NICE National Institute for Health and Care Excellence

# Single Technology Appraisal

# Trifluridine–tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies [ID1507]

# Committee Papers

**NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE**

**SINGLE TECHNOLOGY APPRAISAL**

# Trifluridine–tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies [ID1507]

**Contents:**

The following documents are made available to consultees and commentators:

1. **Response to consultee, commentator and public comments on the Appraisal Consultation Document (ACD)**

2. **Comments on the Appraisal Consultation Document from Servier Laboratories**
   a. ACD response
   b. Additional ACD response
   c. Company's response to ERG's questions on company ACD response

3. **Consultee and commentator comments on the Appraisal Consultation Document from:**
   a. NHS England

4. **Comments on the Appraisal Consultation Document from experts:**
   a. Dr Elizabeth Smyth – clinical expert, nominated by Servier Laboratories Ltd
   b. Dr Wasat Mansoor – clinical expert, nominated by the Royal College of Physicians

   **Comments on the Appraisal Consultation Document received through the NICE website**
   *None*

5. **Evidence Review Group critique of company comments on the ACD**

6. **Weighting analyses based on third line population post-ACM2** by Servier Laboratories Ltd

7. **Clarification questions and company responses post-ACM2:**
   a. Company response 1 – Impact of propensity score on time to event data
   b. Company response 2 – Method of applying weights

8. **ERG critique of company weighting analyses post-ACM2** provided by the School of Health and Related Research (ScHARR)

*Any information supplied to NICE which has been marked as confidential, has been redacted. All personal information has also been redacted.*

# Trifluridine–tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies [ID1507]

### Single Technology Appraisal

## Response to consultee, commentator and public comments on the Appraisal Consultation Document (ACD)

**Type of stakeholder:**

**Consultees –** Organisations that accept an invitation to participate in the appraisal including the companies, national professional organisations, national patient organisations, the Department of Health and the Welsh Government and relevant NHS organisations in England. Consultees can make a submission and participate in the consultation on the appraisal consultation document (ACD; if produced). All non-company consultees can nominate clinical experts and/or patient experts to verbally present their personal views to the Appraisal Committee. Company consultees can also nominate clinical experts. Representatives from NHS England and clinical commissioning groups invited to participate in the appraisal may also attend the Appraisal Committee as NHS commissioning experts. All consultees have the opportunity to consider an appeal against the final recommendations, or report any factual errors, within the final appraisal determination (FAD).

**Clinical and patient experts and NHS commissioning experts** – The Chair of the Appraisal Committee and the NICE project team select clinical experts and patient experts from nominations by consultees and commentators. They attend the Appraisal Committee meeting as individuals to answer questions to help clarify issues about the submitted evidence and to provide their views and experiences of the technology and/or condition. Before they attend the meeting, all experts must either submit a written statement (using a template) or indicate they agree with the submission made by their nominating organisation.

**Commentators –** Commentators can participate in the consultation on the ACD (if produced), but NICE does not ask them to make any submission for the appraisal. Non-company commentator organisations can nominate clinical experts and patient experts to verbally present their personal views to the Appraisal Committee. Commentator organisations representing relevant comparator technology companies can also nominate clinical experts. These organisations receive the FAD and have opportunity to report any factual errors. These organisations include comparator technology companies, Healthcare Improvement Scotland any relevant National Collaborating Centre (a group commissioned by NICE to develop clinical guidelines), other related research groups where appropriate (for example, the Medical Research Council and National Cancer Research Institute); other groups such as the NHS Confederation, the NHS Commercial Medicines Unit, the Scottish Medicines Consortium, the Medicines and Healthcare Products Regulatory Agency, the Department of Health, Social Services and Public Safety for Northern Ireland).

**Public –** Members of the public have the opportunity to comment on the ACD when it is posted on the Institute's web site 5 days after it is sent to consultees and commentators. These comments are usually presented to the appraisal committee in full, but NICE reserves the right to summarise and edit comments received during consultations, or not to publish them at all, where in the reasonable opinion of NICE, the comments are voluminous, publication would be unlawful or publication would be otherwise inappropriate.

**Please note:** Comments received in the course of consultations carried out by NICE are published in the interests of openness and transparency, and to promote understanding of how recommendations are developed. The comments are published as a record of the submissions that NICE has received, and are not endorsed by NICE, its officers or advisory committees.

| Comment number | Type of stakeholder | Organisation name | Stakeholder comment<br>Please insert each new comment in a new row | NICE Response<br>Please respond to each comment |
|---|---|---|---|---|
| 1 | Consultee | Servier Laboratories Limited (Servier) | <u>Population for decision making</u><br><br>The ACD states:<br>*"There were 63% of the full intention-to-treat population who had 3 or more previous treatments. The clinical experts expected this to be less than 5% in clinical practice in England."* Section 3.3, page 7<br><br>In the Committee's preferred population for decision making, the committee have selected the European-only no prior ramucirumab population, yet this population still considers a majority of patients with 3 or more prior lines ([redacted]). As discussed previously, an increased number of prior lines is associated with poorer prognosis, and in practice the majority of patients are expected to be treated in the third-line setting (i.e. 2 prior lines).[1]<br><br>To address this limitation of the preferred population for decision making, Servier has conducted subgroup analyses based on the third-line only population (please see separate Appendix containing full results).<br><br>The ACD also states:<br>*"The company preferred to use data from a TAGS subgroup of people from Europe who had not had ramucirumab, because this is more generalisable to the treatment pathway and population in the NHS in England."* Section 3.3, page 7<br><br>This is factually inaccurate – Servier explicitly emphasized caution be exercised when interpreting results from the European-only no prior ramucirumab population, as in addition to not being a pre-specified subgroup analysis, this subgroup only partly addresses some of the limitations of the intention-to-treat (ITT) population. As discussed within the Appraisal Committee meeting, the most appropriate population for decision making would be a predominantly third-line population with no prior ramucirumab exposure, from the UK; however, the no prior ramucirumab subgroup | Comment noted. At the third meeting, the committee concluded that the third-line, European subgroup from TAGS, including the company's adjustment for imbalances in important characteristics, was acceptable for decision-making (see section 3.3 of the Final Appraisal Document, FAD). Therefore, text relating to the subgroup who had not had prior ramucirumab has been removed from section 3.3 of the FAD. |

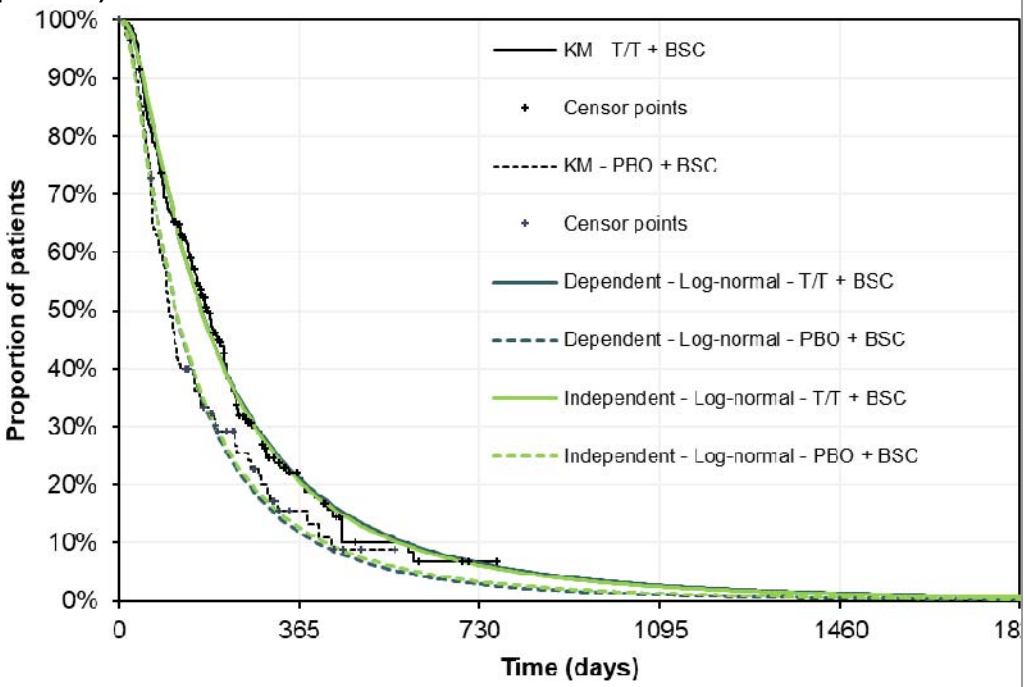
| Comment number | Type of stakeholder | Organisation name | Stakeholder comment<br>Please insert each new comment in a new row | NICE Response<br>Please respond to each comment |
|---|---|---|---|---|
| | | | was selected by Servier to inform its preferred base-case analysis. However, for completeness, Servier has also conducted additional subgroup analysis considering a third-line only population (please see separate Appendix).<br><br>It is relatively uncommon for pivotal trials in the metastatic gastric cancer population to include a largely-European population. NICE has previously assessed two other products in metastatic gastric cancer: trastuzumab (TA208) and ramucirumab (TA378). Trastuzumab was studied in the ToGA trial, in which 55% of patients were from Asian countries.[2] Ramucirumab was studied in the RAINBOW trial, in which 60% of patients were from "Region 1" – defined as Europe, Israel, USA and Australia.[3] Another trial, REGARD, was also considered, and the FAD stated that the Committee was aware that the trial population for REGARD was very similar to that of RAINBOW.[3]<br>• In TA208, the Committee noted that most of the people in the trial were from Asia, but acknowledged subgroup analyses that appeared to confirm a similar overall survival benefit for the group of European people in the trial.[2] As a result of this, the Committee considered it reasonable to make its recommendations on the basis of the ITT population.<br>• In TA378, the Committee was concerned that the drug acquisition costs for ramucirumab and paclitaxel had been underestimated by the company because they were based on the average weight of all people in the RAINBOW trial, about one-third of whom were Asian. In its preferred analysis, Region 1 weight data were used to inform dosing, yet the ITT population was considered to inform estimates of relative efficacy and safety.<br><br>Servier is concerned that the distribution of geographical region for patients in the TAGS trial has been unjustly criticised. Both previous NICE appraisals in metastatic gastric cancer considered studies with much larger Asian populations (and lower European proportions), yet no adjustments to efficacy were made to account for this within the cost-effectiveness analyses conducted. Adjustments were however made for dosing, which were also incorporated into Servier's analysis for the dosing of trifluridine/tipiracil.<br><br>Clinical expert advice at the Appraisal Committee meeting noted that the inclusion of a relatively-small proportion of Asian patients should not impact the results of the clinical trial markedly. However, it should be noted that removal of Asian patients could lead to some imbalances in patient characteristics – for example, consider the proportion of patients with ECOG PS 1:<br>• 'No prior ramucirumab' population: [redacted] (T/T) versus [redacted] (placebo) | |

