. The mapping algorithm used in the company's base case was developed in a gastric cancer population. However, none of the patients were suffering from metastatic cancer and the sample size was small (n=48).
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	The CS met the NICE reference case, although the company makes a case for the end of life criteria being met.
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	The CS met the NICE reference case.[30]
Discount rate	The same annual rate for both costs and health effects (currently 3.5%)	The CS met the NICE reference case.[30]

4.3.3 ERG Critique of the modelling performed by the company

4.3.3.1 Model verification

The ERG checked and verified the implementation of the model and the methods for generating results. During this process, the ERG identified one minor implementation error, which was addressed by the

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

company in their clarification response to question B9. The implemented model appears to be generally in line with its description within the CS. KM curves were available for OS, PFS and TTD and provided in the model.

4.3.3.2 Correspondence of the model inputs and the original sources of parameter values

The ERG is satisfied that model parameters corresponded with their original source values. These were in line also with the parameter values reported in the CS. The only possible exception to this was potential discrepancies in the AE inputs, see Section 4.2.2.4. However, the ERG's exploratory analyses indicated that these issues would not affect the ICER significantly.

The ERG noted that the company's model uses arbitrary values to characterise the uncertainty in NHS Reference Costs by assuming 10% of the mean cost as its standard deviation and dividing it by the number of cases rather than number of data returns. In its response to clarification question B7,[7] the company reverted to 2012-13 NHS Reference Costs database to estimate the ratio of the standard error (SE) to the mean cost from the quartile data. The company subsequently concluded that a ratio of 5% could be used to account for the uncertainty of all NHS Reference Costs included in the model. The ERG highlights the relatively old NHS reference costs version used by the company and that the same ratio was used for all costs. However, the ERG expects these limitations to have minimal impact on the uncertainty in the probabilistic ICERs.

4.3.4 The main issues identified by the critical appraisal

Generally, the model was implemented well and the company provided reasonable responses to the ERG's clarification questions. However, the ERG identified four main issues within the model. These points are summarised in Box 1, with further details provided in the subsequent sections. The small number of issues are testament to the implementation of the decision problem by the company and the relative simple decision problem.

Box 1: Summary of the main issues identified within the company's health economic model

Summary of identified concerns within the company's health economic model:

1) Selection of the appropriate population for the base case analyses
2) Extrapolation of OS and PFS
3) The mapping of EORTC-QLQ30 to EQ-5D-3L
4) Exclusion of oral chemotherapy delivery fees

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

4.3.4.1 Selection of the appropriate population for the base case analyses

As indicated in Section 3.2.2.1, clinical advice sought by the ERG, and provided to NICE, suggested exposure to ramucirumab is not expected to influence the relative efficacy of TFT or prognosis. The company's base case uses the no prior ramucirumab population, which, therefore, may not be the most appropriate estimates for the purpose of decision making, although this population had fewer lines of previous treatment and a lower proportion of Japanese patients. The company's base case also included patients from Japan and the United States; clinical advice to the ERG suggested that a European subgroup may be more appropriate.

4.3.4.2 Extrapolation of OS and PFS

The ERG notes that the company considered a number of approaches in selecting its preferred base-case model, including the use of statistical goodness of fit, a quantile-quantile plot, a cumulative hazard plot, an empirical hazard plot, visual inspection and assessing the plausibility of longer-term projections. For extrapolating OS, the combined modelling approach provided lower AIC/BIC scores when compared with the independent modelling approach. However, the difference in scores were less than 3 points, hence it indicated that both models provided similar statistical goodness of fit to the data. By examining the plots for assessing the appropriateness of the combined modelling approach (with treatment as a covariate), the ERG believes that it was not clear that the combined modelling approach would be more appropriate for the OS. If the OS data were associated with a constant AF over time, the fitted survival curves would theoretically be the same using either the combined modelling or independent modelling approach (though this would be difficult to establish using "real" trial data, owing to limited sample sizes). The ERG notes that when using independent lognormal models increased ICER to above £50,000.

The company fitted spline models for PFS during the clarification process. The ERG notes that the combined modelling approach was used without justification, and the impact on the use of independent modelling approach is unclear.

The company, in its reply to clarification question B4,[7] chose the same parametric curves fitted for the whole TAGS trial population with no prior ramucirumab experience to be the selected curves for the European subpopulation (i.e. dependent lognormal distributions for the OS data, and independent generalised gamma distributions for the PFS data). The company did not mention the rationale their curve choice although AIC and BIC data were contained in the Excel model.