| Comment number | Type of stakeholder | Organisation name | Stakeholder comment<br>Please insert each new comment in a new row | NICE Response<br>Please respond to each comment |
|---|---|---|---|---|
| | | | • 'European only no prior ramucirumab' population: [redacted] (T/T) versus [redacted] (placebo)<br><br>While only a small change is noted, this illustrates that the removal of non-European patients may lead to biased comparisons between treatment groups. There are also several other imbalances between the treatment arms in the Committee's preferred base-case analysis which are known to be of prognostic importance – for example, HER-2 positivity ([redacted] vs. [redacted]), diffuse histology ([redacted] vs. [redacted]), and peritoneal metastases ([redacted] vs. [redacted]).<br><br>Servier also highlights that a European-only population should not be considered a perfect representation of the UK population. For example, Eastern European countries have notably higher incidence and mortality rates, which may also link to differences in histology.[4,5] A study by Sawaki *et al.* considered an in-depth analysis of the placebo arm of the AVAGAST trial (of first-line treatment of patients with advanced gastric cancer).[6] In this study, European patients were grouped according to whether they resided in the USA/western Europe (n = 81) or in Eastern Europe/South America (n = 118), and showed outcomes for patients in the latter of these groups were poorer (median OS: 7.3 versus 9.1 months).[6]<br><br>Servier notes that while a European-only population is possible to consider within the context of the TAGS trial, this introduces several other limitations and still does not account for patients with more than two prior lines of therapy. In the ERG report, it is stated:<br>"[redacted]".<br><br>It is Servier's opinion that the Committee's chosen subgroup (patients with no prior ramucirumab use, excluding non-European patients) has led to unreasonable interpretations of the clinical- and cost-effectiveness evidence presented. Instead, Servier encourages the Committee to reconsider the population most relevant for decision making, in particular the additional evidence provided concerning the third-line only population which is most closely aligned with NHS practice. | |
| 2 | Consultee | Servier Laboratories Limited (Servier) | <u>End of life</u><br><br>The ACD states:<br>*"The committee agreed that, in line with the NICE methods guide, the* | Comment noted. The end-of-life consideration is based on the committee's preferred population from the third-line, European |

| Comment number | Type of stakeholder | Organisation name | Stakeholder comment<br>Please insert each new comment in a new row | NICE Response<br>Please respond to each comment |
|---|---|---|---|---|
| | | | *criterion requiring a 3-month survival gain should only be relaxed in exceptional circumstances, rather than routinely."* Section 3.8, page 12<br><br>Servier is concerned that this statement implies that trifluridine/tipiracil use within the context of metastatic gastric cancer represents a "routine" circumstance. As communicated within Servier's company submission, the baseline prognosis for patients with heavily pre-treated, metastatic gastric cancer is especially poor, and is expected to be one of the poorest prognoses assessed by NICE in recent history (approximately 6 months or less).<br><br>In Servier's base-case analysis, 65.3% of the incremental LY gain (equivalent to 2.7 months) is accrued within the 'progression-free' health state, with the remaining 34.7% accrued within the 'progressed' health state (total = 2.7 months, 1.8 months = PF). As the majority of the survival gain is achieved while maintaining quality of life, the overall benefit of 2.7 months should be considered as clinically meaningful.<br><br>Based on NICE's end-of-life criteria, an intervention which provides a 3-month survival benefit on a baseline of 2 years is equivalent to a 12.5% increase in survival. Trifluridine/tipiracil for patients with metastatic gastric cancer offers more than three times this relative improvement in Servier's base-case analysis (43.94%). When using the Committee's preferred assumptions, the improvement is still more than double the minimum improvement specified within NICE's end-of-life criteria (25.96%). As shown in the additional evidence provided as an appendix to this response, the additional life-year gain associated with trifluridine/tipiracil is improved when excluding patients with more than three prior lines, in order to better reflect NHS practice.<br><br>In the additional subgroup analyses provided as an Appendix to this response (concerning the third-line only population), the survival benefit attributable to trifluridine/tipiracil is in the region of 3.05-3.21 months (depending on the inclusion of non-European patients). This translates to a relative survival improvement of 47.03-50.23%.<br><br>In NICE TA476, the Committee's decision to appraise nab-paclitaxel under the end-of-life criteria was based on a number of considerations. Importantly however, clinical advice provided to the ERG as part of this appraisal noted that patients recruited to pivotal CA046 trial were younger and fitter than the population of patients with metastatic disease treated in the NHS.[7] More specifically, only 10% of the patients recruited to the trial were aged ≥75 years, whereas CRUK statistics suggest that | subgroup in TAGS. The committee carefully considered the end-of-life criteria. The criteria do not require consideration of proportional gains in life years relative to life expectancy with the comparator. The committee considered the decision made in TA476, but it also noted other technology appraisals where the end-of-life criteria had not been accepted because the extension to life was less than 3 months, even though the short life expectancy criterion had been met. The committee concluded that the extension to life criterion had not been met. Please see section 3.11 of the FAD for more details. |

| Comment number | Type of stakeholder | Organisation name | Stakeholder comment<br>Please insert each new comment in a new row | NICE Response<br>Please respond to each comment |
|---|---|---|---|---|
| | | | almost half (47%) of all patients diagnosed with pancreatic cancer are in this age band. In addition, none of the participating treatment centres were based in the UK, and the majority of patients were in North America (~63% of patients).<br><br>In contrast, the population of the TAGS trial was considered reasonably-well matched to the UK NHS population that would be eligible for trifluridine/tipiracil. Baseline patient characteristics were considered similar to NHS patients, the majority of patients were European, and the trial included patients from [redacted] UK sites. The extension to life estimated in Servier's base-case analysis was 2.7 months using dependent lognormal models, and 2.3 months using independent lognormal models. However, in the additional analysis considering the third-line only population, this improvement exceeds 3 months.<br><br>Servier does not agree with the Committee's decision to base decision making on the European-only no prior ramucirumab population (1.7-month survival gain), as this analysis provides a misleading estimate of the likely survival benefits for patients treated with trifluridine/tipiracil in NHS practice. As shown by the additional analyses conducted by Servier, the survival benefit of trifluridine/tipiracil in a third-line only population is expected to exceed 3 months. An extension in survival of this quantity (the majority of which is accrued within a non-progressed disease state) within the context of a population with extremely poor prognosis who can be treated within the community setting should be considered as a highly valuable end-of-life treatment option. Therefore, Servier requests the Committee reconsider its position concerning whether trifluridine/tipiracil meets the end-of-life criteria. | |
| 3 | Consultee | Servier Laboratories Limited (Servier) | <u>Survival models</u><br><br>The ACD states:<br>*"The company maintained its preference for the dependent model in its base-case analysis but accepted that other approaches may also be valid. The committee concluded that the model should use survival curves fitted independently to each trial arm to extrapolate overall survival."* Section 3.5, page 9<br><br>Servier acknowledges that a range of alternative survival extrapolation approaches are important to consider in order to understand the likely survival gains attributable to treatment with trifluridine/tipiracil. However, by considering an overall model for both treatment arms (i.e. the dependent lognormal model), a more robust estimation of the shape of the curves may be produced, with an estimated treatment effect | Comment noted. The committee preferred independent models for overall survival because in many analyses, the treatment curves crossed or almost converged, therefore the treatment effect could not be assumed to be constant over time. See section 3.6 of the FAD for more details. |

| Comment number | Type of stakeholder | Organisation name | Stakeholder comment<br>Please insert each new comment in a new row | NICE Response<br>Please respond to each comment |
|---|---|---|---|---|
| | | | derived using all patients (rather than being inferred from two separate models). In other words, the dependent model addresses the limited number of patients in the BSC arm in the 'no prior ramucirumab' population (n=115) due to the 2:1 randomisation of the TAGS trial and subgroup analysis chosen to reflect the NHS population.<br><br>In the ERG report, it is stated:<br>*"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* [overall survival]*. If the* [overall survival] *data were associated with a constant* [acceleration factor] *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)."* Section 4.3.4.2, page 81<br><br>Servier agrees that in principle, the dependent and independent models would be identical if the acceleration factor (AF) was constant; but as highlighted in the ERG's report this is difficult to establish with "real" trial data. However, the ERG's logic could plausibly be extended to imply that dependent models should never be fitted – i.e. if the AF is constant, then the independent models should theoretically appear identical to the dependent models (and so the dependent models are redundant); whereas if the AF is not constant, then the dependent model should not be fitted.<br><br>The expectation of constant treatment effect over time (determined via a constant AF in Servier's base case) is not unreasonable in light of the evidence available from the TAGS trial and biological plausibility. Trifluridine/tipiracil delays the time to progression by maintaining patient health-related quality of life. Hence, it is plausible to model the treatment effect as a 'shift' in the survival curve based upon this.<br><br>Servier continues to support the use of the dependent lognormal models in its preferred base-case analysis, as this is aligned with clinical advice provided to Servier, the ERG, and NICE; as well as being statistically the best fit to the data. The fits of the dependent and independent models are similar (as previously described by both Servier and the ERG, and as shown in Figure 1). One of the experts explained that in this group of people with disease that has responded well to previous chemotherapy, survival may be over 6 months. The company base-case analysis (dependent model) estimates survival to be 6.2 months for BSC. The ERG's preferred | |

| Comment number | Type of stakeholder | Organisation name | Stakeholder comment<br>Please insert each new comment in a new row | NICE Response<br>Please respond to each comment |
|---|---|---|---|---|
| | | | analysis (independent models) estimated 6.4 months.<br><br>**Figure 1: Comparison of dependent and independent lognormal models for overall survival (Servier's base-case analysis, all no prior ramucirumab patients)**<br>  <br><br><br>**Key:** BSC, best supportive care; KM, Kaplan-Meier; PBO, placebo; T/T, trifluridine/tipiracil.<br><br>The dependent models make use of the totality of the data from the TAGS trial, and avoid clinically-implausible longer-term extrapolations as noted for the independent generalised gamma models using the Committee's preferred analysis ([redacted]). Servier encourages the Committee to consider analyses using both the dependent and independent modelling approaches to inform its decision making, especially when considering non-pre-specified subgroup analyses (for which data from a relatively small number of placebo patients are available to inform | |