4.3.4.3 The mapping of EORTC-QLQ30 to EQ-5D-3L

The ERG noted that the mapping study used by the company, that of Kontodimopoulos *et al.*,[26] was derived from a small population (n=48) and whilst the patients all had gastric cancer, none had

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

metastatic cancer. The estimated utilities appeared to have a lack of face validity compared with those used in previous STAs where EQ-5D-3L data had been collected within the trial and where the patients were less heavily pre-treated (Table 10) as it would be expected that the PD state after third-line treatment would be lower than after first-, or second-line. Whilst the company identify limitations in the utilities collected in TA378 and TA208, (p146-147 of the CS[1]) this would not address the potential face validity concern.

A recent review of mapping algorithms from the EORTC-QLQ30 to the EQ-5D-3L[34] was identified by the ERG which stated that two algorithms were the best performing in external validation studies.[35, 36] These mapping algorithms use more of the EORTC-QLQ30 domains than Kontodimopoulos *et al.*[26] which only uses physical functioning, emotional functioning and global health status as predictors of EQ-5D-3L.

The ERG asked the company to provide ICERs when each of the two mappings were used, in order to inform the committee of the sensitivity of the results to the chosen mapping algorithm. However, this was not undertaken by the company for the following reasons. The company stated that the "*Versteegh et al. study does not utilise a UK tariff (and so is not aligned with the NICE reference case). However, outside of the tariff used, the study was conducted in only haematological cancers (multiple myeloma and non-Hodgkin's lymphoma)*" and that "*Longworth et al. primarily considers patients with multiple myeloma (n=572 of 771), as well as patients with breast or lung cancer (i.e. no patients with a gastrointestinal cancer). For the multiple myeloma cohort, patients were taken from the VISTA trial – a Phase III randomised open-label trial for newly-diagnosed patients. In the other two populations (breast and lung cancer), real-world data were collected from the Vancouver Cancer Clinic.*" The company quote guidance from NICE DSU TSD 10[37] stating that *""... we recommend that careful consideration is given to the generalisability of the mapping function to the target population, including the range of disease severity over which the function was estimated and the potential for systematic differences in the populations that could impact on the health state utility values.*" (Section 3.2.5 of NICE DSU TSD 10[37]).

A similar conclusion is also made by Woodcock and Doble who state that "*The most appropriate mapping algorithm to apply in practice may depend on the disease severity of the patient sample whose utility values are being predicted.*" Both the NICE DSU TSD 10[37] and Woodcock and Doble[34] would lead the ERG to question whether the mapping algorithm from Kontodimopoulos *et al.*[26] which did not include patients with metastatic cancer would be appropriate in a population in which "*all patients had heavily pre-treated (i.e. two or more previous lines of therapy) metastatic gastric cancer*" and whereby the estimated life expectancy under current standard care was in the region of six months. The Kontodimopoulos *et al.* paper[26] states that "*No patients were suffering from metastases of the cancer to*

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

other organs, which could further affect their HRQoL negatively." The ERG notes, however, that an alternative mapping algorithm by Marriott *et al*.[38] which considers a metastatic colorectal cancer population was provided by the company. These values were higher for both PFS and PD than those generated using the mapping of Kontodimopoulos *et al.*[26]

The ERG is not contending that the mapping algorithms produced by Versteegh *et al.*[35] and Longworth *et al*.[36] are unquestionably better than that of Kontodimopoulos *et al.,*[26] and accept the criticisms of the alternative mappings put forward by the company. However, the ERG believes that the ICERs produced when these mappings are used would be informative to the committee and that the sensitivity analyses should have been performed. As the ERG does not have access to the data required to calculate utility estimates based on the alternative mapping algorithms, this remains an area of considerable uncertainty.

The ERG also notes that the compliance rate for filling in the EORTC QLQ-C30 was 84% and thus there may be the potential for responder bias within the study, however, if there was, the extent to which this would influence the results in unknown.

4.3.4.4 Exclusion of oral chemotherapy delivery fees

The company assumed in its base case analysis that there would be no administration costs regularly associated with TFT due to its oral route of administration. However, as NHS England noted in an recent STA[39] "*Trusts will regard [TFT] as chemotherapy and may charge the oral delivery tariff SB11Z (£120) each time [TFT] is given to patients*." This will be in addition to other consultation costs already included in the economic model. The ERG does not know NICE's position on this but contend that this may be seen as a transfer payment and has excluded this from the base case but has explored it within scenario analyses.

4.4 Exploratory analyses undertaken by the ERG

This section presents the methods of the ERG's exploratory analyses.