| Comment number | Type of stakeholder | Organisation name | Stakeholder comment<br>Please insert each new comment in a new row | NICE Response<br>Please respond to each comment |
|---|---|---|---|---|
| | | | extrapolations). | |
| 4 | Consultee | Servier Laboratories Limited (Servier) | Impact of disease on patients and carers<br><br>In the ACD, it is stated:<br>*"The committee recalled that trifluridine–tipiracil was clinically effective compared with best supportive care (see section 3.3), but noted that it had not seen evidence of additional benefits that were not captured in the model."* Section 3.12, page 14<br><br>In addition, in Section 3.1 of the ACD, it is mentioned that there was no patient organisation submission for this appraisal (and consequently, it may also be inferred that no patient organisation attended the Appraisal Committee meeting). This is to be expected, given that this patient group is hugely underserved with currently-available medicines, caused in part by numerous previous clinical trials that failed to reach their primary endpoints.[8] In addition, there are a small number of metastatic gastric cancer patients expected to eligible for treatment with trifluridine/tipiracil due to the epidemiology and survivorship of the disease.<br><br>The ACD states that trifluridine/tipiracil is not considered innovative, and that all relevant benefits associated with the drug are adequately captured in the model. While Servier appreciates the remit of NICE to consider the clinical- and cost-effectiveness of interventions within the scope of its reference case, the mode of administration of trifluridine/tipiracil is nevertheless an important consideration for clinicians and patients.<br><br>As highlighted by one of the clinical experts engaged by NICE, trifluridine/tipiracil is a *"very easy to use treatment"*. Patients can be treated with trifluridine/tipiracil at home, avoiding the need for frequent visits to the hospital for administration appointments. This non-traditional mode of administration for chemotherapy can reduce disruption to a patient's home life, especially towards the end of life. This statement is mirrored by the other expert who noted that trifluridine/tipiracil is *"easily administered as an oral reagent with minimal disruption to the patient or any special requirements for the NHS"*.<br><br>The impact of late-line advanced gastric cancer on the families and carers of patients has not been extensively studied, but is nonetheless substantial. However, a study by Bilgin and Gozum found a statistically-significant improvement in the quality of life of carers following an intervention of nursing care, with the most impacted dimensions | Comment noted. The committee agreed that there was an unmet need for third-line treatment options for this population (see FAD section 3.1). However, it concluded that trifluridine–tipiracil did not represent a step-change in gastric cancer treatment, and therefore could not be considered innovative. See section 3.15 of the FAD.<br><br>The committee noted the potential for quality-of-life improvement of carers and families that may result from delayed disease progression with trifluridine-tipiracil. However, it concluded there was little evidence that the net quality of life gain would be significant and therefore carer utility should not be considered in the model. See section 3.10 of the FAD. |

| Comment number | Type of stakeholder | Organisation name | Stakeholder comment<br>Please insert each new comment in a new row | NICE Response<br>Please respond to each comment |
|---|---|---|---|---|
| | | | being “psychological strain”, “disruption in daily life” and “caregiving responsibility”.[9] These impacts were not captured within Servier’s economic analysis, though Servier highlights that the benefits of delaying progression to patient carers and families are important to consider. | |
| 5 | Clinical Expert | Elizabeth Smyth | **Has all of the relevant evidence been taken into account?**<br><br>No, if all of the evidence is taken into account the committee should consider the overall survival benefit in the trial as a whole, rather than cherrypicking an underpowered and potentially biased subgroup of a subgroup for analysis The TAGS trial power calculation was based on a dataset of 500 patients, and while there may be some scientific value in assessing a pre-specified stratification factor subgroup analysis (e.g. no prior ramicurumab or Japan vs. the rest of the world), a European vs other subgroup analyis was not stratified for and may be subject to significant confounding. As these subgroups of subgroups reduce in size the validityof any conclusion based on these premises disipates due to wide confidence margins. In conclusion, I believe the assessment of the committee in this regard to be incorrect, and ignores the true benefit of the drug in the gastroesophageal cancer population. | Comment noted. In its submission, the company’s base case was for people who had not had prior ramucirumab rather than the full intention-to-treat population. At the first meeting, the committee heard that the overall survival benefit for trifluridine–tipiracil was less than 3 months in all subgroups and the full intention-to-treat analysis. At the third meeting the committee concluded that the third-line, European subgroup, including the company’s adjustment for imbalances in important characteristics, was acceptable for decision-making. It noted that the overall survival benefit was less than 3 months, therefore the end-of-life criteria had not been met. Trifluridine–tipiracil could not be recommended for routine use in the NHS because it was not cost-effective. For more details of the committee discussions please see sections 3.3 and 3.4 of the FAD for the population and 3.11 for end-of-life. |
| 6 | Clinical Expert | Elizabeth Smyth | **Are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence?**<br><br>No, because the analyses are based on a limited dataset which is not prespecified. The efficacy of trifluridine tipiracil has been underestimated based on valid evidence from a phase III randomised trial. The committee has not considered that the an | Comment noted. The committee noted the poor prognosis for this group of patients with gastric cancer who have had previous treatments. The end-of-life criteria were discussed at all meetings, but |

| Comment number | Type of stakeholder | Organisation name | Stakeholder comment<br>Please insert each new comment in a new row | NICE Response<br>Please respond to each comment |
|---|---|---|---|---|
| | | | extension of overall survival of 30% with limited toxicity is meaningful in a group of patients with survial <6 months. | the committee noted that the overall survival benefit for trifluridine–tipiracil was less than 3 months in most analyses, including its preferred subgroup for decision making. It also recognised that there were other technology appraisals where the end of life criteria had not been accepted due to a life extension of less than 3 months, even though the short life expectancy criterion had been met. At the third meeting, the committee discussed treatment-related adverse events such as neutropenia which were likely to have a negative effect on the quality of life of people with metastatic gastric cancer. It concluded that t only 1 of the 2 end-of-life criteria had been met. See section 3.11 of the FAD. |
| 7 | Clinical Expert | Elizabeth Smyth | **Are the recommendations sound and a suitable basis for guidance to the NHS?**<br><br>No, gastroeosphageal cancer patients in the NHS are already denied access to ramurcirumab which adds significantly to survival in the second line setting. Now that NICE has declined funding for this trifluridine, fit NHS patients with gastroesophageal adenocarcinoma can expect to live 4-5 months less than patients outside the UK. This is in the context of a disease where the median survival is less than one year. These incremental gains matter. The UK has amongst the poorest outcomes for oesophageal and gastric cancer in Europe and oesophageal cancer has been raised an a concern by the Chief Medical Officer. This funding decision is at odds with government policy which is to improve survival for patients with gastroesophageal cancer. | Comment noted. The committee noted the poor prognosis for this group of patients with metastatic gastric cancer who have had previous treatments (see section 3.1 of the FAD). However, based on the evidence presented to it, trifluridine–tipiracil was not cost-effective and therefore could not be recommended for routine use. |
| | | | | |
| 8 | Clinical Expert | Wasat Mansoor | **Has all of the relevant evidence been taken into account?**<br><br>No, the sub population selected (only the EU population and those that have not had ramucirumab) makes little sense biologically. ramucirumab does NOT affect the | Comment noted. The committee was aware of clinical expert opinion that ramucirumab was unlikely to affect the relative |

| Comment number | Type of stakeholder | Organisation name | Stakeholder comment<br>Please insert each new comment in a new row | NICE Response<br>Please respond to each comment |
|---|---|---|---|---|
| | | | biology of this cancer or how it reacts to treatment beyond its use in any way known. So, excluding these patients just makes translation of the trials results more difficult to translate for the purpose of appraisal. | treatment effect for trifluridine–tipiracil. At the third meeting, it concluded that the third-line, European subgroup, including the company's adjustment for imbalances in important characteristics, was acceptable for decision-making. See section 3.3 of the FAD for more details. |
| 9 | Clinical Expert | Wasat Mansoor | **Are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence?**<br><br>please see below | Please see comment below. |
| 10 | Clinical Expert | Wasat Mansoor | **Are the recommendations sound and a suitable basis for guidance to the NHS?**<br><br>Problem with End of Life Criteria and Gastric cancer. This cancer will always be disadvantaged by this measure. Post first line treatment for the non resectable or metastatic patient, the global research body has trialled many different class of drugs and many of these drugs have shown efficacy signal. However, the biology of this cancer has not been generous in allowing survival advantages beyond 2.5 months. This seems to be regardless of whether the test agent is tested in 2nd line or 3rd line treatment. It is important to note that a 2.1 month survival advantage offered by a drug for gastric cancer over and above what you get with best supportive care (3 months) is a large relative increase in life (40% increase). Therefore, there may be a survival ceiling for this lethal cancer where each regimen gives a 2-2.5 month survival advantage in favour of the test drug . These increments build up, however, through out the cancer patients life time so that together, the survival increments of the collective treatments becomes much more meaningful .<br><br>What does this mean in real terms for the entire cancer pathway:<br>NICE rejected the use of Ramucirumab in 2nd line treatment for this cancer because the median survival advantage achieved at 2nd line was 2.2 months. So, the rest of the world take advantage of this survival advantage because they use ramucirumab. The UK cannot. At third line (or beyond), Lonsurf offers a further 2.1 months median survival advantage. So, the rest of the world takes advantage of this, but the UK cannot. Patients in the rest of the world now have a much better TOTAL median survival advantage of 4.3 months extra in life compared to the UK population. the UK population therefore falls considerably behind the rest of the world in terms of | Comment noted. The committee noted the poor prognosis for this group of patients with gastric cancer who have had previous treatments. The end-of-life criteria were discussed at all meetings, but the committee noted that the overall survival benefit for trifluridine–tipiracil was less than 3 months in most analyses, including its preferred subgroup for decision making. It also recognised that there were other technology appraisals where the end of life criteria had not been accepted due to a life extension of less than 3 months, even though the short life expectancy criterion had been met. It concluded that only 1 of the 2 end-of-life criteria had been met. See section 3.11 of the FAD. |

| Comment number | Type of stakeholder | Organisation name | Stakeholder comment<br>Please insert each new comment in a new row | NICE Response<br>Please respond to each comment |
|---|---|---|---|---|
| | | | survival. | |
| | | | | |
| 11 | Consultee | NHS England | NHS England regards the benefits of chemotherapy with trifluridine/tipiracil in this 3rd line gastric cancer indication as being very small (a 2 month increment in median overall survival) with a large part of this additional survival being spent on chemotherapy with its accompanying and significant side effects. NHS England notes that the incremental life year gain and the incremental QALY gain are very small. | Comment noted. The committee agreed that the survival gain with trifluridine–tipiracil was minimal and that only 1 of the 2 end-of-life criteria had been met. At the third meeting, the committee discussed treatment-related adverse events such as neutropenia which were likely to have a negative effect of the quality of life of people with metastatic gastric cancer. See section 3.11 of the FAD. |
| 12 | Consultee | NHS England | NHS England is wary of the multiple subgroup analyses (eg inclusion of US and European patients and only European patients eg previous treatment with ramucirumab) employed by the company to achieve with greater QALY gains | Comment noted. Sections 3.3 and 3.4 of the FAD include discussion of the limitations of using subgroups. |
| 13 | Consultee | NHS England | NHS England sees no reason for the Appraisal Committee to consider the CDF for this appraisal given the maturity of the trial results. | Comment noted. The committee agreed that trifluridine–tipiracil did not meet the criteria for inclusion in the Cancer Drugs Fund as there was no plausible potential to be cost-effective and the survival data was relatively mature. See FAD section 3.14. |