4.4.1 Exploratory analyses based on and whether it should be assumed that the prognoses of the patients and the efficacy of TFT are affected by prior ramucirumab use and amending the chosen geographical region for the analyses

The ERG has explored the impacts of alternative assumptions relating to prognoses of patients considered for TFT, whether prior ramucirumab use affects the HR for TFT compared to BSC, and which geographical region is most appropriate. Each of the three components had two choices, which leads to eight potential scenarios. These are summarised in Section 4.1.4.4, however beforehand each component will be detailed.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

4.4.1.1 Exploring the relationship between prior ramucirumab treatment and prognoses

The company's base case assumes that prior ramucirumab treatment affects the survival of patients who would be considered for TFT. As such, the company use the OS data in its base case for patients who have not had prior ramucirumab treatment. Clinical advice provided to the ERG suggested that it was unclear whether prior ramucirumab leads to a different prognosis and that the OS related to all patients in the TAGS study[2] may be more appropriate. (See Section 3.2.2.2).

4.4.1.2 Exploring the relationship between prior ramucirumab treatment and the relative efficacy of TFT

The company's base case assumes that the most appropriate estimation of the relative efficacy of TFT is derived from the no prior ramucirumab group. As such, the company use the AF in its base case for patients who have not had prior ramucirumab treatment, which matches the prognosis group in Section 4.4.1.1. Clinical advice provided to the ERG and to NICE suggested that the HR or AF would be expected to be independent of prior ramucirumab use (See Section 3.2.2.2). The ERG believes that a more accurate estimate of the efficacy of TFT would therefore come from all patients irrespective of prior ramucirumab use.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

4.4.1.3 Exploring the relationship between prior ramucirumab treatment and TFT relative efficacy on different populations

The company's base case assumes that the HR or AF would be independent of geographical region and combines patients from Japan and the ROW and uses this in its base case. However, it is argued in the CS that gastric cancer operates differently in Japanese patients and it is noted that results from Japanese patients cannot be assumed to be generalisable to non-Asian patients. (Section 3.1.5). Furthermore, the clinical advice provided to the ERG was that the EU subgroup would be more generalisable than a group including Japanese and American patients (see Section 3.2.2.3). Whilst an analysis of the Europe only geographical area breaks the stratification within the TAGS study,[2] the ERG notes that the study was stratified on Japanese vs the ROW and that the European component was approximately 95% of the ROW so it is anticipated that the inaccuracy caused by this limitation may not be large.

4.4.1.4 Summarising the eight potential scenarios

The eight scenarios are shown in Table 14. The first scenario is the company's base case where prior ramucirumab is assumed to affect both prognosis and the relative efficacy of TFT and all geographical regions are used. The second scenario is a company scenario analysis, which uses all patients from the TAGS study. The third scenario assumes that prior ramucirumab treatment may affect disease prognosis, but does not affect the relative efficacy of TFT. In the fourth scenario, prior ramucirumab treatment does not affect disease prognosis but impacts on TFT relative efficacy. To run scenarios 3 and 4, an estimate of the treatment effect is taken from scenarios 2 and 1 respectively, meaning that only the use of dependent models could be explored. Scenarios 1 to 4 are replicated for the European population in Scenarios 5-8. However, the ERG did not have the data required to explore Scenarios 6 to 8. The company provided the data for Scenario 5 in its response to clarification question B4.[7]

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

Table 14: The eight scenarios defined by the ERG

Scenario	Is prognosis from non-ramucirumab patients most appropriate?	Is the HR or AF from non-ramucirumab patients most appropriate?	Is the entire TAGS study population more appropriate than the European geographical area?
1*	✓	✓	✓
2†	✗	✗	✓
3	✓	✗	✓
4	✗	✓	✓
5†	✓	✓	✗
6☯	✗	✗	✗
7	✓	✗	✗
8	✗	✓	✗

*The company base case; †A company scenario analysis; ☯Tentative ERG base case.

4.4.2 Analyses exploring the uncertainty in survival curve fits

The ERG selected the three best fitting survival distributions to run the analysis for each of the eight scenarios defined in Table 14. These were determined based on the AIC/BIC scores provided in Table 21 of the CS for the overall population with no prior ramucirumab, and in the revised model for the European subgroup. These were the lognormal, the Weibull and the log-logistic for the all geographical region analyses, and the lognormal, the log-logistic and the generalised gamma for the European population. Where possible, independent models were evaluated as well as dependent models.

4.4.3 Impact of alternative mapping studies

As indicated in Section 4.3.4.3, the impact of using different mapping algorithms was not explored by the company. The ERG noted that the company's preferred mapping study did not include metastatic patients; the ERG believes that it is likely that these patients would have a lower utility than patients without metastases. The ERG performed an analysis using the values from the most recent STA for advanced gastric cancer or gastro-oesophageal junction adenocarcinoma which were 0.729 for those in the PFS and 0.587 for those in the PD state. (Table 10)

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

4.4.4 Impact of including oral chemotherapy delivery fees

As indicated in Section 4.3.4.4, the ERG explored the inclusion of the oral delivery tariff for chemotherapy (SB11Z) for outpatient setting in its scenario analyses. This resulted in a cost of £131.61 applied every 28 days for patients receiving TFT.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

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

This section presents the results of the ERG's exploratory analyses.

5.1 The impact of selecting different populations and using different curve fits

Table 15 shows the results for the scenarios 1-5 (as defined in Table 14). The ICERs varied considerably among different scenarios and models. Generally, independent models gave higher ICERs, whereas log-logistic models produced lower ICERs followed by lognormal and Weibull models. The company selected a dependent lognormal to use within the base case as it had the lowest AIC, and the ERG agrees that the lognormal is more likely to be appropriate than the log-logistic or the Weibull although the other two distributions remain plausible. The ERG prefers the use of independent models because the use of independent modelling approaches avoids making any assumptions about constant HR or AF over time, and would provide the same fitted curves as to the combined modelling approach if either assumption holds.

The ERG could not produce ICERs for its tentative base case (Scenario 6) but notes that when moving from Scenario 1 to 2, the ICER increases by £4,000 to £5,000. In the absence of further evidence, it may be appropriate to assume that this level of increase would also apply when moving from Scenario 5 to Scenario 6. Such calculations would indicate ICERs of over £64,000 when using independent models and in excess of £50,000 per QALY gained when assuming dependent models. However, the analyses would need to be undertaken to provide an accurate estimation. As the model appeared linear, only deterministic analyses have been run by the ERG.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

Table 15: ERG's exploratory analysis regarding the impact of prior ramucirumab treatment and geographical region

Scenario	Independent models			Dependent models		
	Lognormal	Log-logistic	Weibull	Lognormal	Log-logistic	Weibull
1	£51,642	£46,942	£61,310	£45,164*	£42,208	£58,363
2	£55,600	£52,655	£66,137	£50,191	£47,449	£64,318
3				£50,278	£47,750	£65,129
4				£45,076	£41,926	£57,652
	Lognormal	Log-logistic	Generalised gamma	Lognormal	Log-logistic	Generalised gamma
5	£68,061	£59,564	£169,370	£49,067	£45,068	£46,024
6	Not evaluable due to data unavailability					
7						
8						

*Company's base case

Appendix 2 presents the results of Table 15 in terms of differential costs and QALYs. The different scenarios and model selection have little impact on cost differences but have a proportionately higher impact of the difference in QALYs. In Scenario 5 the ICER from the independent generalised gamma models was markedly larger than other fits

[redacted]

5.2 Impact of decrementing utility values due to patients having metastatic disease

The impact of using utility values from TA378 increased the ICER. The company's base case ICER increased from £45,164 to £47,857 and in Scenario 5, using independent lognormal models, from £68,061 to £70,905 per QALY gained.

5.3 Impact of including the oral administration delivery fees

In a scenario analysis, the ERG explored the impact of adding the delivery fees as detailed in Section 4.4.3. This increased the differential costs between the two compared interventions by approximately [redacted] and increased the company's base case ICER from £45,164 to £48,592.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

6 END OF LIFE

The company puts forward the case, in Section B.2.13 of its CS, that TFT meets the NICE End of Life criteria. These criteria are:

- The treatment is indicated for patients with a short life expectancy, normally less than 24 months and;
- There is sufficient evidence to indicate that the treatment offers an extension to life, normally of at least an additional 3 months, compared to current NHS treatment.

The ERG believes that TFT meets the first criterion because the mean life years associated with BSC in the model was 0.514 years (6.2 months) for patients without prior ramucirumab use.

Whether TFT meets the second criterion is less straightforward as the life extension associated with TFT estimated in the model is 0.226 years (2.7 months) which is below the 3 months normally required. The company cite precedent in two prior NICE appraisals to support their case, although only one had a survival extension of less than 3 months. This is nab-paclitaxel for (untreated) metastatic pancreatic cancer (TA476[40]) which was estimated to have a mean life extension of 2.4 months, but was assumed to meet the end of life criteria due to the short median life expectancy without treatment of 6.6 months. The company states that the proportional life improvement associated with TFT in mGC and GEJ is superior to that estimated in TA476. For completeness the ERG has reproduced the text from TA476. "*The committee noted that the survival data were mature and therefore considered that the survival gain estimate was robust. It recognised that this survival gain should be considered in the context of the very poor prognosis for metastatic pancreatic cancer. The committee noted that the survival gain was below what is normally considered appropriate for the extension-to-life criterion to be met (that is, it was less than 3 months). However, it agreed that the survival gain was particularly important relative to the average survival of people with this condition, and therefore this criterion could be accepted as met in this circumstance. The committee concluded that, for the comparison with gemcitabine monotherapy, nab-paclitaxel plus gemcitabine met the criteria to be considered a life-extending end-of-life treatment.*" The ERG leaves the decision on whether TFT meets the second criterion to the NICE Appraisal Committee.