NICE National Institute for Health and Care Excellence

**Trifluridine–tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies [ID1507]**

**Consultation on the appraisal consultation document – deadline for comments: 5pm** on **Friday 24 January 2020. Email:** TACommC@nice.org.uk / NICE DOCS

Please read the checklist for submitting comments at the end of this form. We cannot accept forms that are not filled in correctly.

The Appraisal Committee is interested in receiving comments on the following:

- has all of the relevant evidence been taken into account?
- are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence?
- are the provisional recommendations sound and a suitable basis for guidance to the NHS?

NICE is committed to promoting equality of opportunity, eliminating unlawful discrimination and fostering good relations between people with particular protected characteristics and others. Please let us know if you think that the preliminary recommendations may need changing in order to meet these aims. In particular, please tell us if the preliminary recommendations:

- could have a different impact on people protected by the equality legislation than on the wider population, for example by making it more difficult in practice for a specific group to access the technology;
- could have any adverse impact on people with a particular disability or disabilities.

Please provide any relevant information or data you have regarding such impacts and how they could be avoided or reduced.

| | |
|---|---|
| **Organisation name – Stakeholder or respondent** (if you are responding as an individual rather than a registered stakeholder please leave blank): | Servier Laboratories Limited (Servier) |
| **Disclosure**<br>Please disclose any past or current, direct or indirect links to, or funding from, the tobacco industry. | Servier does not have any past or current, direct or indirect links to, or funding from, the tobacco industry. |
| **Name of commentator person completing form:** | [redacted] |

| Comment number | Comments<br>Insert each comment in a new row.<br>Do not paste other tables into this table, because your comments could get lost – type directly into this table. |
|---|---|
| 1 | Population for decision making<br><br>The ACD states:<br>*"There were 63% of the full intention-to-treat population who had 3 or more previous treatments. The clinical experts expected this to be less than 5% in clinical practice in England."* Section 3.3, page 7<br><br>In the Committee's preferred population for decision making, the committee have selected the European-only no prior ramucirumab population, yet this population still considers a majority of patients with 3 or more prior lines ([redacted]). As discussed previously, an increased number of prior lines is associated with poorer prognosis, and in practice the majority of patients are expected to be treated in the third-line setting (i.e. 2 prior lines).[1]<br><br>To address this limitation of the preferred population for decision making, Servier has conducted subgroup analyses based on the third-line only population (please see separate Appendix containing full results).<br><br>The ACD also states:<br>*"The company preferred to use data from a TAGS subgroup of people from Europe who had not had ramucirumab, because this is more generalisable to the treatment pathway and population in the NHS in England."* Section 3.3, page 7 |

Please return to: **NICE DOCS**

**Trifluridine–tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies [ID1507]**

**Consultation on the appraisal consultation document – deadline for comments: 5pm** on **Friday 24 January 2020. Email:** TACommC@nice.org.uk / NICE DOCS

| | |
|---|---|
| | This is factually inaccurate – Servier explicitly emphasized caution be exercised when interpreting results from the European-only no prior ramucirumab population, as in addition to not being a pre-specified subgroup analysis, this subgroup only partly addresses some of the limitations of the intention-to-treat (ITT) population. As discussed within the Appraisal Committee meeting, the most appropriate population for decision making would be a predominantly third-line population with no prior ramucirumab exposure, from the UK; however, the no prior ramucirumab subgroup was selected by Servier to inform its preferred base-case analysis. However, for completeness, Servier has also conducted additional subgroup analysis considering a third-line only population (please see separate Appendix).<br><br>It is relatively uncommon for pivotal trials in the metastatic gastric cancer population to include a largely-European population. NICE has previously assessed two other products in metastatic gastric cancer: trastuzumab (TA208) and ramucirumab (TA378). Trastuzumab was studied in the ToGA trial, in which 55% of patients were from Asian countries.[2] Ramucirumab was studied in the RAINBOW trial, in which 60% of patients were from "Region 1" – defined as Europe, Israel, USA and Australia.[3] Another trial, REGARD, was also considered, and the FAD stated that the Committee was aware that the trial population for REGARD was very similar to that of RAINBOW.[3]<br>• In TA208, the Committee noted that most of the people in the trial were from Asia, but acknowledged subgroup analyses that appeared to confirm a similar overall survival benefit for the group of European people in the trial.[2] As a result of this, the Committee considered it reasonable to make its recommendations on the basis of the ITT population.<br>• In TA378, the Committee was concerned that the drug acquisition costs for ramucirumab and paclitaxel had been underestimated by the company because they were based on the average weight of all people in the RAINBOW trial, about one-third of whom were Asian. In its preferred analysis, Region 1 weight data were used to inform dosing, yet the ITT population was considered to inform estimates of relative efficacy and safety.<br><br>Servier is concerned that the distribution of geographical region for patients in the TAGS trial has been unjustly criticised. Both previous NICE appraisals in metastatic gastric cancer considered studies with much larger Asian populations (and lower European proportions), yet no adjustments to efficacy were made to account for this within the cost-effectiveness analyses conducted. Adjustments were however made for dosing, which were also incorporated into Servier's analysis for the dosing of trifluridine/tipiracil.<br><br>Clinical expert advice at the Appraisal Committee meeting noted that the inclusion of a relatively-small proportion of Asian patients should not impact the results of the clinical trial markedly. However, it should be noted that removal of Asian patients could lead to some imbalances in patient characteristics – for example, consider the proportion of patients with ECOG PS 1:<br>• 'No prior ramucirumab' population: [redacted] (T/T) versus [redacted] (placebo)<br>• 'European only no prior ramucirumab' population: [redacted] (T/T) versus [redacted] (placebo)<br><br>While only a small change is noted, this illustrates that the removal of non-European patients may lead to biased comparisons between treatment groups. There are also several other imbalances between the treatment arms in the Committee's preferred base-case analysis which are known to be of prognostic importance – for example, HER-2 positivity ([redacted] vs. [redacted]), diffuse histology ([redacted] vs. [redacted]), and peritoneal metastases ([redacted] vs. [redacted]).<br><br>Servier also highlights that a European-only population should not be considered a perfect representation of the UK population. For example, Eastern European countries have notably higher incidence and mortality rates, which may also link to differences in histology.[4,5] A study by Sawaki *et al.* considered an in-depth analysis of the placebo arm of the AVAGAST trial (of first-line treatment of patients with advanced gastric cancer).[6] In this study, European patients were grouped according to whether they resided in the USA/western Europe (n = 81) or in Eastern |

**Trifluridine–tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies [ID1507]**

**Consultation on the appraisal consultation document – deadline for comments: 5pm** on **Friday 24 January 2020. Email:** TACommC@nice.org.uk / NICE DOCS

| | |
|---|---|
| | Europe/South America (n = 118), and showed outcomes for patients in the latter of these groups were poorer (median OS: 7.3 versus 9.1 months).[6]<br><br>Servier notes that while a European-only population is possible to consider within the context of the TAGS trial, this introduces several other limitations and still does not account for patients with more than two prior lines of therapy. In the ERG report, it is stated: "[redacted]".<br><br>It is Servier's opinion that the Committee's chosen subgroup (patients with no prior ramucirumab use, excluding non-European patients) has led to unreasonable interpretations of the clinical- and cost-effectiveness evidence presented. Instead, Servier encourages the Committee to reconsider the population most relevant for decision making, in particular the additional evidence provided concerning the third-line only population which is most closely aligned with NHS practice. |
| 2 | <u>End of life</u><br><br>The ACD states:<br>*"The committee agreed that, in line with the NICE methods guide, the criterion requiring a 3-month survival gain should only be relaxed in exceptional circumstances, rather than routinely."* Section 3.8, page 12<br><br>Servier is concerned that this statement implies that trifluridine/tipiracil use within the context of metastatic gastric cancer represents a "routine" circumstance. As communicated within Servier's company submission, the baseline prognosis for patients with heavily pre-treated, metastatic gastric cancer is especially poor, and is expected to be one of the poorest prognoses assessed by NICE in recent history (approximately 6 months or less).<br><br>In Servier's base-case analysis, 65.3% of the incremental LY gain (equivalent to 2.7 months) is accrued within the 'progression-free' health state, with the remaining 34.7% accrued within the 'progressed' health state (total = 2.7 months, 1.8 months = PF). As the majority of the survival gain is achieved while maintaining quality of life, the overall benefit of 2.7 months should be considered as clinically meaningful.<br><br>Based on NICE's end-of-life criteria, an intervention which provides a 3-month survival benefit on a baseline of 2 years is equivalent to a 12.5% increase in survival. Trifluridine/tipiracil for patients with metastatic gastric cancer offers more than three times this relative improvement in Servier's base-case analysis (43.94%). When using the Committee's preferred assumptions, the improvement is still more than double the minimum improvement specified within NICE's end-of-life criteria (25.96%). As shown in the additional evidence provided as an appendix to this response, the additional life-year gain associated with trifluridine/tipiracil is improved when excluding patients with more than three prior lines, in order to better reflect NHS practice.<br><br>In the additional subgroup analyses provided as an Appendix to this response (concerning the third-line only population), the survival benefit attributable to trifluridine/tipiracil is in the region of 3.05-3.21 months (depending on the inclusion of non-European patients). This translates to a relative survival improvement of 47.03-50.23%.<br><br>In NICE TA476, the Committee's decision to appraise nab-paclitaxel under the end-of-life criteria was based on a number of considerations. Importantly however, clinical advice provided to the ERG as part of this appraisal noted that patients recruited to pivotal CA046 trial were younger and fitter than the population of patients with metastatic disease treated in the NHS.[7] More specifically, only 10% of the patients recruited to the trial were aged ≥75 years, whereas CRUK statistics suggest that almost half (47%) of all patients diagnosed with pancreatic cancer are in this age band. In addition, none of the participating treatment centres were based in the UK, and the majority of patients were in North America (~63% of patients). |