Table 16 presents the life expectancy gains associated with TFT and other relevant technologies appraised in different scenarios.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

Table 16: Survival gain associated with TFT and with the precedent cited by the company

Scenario	OS associated with SoC (months)	OS associated with the appraised technology (months)	OS gained with the technology (months, % gained)
The TAGS trial (no prior ramucirumab, whole population)	Median: 3.3	Median: 6.0	2.7 (82%)
The TAGS trial (whole population regardless of ramucirumab use)	Median: 3.6	Median: 5.7	2.1 (58%)
The company's base case model (no prior ramucirumab, whole population)	Mean: 6.2	Mean: 8.9	2.7 (44%)
The company's model (whole population regardless of ramucirumab use)	Mean: 6.2	Mean: 8.5	2.3 (37%)
TA476 main trial results	Median: 6.6	Median: 8.7	2.1 (32%)
TA476 economic model results	Mean: 8.7	Mean: 11.1	2.4 (28%)

OS, overall survival; SoC, standard of care

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

7 OVERALL CONCLUSIONS

The TAGS study reported a HR 0.69 (95% CI 0.56–0.85) for OS, and a difference in median survival of 2.1 months between arms. Analyses adjusted for relevant prognostic factors gave a similar HR. PFS was also positively affected, with a HR 0.57 (95% CI 0.47–0.70) and a 0.2 month difference between arms. Small benefits were reported for response rates and duration of response as may be expected given the stage of disease, however, the disease control rate was significantly improved. Health related quality of life was shown to be largely maintained with TFT treatment.

In subgroup analyses, for OS, patients with prior ramucirumab treatment had a HR of 0.76 (95% CI 0.53-1.09) and those without a HR of 0.66 (95% CI 0.51-0.85). Patients from Japan had a HR of 0.77 (95% CI 0.46-1.30), and those from ROW had a HR of 0.68; (95% CI 0.54-0.85). Patients from Europe had a HR of 0.67 (95% CI 0.53-0.86). The ERG requested an analysis of patients from Europe without prior ramucirumab treatment, for which a HR was not reported, [redacted]. PFS subgroup analyses were largely similar to the main analysis of PFS.

Clinical advice to the ERG and NICE suggested that there is no strong indication that prior ramucirumab treatment affects prognosis, and was unlikely to affect the efficacy of TFT. Clinical advice also indicated that European patients would have the highest generalisability to the decision problem due to biological and/or treatment pathway differences between Europe, the USA and particularly, Japan.

As shown within the ERG exploratory analyses the company's base case ICER is one of the lower estimates amongst the analyses undertaken by the company and the ERG. Factors that increase the ICER include: the use of independent rather than dependent models; assuming that prior use of ramucirumab does not affect prognosis; assuming that prior use of ramucirumab does not affect the efficacy of TFT; and assuming a European geographical area rather than the full TAGS study. The clinical advice provided to the ERG and NICE resulted in a tentative base case being put forward by the ERG (Scenario 6). This scenario could not be evaluated, but it is expected to have ICERs which are higher than those for Scenario 5 whereby the ICER was approximately £68,000 when an independent lognormal model was used and £49,000 when a dependent lognormal model was used. The ERG prefers the use of independent curves rather than dependent ones.

The ERG notes that the company explored an alternative mapping in sensitivity analysis, but declined to explore the alternative mappings proposed by the ERG. The ERG believes that sensitivity analyses should have been performed that assessed the impact of these mapping algorithms on the ICER. The

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

ERG performed exploratory analyses that showed that reducing the assumed utility value in the PFS and PD state increased the ICER.

In summary, Based on the analyses provided by the company and the ERG's exploratory analyses the ERG believes that the cost per QALY gained of TFT compared with BSC is likely to be in excess of £50,000. Whilst, the ERG's tentatively preferred scenario could not be evaluated, many component factors such as: using independent curves; assuming that prior ramucirumab use does not affect prognosis; assuming that prior ramucirumab use does not affect the relative treatment effect of TFT; using a European population; and reducing utility values, all increase the ICER (£45,164). The ERG notes that some of these factors, in isolation, increase the ICER to greater than £50,000 per QALY gained.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