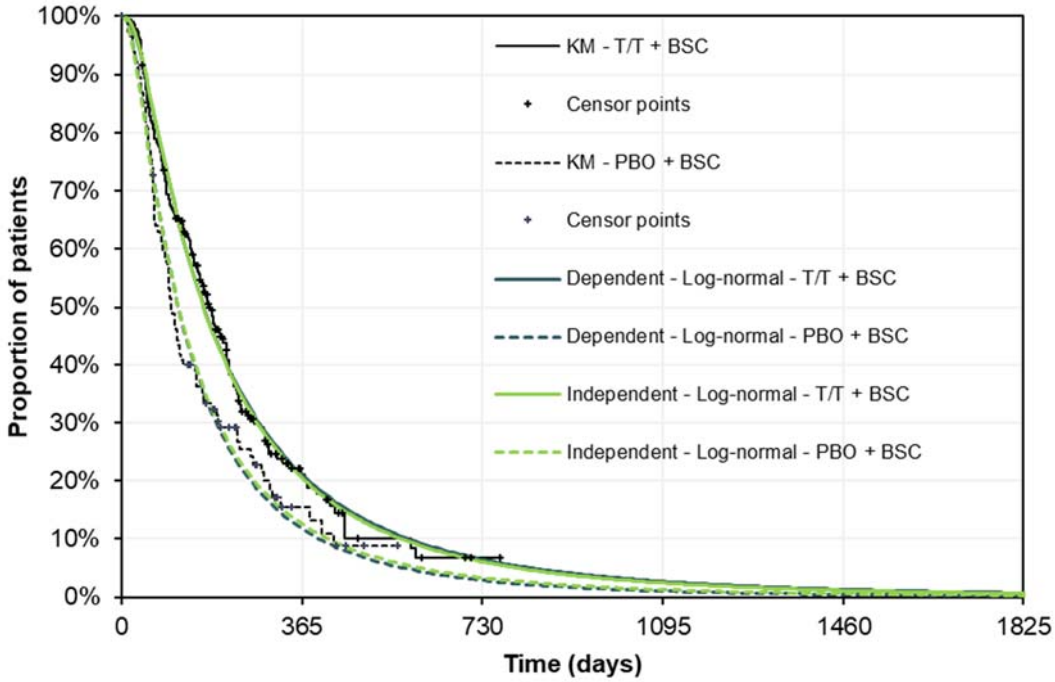
**Trifluridine–tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies [ID1507]**

**Consultation on the appraisal consultation document – deadline for comments: 5pm** on **Friday 24 January 2020. Email:** TACommC@nice.org.uk / NICE DOCS

| | |
|---|---|
| | In contrast, the population of the TAGS trial was considered reasonably-well matched to the UK NHS population that would be eligible for trifluridine/tipiracil. Baseline patient characteristics were considered similar to NHS patients, the majority of patients were European, and the trial included patients from [redacted] UK sites. The extension to life estimated in Servier's base-case analysis was 2.7 months using dependent lognormal models, and 2.3 months using independent lognormal models. However, in the additional analysis considering the third-line only population, this improvement exceeds 3 months.<br><br>Servier does not agree with the Committee's decision to base decision making on the European-only no prior ramucirumab population (1.7-month survival gain), as this analysis provides a misleading estimate of the likely survival benefits for patients treated with trifluridine/tipiracil in NHS practice. As shown by the additional analyses conducted by Servier, the survival benefit of trifluridine/tipiracil in a third-line only population is expected to exceed 3 months. An extension in survival of this quantity (the majority of which is accrued within a non-progressed disease state) within the context of a population with extremely poor prognosis who can be treated within the community setting should be considered as a highly valuable end-of-life treatment option. Therefore, Servier requests the Committee reconsider its position concerning whether trifluridine/tipiracil meets the end-of-life criteria. |
| 3 | Survival models<br><br>The ACD states:<br>*"The company maintained its preference for the dependent model in its base-case analysis but accepted that other approaches may also be valid. The committee concluded that the model should use survival curves fitted independently to each trial arm to extrapolate overall survival."* Section 3.5, page 9<br><br>Servier acknowledges that a range of alternative survival extrapolation approaches are important to consider in order to understand the likely survival gains attributable to treatment with trifluridine/tipiracil. However, by considering an overall model for both treatment arms (i.e. the dependent lognormal model), a more robust estimation of the shape of the curves may be produced, with an estimated treatment effect derived using all patients (rather than being inferred from two separate models). In other words, the dependent model addresses the limited number of patients in the BSC arm in the 'no prior ramucirumab' population (n=115) due to the 2:1 randomisation of the TAGS trial and subgroup analysis chosen to reflect the NHS population.<br><br>In the ERG report, it is stated:<br>*"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* [overall survival]*. If the* [overall survival] *data were associated with a constant* [acceleration factor] *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)."* Section 4.3.4.2, page 81<br><br>Servier agrees that in principle, the dependent and independent models would be identical if the acceleration factor (AF) was constant; but as highlighted in the ERG's report this is difficult to establish with "real" trial data. However, the ERG's logic could plausibly be extended to imply that dependent models should never be fitted – i.e. if the AF is constant, then the independent models should theoretically appear identical to the dependent models (and so the dependent models are redundant); whereas if the AF is not constant, then the dependent model should not be fitted. |

**Trifluridine–tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies [ID1507]**

**Consultation on the appraisal consultation document – deadline for comments: 5pm** on **Friday 24 January 2020. Email:** TACommC@nice.org.uk / NICE DOCS

| | |
|---|---|
| | The expectation of constant treatment effect over time (determined via a constant AF in Servier's base case) is not unreasonable in light of the evidence available from the TAGS trial and biological plausibility. Trifluridine/tipiracil delays the time to progression by maintaining patient health-related quality of life. Hence, it is plausible to model the treatment effect as a 'shift' in the survival curve based upon this.<br><br>Servier continues to support the use of the dependent lognormal models in its preferred base-case analysis, as this is aligned with clinical advice provided to Servier, the ERG, and NICE; as well as being statistically the best fit to the data. The fits of the dependent and independent models are similar (as previously described by both Servier and the ERG, and as shown in Figure 1). One of the experts explained that in this group of people with disease that has responded well to previous chemotherapy, survival may be over 6 months. The company base-case analysis (dependent model) estimates survival to be 6.2 months for BSC. The ERG's preferred analysis (independent models) estimated 6.4 months.<br><br>**Figure 1: Comparison of dependent and independent lognormal models for overall survival (Servier's base-case analysis, all no prior ramucirumab patients)**<br>  <br><br>**Key:** BSC, best supportive care; KM, Kaplan-Meier; PBO, placebo; T/T, trifluridine/tipiracil.<br><br>The dependent models make use of the totality of the data from the TAGS trial, and avoid clinically-implausible longer-term extrapolations as noted for the independent generalised gamma models using the Committee's preferred analysis (). Servier encourages the Committee to consider analyses using both the dependent and independent modelling approaches to inform its decision making, especially when considering non-pre-specified subgroup analyses (for which data from a relatively small number of placebo patients are available to inform extrapolations). |
| | Impact of disease on patients and carers<br><br>In the ACD, it is stated: |

**Trifluridine–tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies [ID1507]**

**Consultation on the appraisal consultation document – deadline for comments: 5pm** on **Friday 24 January 2020. Email:** TACommC@nice.org.uk / NICE DOCS

| | |
|---|---|
| | *"The committee recalled that trifluridine–tipiracil was clinically effective compared with best supportive care (see section 3.3), but noted that it had not seen evidence of additional benefits that were not captured in the model."* Section 3.12, page 14<br><br>In addition, in Section 3.1 of the ACD, it is mentioned that there was no patient organisation submission for this appraisal (and consequently, it may also be inferred that no patient organisation attended the Appraisal Committee meeting). This is to be expected, given that this patient group is hugely underserved with currently-available medicines, caused in part by numerous previous clinical trials that failed to reach their primary endpoints.[8] In addition, there are a small number of metastatic gastric cancer patients expected to eligible for treatment with trifluridine/tipiracil due to the epidemiology and survivorship of the disease.<br><br>The ACD states that trifluridine/tipiracil is not considered innovative, and that all relevant benefits associated with the drug are adequately captured in the model. While Servier appreciates the remit of NICE to consider the clinical- and cost-effectiveness of interventions within the scope of its reference case, the mode of administration of trifluridine/tipiracil is nevertheless an important consideration for clinicians and patients.<br><br>As highlighted by one of the clinical experts engaged by NICE, trifluridine/tipiracil is a *"very easy to use treatment"*. Patients can be treated with trifluridine/tipiracil at home, avoiding the need for frequent visits to the hospital for administration appointments. This non-traditional mode of administration for chemotherapy can reduce disruption to a patient's home life, especially towards the end of life. This statement is mirrored by the other expert who noted that trifluridine/tipiracil is *"easily administered as an oral reagent with minimal disruption to the patient or any special requirements for the NHS"*.<br><br>The impact of late-line advanced gastric cancer on the families and carers of patients has not been extensively studied, but is nonetheless substantial. However, a study by Bilgin and Gozum found a statistically-significant improvement in the quality of life of carers following an intervention of nursing care, with the most impacted dimensions being "psychological strain", "disruption in daily life" and "caregiving responsibility".[9] These impacts were not captured within Servier's economic analysis, though Servier highlights that the benefits of delaying progression to patient carers and families are important to consider. |

**Checklist for submitting comments**

- Use this comment form and submit it as a Word document (not a PDF).
- Complete the disclosure about links with, or funding from, the tobacco industry.
- Combine all comments from your organisation into 1 response. We cannot accept more than 1 set of comments from each organisation.
- Do not paste other tables into this table – type directly into the table.
- Please underline all confidential information, and separately highlight information that is submitted under 'commercial in confidence' in turquoise and all information submitted under 'academic in confidence' in yellow. If confidential information is submitted, please also send a 2nd version of your comment with that information replaced with the following text: 'X'. See the Guide to the processes of technology appraisal (section 3.1.23 to 3.1.29) for more information.
- Do not include medical information about yourself or another person from which you or the person could be identified.
- Do not use abbreviations
- Do not include attachments such as research articles, letters or leaflets. For copyright reasons, we will have to return comments forms that have attachments without reading them. You can resubmit your comments form without attachments, it must send it by the deadline.
- If you have received agreement from NICE to submit additional evidence with your comments on the appraisal consultation document, please submit these separately.

Please return to: **NICE DOCS**

NICE National Institute for Health and Care Excellence

**Trifluridine–tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies [ID1507]**

**Consultation on the appraisal consultation document – deadline for comments: 5pm** on **Friday 24 January 2020. Email:** TACommC@nice.org.uk / NICE DOCS

**Note:** We reserve the right to summarise and edit comments received during consultations, or not to publish them at all, if we consider the comments are too long, or publication would be unlawful or otherwise inappropriate.

Comments received during our consultations are published in the interests of openness and transparency, and to promote understanding of how recommendations are developed. The comments are published as a record of the comments we received, and are not endorsed by NICE, its officers or advisory committees.

Please return to: **NICE DOCS**

**Trifluridine–tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies [ID1507]**

**Consultation on the appraisal consultation document – deadline for comments: 5pm** on **Friday 24 January 2020. Email:** TACommC@nice.org.uk / NICE DOCS

**References**


1. 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. Clin Colorectal Cancer. 2018;17(3):223–30.

2. National Institute for Health and Care Excellence (NICE). Final appraisal determination: Trastuzumab for the treatment of HER2-positive metastatic gastric cancer [Internet]. 2010 Sep [cited 2020 Jan 13]. Available from: https://www.nice.org.uk/guidance/ta208/documents/gastric-cancer-her2positive-metastatic-trastuzumab-final-appraisal-determination3

3. National Institute for Health and Care Excellence (NICE). Final appraisal determination: Ramucirumab for treating advanced gastric cancer or gastro–oesophageal junction adenocarcinoma previously treated with chemotherapy [Internet]. 2015 Nov [cited 2020 Jan 13]. Available from: https://www.nice.org.uk/guidance/ta378/documents/final-appraisal-determination-document

4. Morais S, Ferro A, Bastos A, Castro C, Lunet N, Peleteiro B. Trends in gastric cancer mortality and in the prevalence of Helicobacter pylori infection in Portugal. Eur J Cancer Prev Off J Eur Cancer Prev Organ ECP. 2016;25(4):275–81.

5. European Network of Cancer Registries (ENCR). Stomach cancer (SC) Factsheet [Internet]. 2017 Apr [cited 2020 Jan 15]. Available from: https://www.encr.eu/sites/default/files/factsheets/ENCR_Factsheet_Stomach_2017.pdf

6. Sawaki A, Yamada Y, Yamaguchi K, Nishina T, Doi T, Satoh T, et al. Regional differences in advanced gastric cancer: exploratory analyses of the AVAGAST placebo arm. Gastric Cancer Off J Int Gastric Cancer Assoc Jpn Gastric Cancer Assoc. 2018 May;21(3):429–38.

7. National Institute for Health and Care Excellence (NICE). Single Technology Appraisal: Paclitaxel as albumin-bound nanoparticles with gemcitabine for untreated metastatic pancreatic cancer [ID1058] - Pre-meeting briefing [Internet]. 2017 [cited 2020 Jan 13]. Available from: https://www.nice.org.uk/guidance/ta476/documents/committee-papers

8. Smyth EC, Moehler M. Late-line treatment in metastatic gastric cancer: today and tomorrow. Ther Adv Med Oncol. 2019 Jan 1;11:1758835919867522.

9. Bilgin S, Gozum S. Effect of nursing care given at home on the quality of life of patients with stomach cancer and their family caregivers' nursing care. Eur J Cancer Care (Engl). 2018;27(2):e12567.

# Company response to the Appraisal Consultation Document – Additional analysis

## Introduction

On 30 December 2019, the National Institute for Health and Care Excellence (NICE) published its preliminary guidance concerning the use of trifluridine/tipiracil for patients with metastatic gastric cancer or gastro-oesophageal junction adenocarcinoma after 2 or more therapies. The Appraisal Consultation Document (ACD) states that trifluridine/tipiracil is not recommended, within its marketing authorisation, for treating metastatic gastric cancer or gastro-oesophageal junction adenocarcinoma in adults who have had 2 or more systemic treatment regimens. In describing the reasons behind the Committee's decision, the ACD states:

> *"The clinical evidence suggests that people having trifluridine–tipiracil live longer compared with best supportive care. But the evidence also suggests that it is unlikely to extend how long people live by at least 3 months, particularly in the people who are most relevant to the NHS (the subgroup of people in Europe who have not had ramucirumab). This means trifluridine–tipiracil does not meet NICE's criterion to be considered a life-extending treatment at the end of life."*
> Page 3, Section 1.2

Later in the ACD, the population of most relevance for decision making is discussed. Three patient characteristics are explicitly referred to with respect to the population for use within the model:

- **Geographical region** (i.e. whether patients are from Europe, the United States, or Japan; acknowledging within the pivotal TAGS trial patients were stratified according to whether or not they were from Japan [i.e. Japan versus rest-of-the-world])
- **Prior use of ramucirumab** (a treatment option in the second-line setting, which is not currently recommended by NICE for use in routine National Health Service [NHS] practice, and was also included as a stratification factor within the TAGS trial)
- **Number of previous lines of treatment** (in the TAGs trial all patients had received at least 2 prior lines, but the majority had received 3 or more prior lines. The number of prior lines was not included as a stratification factor within the TAGS trial, yet this is expected to be correlated with both geographical region and prior use of ramucirumab)

In previous communication to NICE, Servier has provided analyses for the following patient populations:

- Intention-to-treat (ITT) population
- Patients with no prior ramucirumab use
- Patients with no prior ramucirumab use, residing in Europe

In Servier's preferred base-case analysis, the no prior ramucirumab subgroup was previously selected. This subgroup was selected based on alignment with the treatment pathway in the UK, and to address some differences between the ITT population and the population expected to be eligible for treatment with trifluridine/tipiracil in NHS practice. While not a perfect representation of the UK population, this subgroup was considered by Servier to be the most appropriate choice of all possible pre-specified subgroup analyses, and is fully aligned with the treatment pathway. Other choices of subgroup analysis would unavoidably introduce the risk of confounding as a trade-off for improving the generalisability of the subgroup to the patient population expected to treated in NHS practice.

The Committee preferred the European-only, no prior ramucirumab population. Servier emphasised caution with respect to the third population when providing this in response to a priority clarification question asked by the Evidence Review Group (ERG), as this comprises a non-pre-specified subgroup

analysis that is subject to a risk of confounding owing to imbalances in patient characteristics. In addition, this subgroup adjusts for geographical region, but does not adjust for the number of prior lines. As found by Davidson *et al.*, an increased number of prior lines of treatment for patients with advanced gastric cancer is associated with poorer prognosis.[*] As such, both Servier's and the Committee's preferred populations are not entirely representative of the population expected to benefit from treatment with trifluridine/tipiracil in NHS practice.

At the Committee meeting, the relevant comparator for this appraisal was confirmed to be best supportive care (BSC). This is because in practice, patients rarely receive chemotherapy in the third-line setting due to a lack of recommended treatment options with evidence of benefit in this setting. As such, the majority of NHS patients would be expected to currently receive BSC as their "third-line" treatment. Given that the majority of patients in the TAGS trial were treated in the fourth-line or beyond setting, associated with poorer prognosis versus third-line, the cost-effectiveness of trifluridine/tipiracil may be under-estimated by both Servier's and the Committee's base-case analyses.

This report details the findings of additional subgroup analyses focusing on the population of patients that were treated in the third-line setting only. These analyses were not previously provided to the Committee as exposure to ramucirumab was intended to serve as a proxy for the number of prior treatment lines. However, as explained above and discussed at the Committee meeting, neither of these subgroups are wholly representative of the population expected to be treated in NHS practice. Consequently, the provision of these analyses is intended to demonstrate the cost-effectiveness of trifluridine/tipiracil in the most relevant population of patients for decision making.

## Methods

In the TAGS trial, a total of n=337 trifluridine/tipiracil patients and n=170 placebo patients were randomised to receive treatment. Of the n=337 trifluridine/tipiracil patients, [redacted] were third-line (i.e. had 2 prior lines), leaving [redacted] patients that were fourth-line or beyond (i.e. 3 or more prior lines), and thus excluded from this analysis. In the third-line population, [redacted] were non-European and so a sensitivity analysis was conducted wherein these patients were excluded. In the placebo arm of the TAGS trial, [redacted] were treated in the third-line setting, and [redacted] third-line patients were non-European. The numbers of patients are summarised in Figure 1.

**Figure 1: Breakdown of patients by treatment line and region**

[redacted]

Based on the breakdown of patients, two subgroups were considered within this analysis:

- **Third-line only:** Patients with two prior lines of treatment, regardless of geographical region
- **Third-line, European only:** Patients with two prior lines of treatment, residing in Europe

[*] Davidson M *et al.* Survival in Advanced Esophagogastric Adenocarcinoma Improves With Use of Multiple Lines of Therapy: Results From an Analysis of More Than 500 Patients. Clin Colorectal Cancer. 2018 Sep;17(3):223-230.

An important difference in these analyses compared with the analyses previously presented by Servier is that patients were not excluded based on prior exposure to ramucirumab. In the third-line analyses, both the number of treatment lines and geographical region were explicitly accounted for when determining eligible patients, and so prior ramucirumab use is no longer necessary to consider as a proxy for these parameters. Clinical expert opinion provided to Servier, the ERG, and the Committee was that prior exposure to ramucirumab is not expected to be a treatment effect modifier or of prognostic importance (in isolation of all other patient characteristics).