8 REFERENCES

1. Servier Laboraties. Trifluridine–tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies [ID1507]. Company evidence submission. 2019.
2. Shitara K, Doi T, Dvorkin M, Mansoor W, Arkenau H-T, Prokharau A, *et al.* Trifluridine/tipiracil versus placebo in patients with heavily pretreated metastatic gastric cancer (TAGS): a randomised, double-blind, placebo-controlled, phase 3 trial. *The Lancet Oncology* 2018;19:1437-48.
3. Davidson M, Cafferkey C, Goode EF, Kouvelakis K, Hughes D, Reguera P, *et al.* Survival in Advanced Esophagogastric Adenocarcinoma Improves With Use of Multiple Lines of Therapy: Results From an Analysis of More Than 500 Patients. *Clinical Colorectal Cancer* 2018;17:223-30.
4. Kang Y-K, Boku N, Satoh T, Ryu M-H, Chao Y, Kato K, *et al.* Nivolumab in patients with advanced gastric or gastro-oesophageal junction cancer refractory to, or intolerant of, at least two previous chemotherapy regimens (ONO-4538-12, ATTRACTION-2): a randomised, double-blind, placebo-controlled, phase 3 trial. *The Lancet* 2017;390:2461-71.
5. National Institute for Health and Care Excellence. Trifluridine–tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies. Final scope. 2019.
6. British Medical Association. British National Formulary: March -September 2019. 2019.
7. Servier Laboraties. Trifluridine–tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies [ID1507]. Clarification Response. 2019.
8. Centre for Reviews and Dissemination. CRD's guidance for undertaking reviews in health care.: University of York.; 2008.
9. Li J, Qin S, Xu J, Xiong J, Wu C, Bai Y, *et al.* Randomized, Double-Blind, Placebo-Controlled Phase III Trial of Apatinib in Patients With Chemotherapy-Refractory Advanced or Metastatic Adenocarcinoma of the Stomach or Gastroesophageal Junction. *Journal of Clinical Oncology* 2016;34:1448-54.
10. NICE. Single technology appraisal: User guide for company evidence submission template. Instructions for companies.; 2017.
11. Higgins JPT, Savovic J, Page MJ, Sterne JAC. Revised Cochrane risk-of-bias tool for randomized trials (RoB 2) TEMPLATE FOR COMPLETION. *https://sitesgooglecom/site/riskofbiastool/welcome/rob-2-0-tool* 2019.
12. Li J, Qin S, Xu J, Guo W, Xiong J, Bai Y, *et al.* Apatinib for chemotherapy-refractory advanced metastatic gastric cancer: results from a randomized, placebo-controlled, parallel-arm, phase II trial. *J Clin Oncol* 2013;31:3219-25.
13. Bando H, Doi T, Muro K, Yasui H, Nishina T, Yamaguchi K, *et al.* A multicenter phase II study of TAS-102 monotherapy in patients with pre-treated advanced gastric cancer (EPOC1201). *European Journal of Cancer* 2016;62:46-53.
14. Laboratories S. Clinical Study Report TAGS. 2018.
15. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. *Journal of clinical oncology* 1998;16:139-44.
16. Mansoor W. Trifluridine–tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies [ID1507]. Clinical expert statement. 2019.
17. Smyth E. Trifluridine–tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies [ID1507]. Clinical expert statement. 2019.
18. Office of National Statistics. Ethnicity and National Identity in England and Wales: 2011.; 2012. https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/articles/ethnicityandnationalidentityinenglandandwales/2012-12-11 (Accessed 06.08.19).
19. CTCAE Files.; 2019. https://evs.nci.nih.gov/ftp1/CTCAE/About.html (Accessed 07.08.19).
20. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. 2013.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