To inform the economic model, parametric survival models were fitted for the outcomes of overall survival (OS), progression-free survival (PFS), and time on treatment (ToT). For consistency with the preferences stated by the Committee within the ACD, the following parametric functions were considered:

- **OS:** Independent lognormal models
- **PFS:** Independent generalised gamma models
- **ToT:** Generalised gamma model

The fitted models were incorporated into the cost-effectiveness model previously submitted to NICE, and the impact on the cost-effectiveness results was recorded. All other model settings are aligned with the Committee's preferred assumptions used to inform the base-case ICER presented in the ACD of £68,061. For comparison purposes, the Committee's preferred base-case cost-effectiveness results are presented in Table 1.

**Table 1: Cost-effectiveness results: No prior ramucirumab, European only population (ACD base-case)**

| Arm | Total | | | Incremental | | | ICER |
|---|---|---|---|---|---|---|---|
| | Costs | QALYs | LYs | Costs | QALYs | LYs | |
| BSC | [redacted] | 0.363 | 0.538 | | | | |
| T/T | [redacted] | 0.462 | 0.677 | [redacted] | 0.099 | 0.140 | £68,061 |

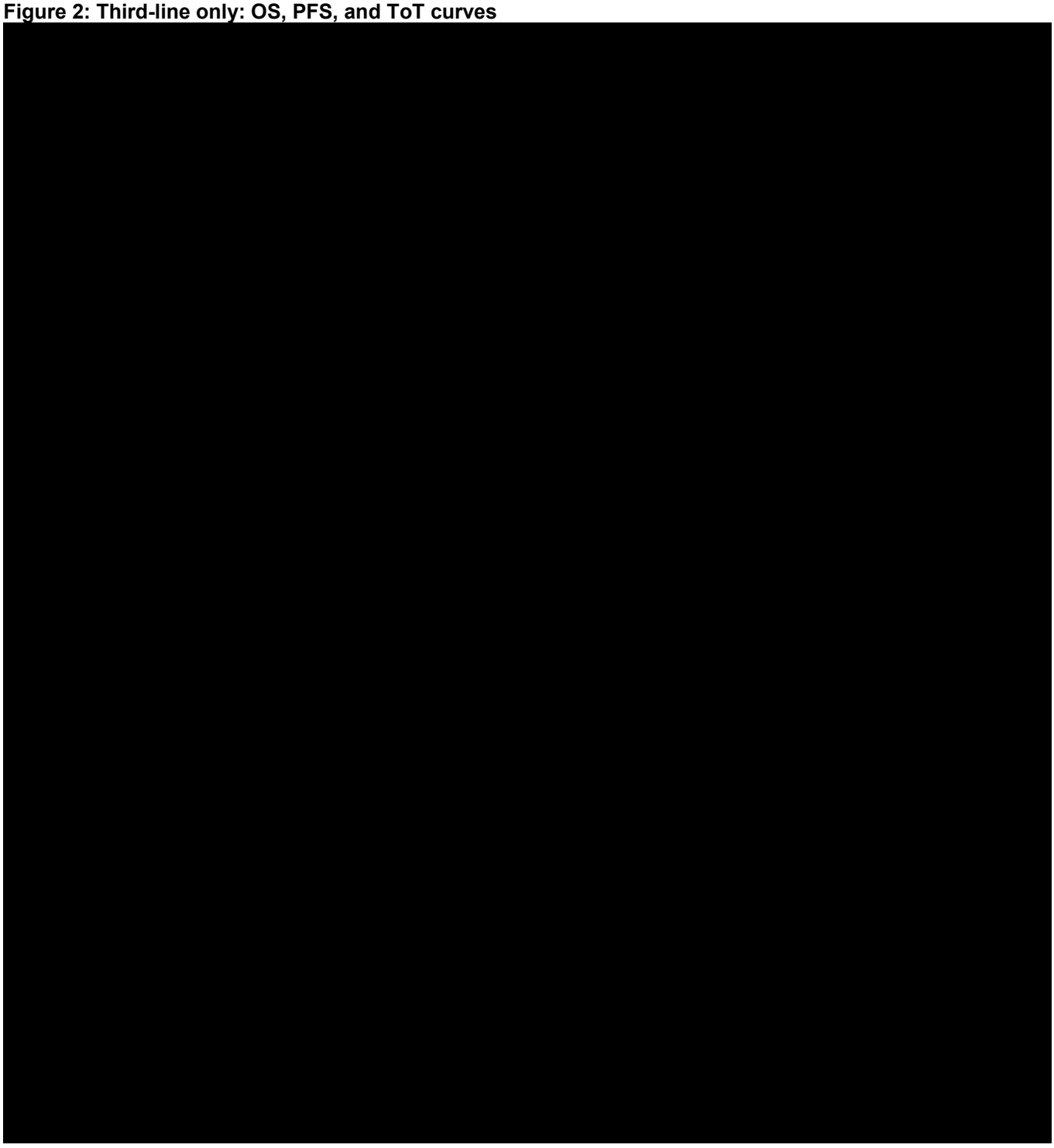
**Key:** ACD, Appraisal Consultation Document; BSC, best supportive care; ICER, incremental cost-effectiveness ratio; LY, life-year; QALY, quality-adjusted life year; T/T, trifluridine/tipiracil.

## Results

### Third-line only

The Kaplan-Meier curves for the outcomes of OS, PFS, and ToT for the third-line only population are presented in Figure 2 alongside the fitted survival models. The parametric functions used (per the Committee's preferred assumptions for the no prior ramucirumab populations) exhibit a good visual fit to the data. However, due to the small number of patients at risk after 1 year for the outcome of OS (approximately [redacted] trifluridine/tipiracil patients and [redacted] placebo patients), the impact of several events on the trifluridine/tipiracil arm causes the curves to cross.

**Figure 2: Third-line only: OS, PFS, and ToT curves**



These curves were then used to inform the cost-effectiveness model, and the impact on results was recorded (presented in Table 2). Through excluding fourth-line and beyond patients, the survival benefit attributable to trifluridine/tipiracil increases to 0.268 life-years (LYs) – equivalent to a benefit of approximately 3.21 months. The improvement in survival translates to an increased quality-adjusted life year (QALY) gain of 0.179, leading to an incremental cost-effectiveness ratio (ICER) of £43,052 (including the [redacted] simple patient access scheme [PAS] discount).

**Table 2: Cost-effectiveness results: Third-line only population**

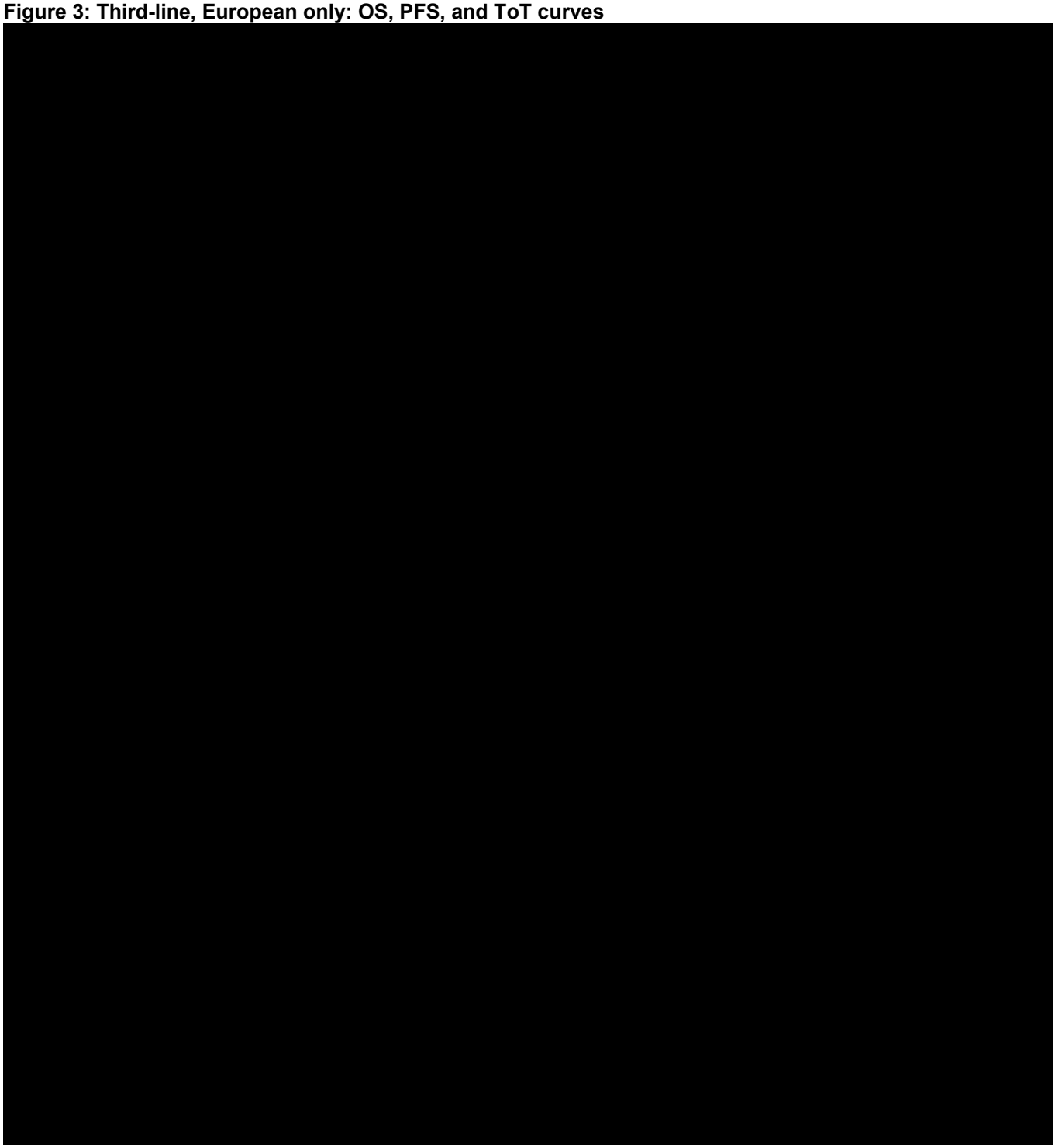
| Arm | Total | | | Incremental | | | ICER |
|---|---|---|---|---|---|---|---|
| | Costs | QALYs | LYs | Costs | QALYs | LYs | |
| BSC | [redacted] | 0.363 | 0.533 | | | | |
| T/T | [redacted] | 0.542 | 0.801 | [redacted] | 0.179 | 0.268 | £43,052 |

**Key:** BSC, best supportive care; ICER, incremental cost-effectiveness ratio; LY, life-year; QALY, quality-adjusted life year; T/T, trifluridine/tipiracil.