21. Latimer NR. NICE DSU Technical Support Document 14: Survival Analysis For Economic Evaluations Alongside Clinical Trials - Extrapolation With Patient-Level Data. School of Health and Related Research, University of Sheffield, UK 2011.
22. Jackson C, Metcalfe P, Amdahl J. Flexible Parametric Survival and Multi-State Models. R package version 1.1.1. 2019.
23. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. . 2013.
24. Office of National Statistics. Dataset: National life tables: UK. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesunitedkingdomreferencetables (Accessed June).
25. Royston P, Parmar M. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. . *Stat Med* 2002;21:2175-97.
26. Kontodimopoulos N, Aletras VH, Paliouras D, Niakas D. Mapping the cancer-specific EORTC QLQ-C30 to the preference-based EQ-5D, SF-6D, and 15D instruments. *Value Health* 2009;12:1151-7.
27. Health Economics Research Centre. HERC Database of mapping studies. http://www.herc.ox.ac.uk/downloads/herc-database-of-mapping-studies. (Accessed June).
28. Søren H, Ulrich H, Yan JP. Package 'geepack': Generalized Estimating Equation Package Version 1.2-1. 2016.
29. NICE. Ramucirumab for treating advanced gastric cancer or gastro–oesophageal junction adenocarcinoma previously treated with chemotherapy [TA378]. 2016. https://www.nice.org.uk/guidance/ta378/resources/ramucirumab-for-treating-advanced-gastric-cancer-or-gastrooesophageal-junction-adenocarcinoma-previously-treated-with-chemotherapy-82602795958981 (Accessed May).
30. NHS Improvement. Reference costs. 2018. https://improvement.nhs.uk/resources/reference-costs/ (Accessed 29 July 2019).
31. Personal Social Services Research Unit. Unit Costs of Health and Social Care. 2015.
32. Wehler E, Zhao Z, Pinar Bilir S, Munakata J, Barber B. Economic burden of toxicities associated with treating metastatic melanoma in eight countries. *The European journal of health economics : HEPAC : health economics in prevention and care* 2017;18:49-58.
33. Round J, Jones L, Morris S. Estimating the cost of caring for people with cancer at the end of life: A modelling study. *Palliative medicine* 2015;29:899-907.
34. Woodcock F, Doble B, Consortium C. Mapping the EORTC-QLQ-C30 to the EQ-5D-3L: An Assessment of Existing and Newly Developed Algorithms. *Med Decis Making* 2018;38:954-67.
35. Versteegh M, Leunis A, Luime J, Boggild M, Uyl-de Groot C, Stolk E. Mapping QLQ-C30, HAQ, and MSIS-29 on EQ-5D. . *Med Decis Making* 2012;32:554-68.
36. Longworth L, Yang Y, Young T, Mulhern B, Hernández Alava M, Mukuria C*, et al.* Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey. . *Health Technol Assess* 2014;18:1-224.
37. Longworth L, Rowen D. NICE DSU Technical Support Document 10: The Use of Mapping Methods to Estimate Health State Utility Values 2011.
38. Marriott E-R, van Hazel G, Gibbs P, Hatswell A. Mapping EORTC-QLQ-C30 to EQ-5D-3L in patients with colorectal cancer. . *J Med Econ* 2017;20:193-9.
39. NICE. Olaparib for maintenance treatment of recurrent, platinum-sensitive ovarian, fallopian tube and peritoneal cancer that has responded to platinum-based chemotherapy (including a review of technology appraisal no. 381) [ID1296]. 2019. https://www.nice.org.uk/guidance/indevelopment/gid-ta10303 (Accessed 13.08.19).
40. NICE. Paclitaxel as albumin-bound nanoparticles with gemcitabine for untreated metastatic pancreatic cancer. TA476.; 2017. https://www.nice.org.uk/guidance/TA476 (Accessed 29 July 2019).

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

9 APPENDICES

Appendix 1: Summary of Medical Resource Costs used within the model

Table 17 summarises the frequencies of MRU used in the company's base case and scenario analyses and the resultant costs applied per 28-day treatment cycle. Differences between the scenario analysis and the base case are underlined in the scenario analysis data.

Table 17: MRU frequencies used in the company's base case and scenario analyses per treatment and progression status

MRU item	Company's base case				Company's scenario analysis (TA378)			
	TFT + BSC		Placebo + BSC		TFT + BSC		Placebo + BSC	
	PF on treatment	PF off treatment or PP	PF on treatment	PF off treatment or PP	PF on treatment	PF off treatment or PP	PF on treatment	PF off treatment or PP
Oncologist consultations	1.00	0.33	0.33	0.33	1.00	0.33	0.33	0.33
CT scan	0.50	0.00	0.00	0.00	0.33	0.00	0.00	0.00
FBC	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
LFT	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
RFT	1.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Pain control*	0.00	0.00	0.00	0.00	471.5	704.5	704.5	704.5
Distress management†	0.00	0.00	0.00	0.00	2.52	4.06	4.06	4.06
Blood transfusion**	0.00	0.00	0.00	0.00	0.08	0.22	0.22	0.22
Radiotherapy††	0.00	0.00	0.00	0.00	0.13	0.11	0.11	0.11
Total MRU cost***	£210.88	£54.02	£54.02	£54.02	£395.65	£324.45	£324.45	£324.45

BSC, best supportive care; CT, computerised tomography; FBC, full blood count; LFT, liver function test; PF, progression-free; PP, post-progression; RFT, renal function test; TFT, trifluridine/tipiracil

* Average number of mg of morphine required per patient per 28-day treatment cycle. This is based on 42.1% of patients on TFT requiring 40 mg of morphine per day versus 62.9% of those who are not on TFT.

† Average number of cognitive behavioural therapy (CBT) sessions undergone per patient per 28-day treatment cycle. This is based on 10.5% of patients on TFT undergoing six CBT sessions per week versus 16.9% of those who are not on TFT.

** Average number of red blood cell (RBC) transfusions required per patient per 28-day treatment cycle. This is based on 8.8% of patients on TFT requiring 1 RBC transfusion per month versus 23.8% of those who are not on TFT.

†† Average number of fractions of radiotherapy required per patient per 28-day treatment cycle. This is based on 14.0% of patients on TFT requiring 1 radiotherapy fraction per month versus 11.9% of those who are not on TFT.

*** Total cost was calculated by multiplying the item cost by its frequency of usage.

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

Appendix 2: Detailed Results for the exploratory analyses undertaken by the ERG

Table 18 provides more detailed results for each of the scenarios and statistical fits explored by the ERG than in the main document.

Table 18: Granular results for each of the scenarios and statistical fits explored by the ERG

Treatment	Total QALYs	Total Costs	ICER (£ per QALY gained)
Scenario 1 (independent lognormal models)			
Placebo + BSC	0.360	[redacted]	
TFT + BSC	0.493	[redacted]	£51,642
Scenario 1 (independent log-logistic models)			
Placebo + BSC	0.377	[redacted]	
TFT + BSC	0.524	[redacted]	£46,942
Scenario 1 (independent Weibull models)			
Placebo + BSC	0.328	[redacted]	
TFT + BSC	0.440	[redacted]	£61,310
Scenario 1 (dependent lognormal models)			
Placebo + BSC	0.349	[redacted]	
TFT + BSC	0.502	[redacted]	£51,642
Scenario 1 (dependent log-logistic models)			
Placebo + BSC	0.367	[redacted]	
TFT + BSC	0.531	[redacted]	£42,208
Scenario 1 (dependent Weibull models)			
Placebo + BSC	0.326	[redacted]	
TFT + BSC	0.444	[redacted]	£58,363
Scenario 2 (independent lognormal models)			
Placebo + BSC	0.354	[redacted]	
TFT + BSC	0.471	[redacted]	£55,600
Scenario 2 (independent log-logistic models)			
Placebo + BSC	0.372	[redacted]	
TFT + BSC	0.496	[redacted]	£52,655
Scenario 2 (independent Weibull models)			
Placebo + BSC	0.329	[redacted]	
TFT + BSC	0.427	[redacted]	£66,137
Scenario 2 (dependent lognormal models)			
Placebo + BSC	0.346	[redacted]	

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

TFT + BSC	0.476	[redacted]	£50,191
Scenario 2 (dependent log-logistic models)			
Placebo + BSC	0.364	[redacted]	
TFT + BSC	0.501	[redacted]	£47,449
Scenario 2 (dependent Weibull models)			
Placebo + BSC	0.328	[redacted]	
TFT + BSC	0.429	[redacted]	£64,318
Scenario 3 (dependent lognormal models)			
Placebo + BSC	0.347	[redacted]	
TFT + BSC	0.477	[redacted]	£50,278
Scenario 3 (dependent log-logistic models)			
Placebo + BSC	0.365	[redacted]	
TFT + BSC	0.502	[redacted]	£47,750
Scenario 3 (dependent Weibull models)			
Placebo + BSC	0.324	[redacted]	
TFT + BSC	0.423	[redacted]	£65,129
Scenario 4 (dependent lognormal models)			
Placebo + BSC	0.348	[redacted]	
TFT + BSC	0.502	[redacted]	£45,076
Scenario 4 (dependent log-logistic models)			
Placebo + BSC	0.366	[redacted]	
TFT + BSC	0.531	[redacted]	£41,926
Scenario 4 (dependent Weibull models)			
Placebo + BSC	0.330	[redacted]	
TFT + BSC	0.449	[redacted]	£57,652
Scenario 5 (independent lognormal models)			
Placebo + BSC	0.363	[redacted]	
TFT + BSC	0.462	[redacted]	£68,061
Scenario 5 (independent log-logistic models)			
Placebo + BSC	0.379	[redacted]	
TFT + BSC	0.493	[redacted]	£59,564
Scenario 5 (independent generalised gamma models)			
Placebo + BSC	0.417	[redacted]	
TFT + BSC	0.457	[redacted]	£169,370
Scenario 5 (dependent lognormal models)			

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved

Placebo + BSC	0.342	[redacted]	
TFT + BSC	0.480	[redacted]	£49,067
Scenario 5 (dependent log-logistic models)			
Placebo + BSC	0.358	[redacted]	
TFT + BSC	0.508	[redacted]	£45,068
Scenario 5 (dependent generalised gamma models)			
Placebo + BSC	0.356	[redacted]	
TFT + BSC	0.503	[redacted]	£46,024

BSC, best supportive care; ICER, incremental cost-effectiveness ratio; QALY, quality adjusted life year; TFT, trifluridine/tipiracil

Copyright 2020 Queen's Printer and Controller of HMSO. All rights reserved