### Third-line, European only

The Kaplan-Meier curves for the outcomes of OS, PFS, and ToT for the third-line, European only population are presented in Figure 3 alongside the fitted survival models. As with the third-line only population, due to the small number of patients at risk the OS curves cross after 1 year.

**Figure 3: Third-line, European only: OS, PFS, and ToT curves**



The corresponding cost-effectiveness results are presented in Table 3. The survival benefit associated with trifluridine/tipiracil is similar to that of the third-line only population at 0.255 LYs – equivalent to a benefit of approximately 3.05 months. The improvement in survival translates to a QALY gain of 0.172, leading to an ICER of £46,731 (including the [redacted] simple PAS discount).

**Table 3: Cost-effectiveness results: Third-line, European only population**

| Arm | Total | | | Incremental | | | ICER |
|---|---|---|---|---|---|---|---|
| | Costs | QALYs | LYs | Costs | QALYs | LYs | |
| BSC | [redacted] | 0.369 | 0.541 | | | | |
| T/T | [redacted] | 0.541 | 0.796 | [redacted] | 0.172 | 0.255 | £46,731 |

**Key:** BSC, best supportive care; ICER, incremental cost-effectiveness ratio; LY, life-year; QALY, quality-adjusted life year; T/T, trifluridine/tipiracil.

## Discussion

This analysis presents cost-effectiveness results for the third-line population studied within the pivotal TAGS clinical trial. Aligned with clinical expectation, the results of the analysis demonstrate that the survival benefit associated with trifluridine/tipiracil is greater in a third-line only population versus a population including patients treated at later lines. The results of the analysis were shown to be consistent when including patients from all regions or restricting the analysis to consider only those from Europe (aligned with the Committee's preferred base-case analysis).

When considering the third-line only population, the survival benefit associated with trifluridine/tipiracil is shown to exceed 3 months – the extension in survival normally required in order for a treatment to be considered a life-extending treatment at the end of life. In a population where prognosis is extremely poor, a benefit of 3.05-3.21 months translates to a relative survival improvement of 47.03-50.23%. This larger estimate of increased survival (compared with the Committee's preferred base-case, and the results for the no prior ramucirumab subgroup) is unsurprising, given the improved prognosis of patients at the third-line versus those at the fourth-line and beyond (and thus the increased capacity to derive benefit from trifluridine/tipiracil).

The analysis presented is not without its limitations. As previously described, the TAGS trial recruited patients in the third-line and beyond treatment setting, and the majority of these patients had 3 or more prior lines (i.e. were fourth-line and beyond). Consequently, the sample size available to inform this analysis is smaller than ideal. However, there are still [redacted] patients treated in this setting, [redacted] of which were European. In addition, as per the Committee's preferred analysis, this subgroup was not pre-specified within the TAGS trial protocol, and so the results are unavoidably also at risk of confounding. Nevertheless, the number of treatment lines is associated with all three stratification factors considered within the TAGS trial, mitigating this risk to an extent.

In light of the evidence provided as part of this document, as well as Servier's proforma response to the ACD, Servier kindly requests the Committee to reconsider its positions relating to both the population most suitable for decision making and trifluridine/tipiracil meeting the life extension criterion in order to be considered a life-extending end-of-life treatment. The population of patients treated in the third-line setting is objectively the most relevant to NHS practice, and comprises [redacted] patients enrolled within the TAGS trial. When considering the difference in survival between the treatment arms in the third-line setting, the survival benefit associated with trifluridine/tipiracil exceeds 3 months, illustrating the value of trifluridine/tipiracil as a life-extending treatment option for patients at the end of life.

# Company response to ERG's clarification questions concerning company response to the Appraisal Consultation Document – Additional analysis

1. Please provide baseline characteristics of the two arms (for both 3L full population, and 3L Europe). We know randomisation has been broken, which may or may not cause confounding. Rather than state this should be ok as they are associated with the stratification factors, the company should explicitly show the comparison.

Please note that Servier did not claim the analysis "should be ok" as certain variables are associated with stratification factors, as this question implies. Rather, Servier acknowledged that the number of treatment lines is associated with the stratification factors of the trial which mitigates the risk of confounding to an extent. The explanation provided concerning the risk of confounding is provided in full below for completeness:

> *"The analysis presented is not without its limitations. As previously described, the TAGS trial recruited patients in the third-line and beyond treatment setting, and the majority of these patients had 3 or more prior lines (i.e. were fourth-line and beyond). Consequently, the sample size available to inform this analysis is smaller than ideal. However, there are still* [redacted] *patients treated in this setting,* [redacted] *of which were European. In addition, as per the Committee's preferred analysis, this subgroup was not pre-specified within the TAGS trial protocol, and so the results are unavoidably also at risk of confounding. Nevertheless, the number of treatment lines is associated with all three stratification factors considered within the TAGS trial, mitigating this risk to an extent."*

Please find the requested baseline patient characteristics in Table 1 and Table 2 for the full third-line only and third-line, European only populations, respectively.

## Table 1: Baseline patient characteristics – third-line only population

| | Trifluridine/tipiracil (n=126) | Placebo (n=64) |
|---|---|---|
| **Age (years)**<br>Median (IQR)<br><65<br>≥65 | [redacted] | [redacted] |
| **Sex**<br>Male<br>Female | [redacted] | [redacted] |
| **Ethnicity**<br>White<br>Asian<br>Other<br>Not available | [redacted] | [redacted] |
| **Region**<br>USA<br>Europe*<br>Japan | [redacted] | [redacted] |
| **ECOG performance status**<br>0<br>1 | [redacted] | [redacted] |
| **Primary site**<br>Gastric<br>GEJ<br>Both | [redacted] | [redacted] |
| **Measurable disease** | [redacted] | [redacted] |
| **Histology**<br>Diffused<br>Intestinal<br>Mixed<br>Unknown<br>Not available | [redacted] | [redacted] |
| **HER2 status**<br>Positive<br>Negative<br>Not assessed or unknown | [redacted] | [redacted] |
| **No. of metastatic sites**<br>1–2<br>≥3 | [redacted] | [redacted] |
| **Peritoneal metastases** | [redacted] | [redacted] |
| **Previous gastrectomy** | [redacted] | [redacted] |
| **No. of prior regimens**<br>2<br>3<br>≥4 | [redacted] | [redacted] |
| **Prior systemic cancer therapeutic agents**<br>Platinum<br>Fluoropyrimidine | [redacted] | [redacted] |

| | | |
|---|---|---|
| Taxane<br>Irinotecan<br>Ramucirumab<br>Anti-HER2 therapy†<br>Immunotherapy (anti–PD-1/PD-L1)†<br>Other† | | |

**Key:** 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. †Servier could not identify these values at this time.

**Table 2: Baseline patient characteristics – third-line, European only population**

| | Trifluridine/tipiracil (n=111) | Placebo (n=60) |
|---|---|---|
| **Age (years)**<br>Median (IQR)<br><65<br>≥65 | [redacted] | [redacted] |
| **Sex**<br>Male<br>Female | [redacted] | [redacted] |
| **Ethnicity**<br>White<br>Asian<br>Other<br>Not available | [redacted] | [redacted] |
| **Region**<br>USA<br>Europe*<br>Japan | [redacted] | [redacted] |
| **ECOG performance status**<br>0<br>1 | [redacted] | [redacted] |
| **Primary site**<br>Gastric<br>GEJ<br>Both | [redacted] | [redacted] |
| **Measurable disease** | [redacted] | [redacted] |
| **Histology**<br>Diffused<br>Intestinal<br>Mixed<br>Unknown<br>Not available | [redacted] | [redacted] |
| **HER2 status**<br>Positive<br>Negative<br>Not assessed or unknown | [redacted] | [redacted] |
| **No. of metastatic sites**<br>1–2<br>≥3 | [redacted] | [redacted] |
| **Peritoneal metastases** | [redacted] | [redacted] |
| **Previous gastrectomy** | [redacted] | [redacted] |
| **No. of prior regimens**<br>2<br>3<br>≥4 | [redacted] | [redacted] |
| **Prior systemic cancer therapeutic agents**<br>Platinum<br>Fluoropyrimidine | [redacted] | [redacted] |

| | | |
|---|---|---|
| Taxane<br>Irinotecan<br>Ramucirumab<br>Anti-HER2 therapy†<br>Immunotherapy (anti–PD-1/PD-L1)<br>Other† | | |

**Key:** 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. †Servier could not identify these values at this time.

2. If there is known confounding (although there may be unobserved confounding even if the data look comparable) then state how this can be addressed

Servier acknowledges that this subgroup analysis (as per any other non-pre-specified subgroup analysis) is subject to the risk of confounding. As highlighted by the ERG, even if the groups appear balanced, there may be unobserved confounding, which Servier appreciates is a limitation of the analysis presented.

In response to Question 1, Servier has provided the baseline characteristics of the third-line only and third-line, European only populations. In general, the groups are reasonably well balanced. However, the following characteristics in particular should be noted:

[redacted] Servier understands that in principal, further analysis could be performed to re-weight patients on one treatment arm with the objective of re-balancing patient characteristics. However, such an analysis requires careful consideration of which parameters to place the greatest importance upon. Servier does not consider any of the imbalances highlighted to have a large influence on the overall results, given that any differences in the proportions of patients exhibiting a given characteristic are based on a relatively small sample size (compared with the ITT population).

3. Perform full curve fitting for the new data, the old functions may not be applicable. Provide AIC/CIC and other relevant info

Servier appreciates that a full parametric survival model analysis is important to consider when determining optimal model selection. Alongside provision of this response document, an updated economic model file has been developed to allow the ERG to explore alternative parametric curve fits for the outcomes of OS, PFS, and ToT.

In the interest of providing the updated model to the ERG in a timely manner, Servier has included the full set of parametric survival curves by adapting pre-existing options within the model. To explore the full set of models, please use the following approach:

1. On the 'Controls' tab, please tick the box labelled *"Consider 3L-only population? FULL ANALYSIS"*
2. On the 'Controls' tab, please set both of the ranges *c_lonsurf_data* and *c_bsc_data* to "Prior ramucirumab"
3. On the 'Post-ACD' tab, select whether to consider the third-line only or third-line, European only population by using the drop-down menu in range H22

Servier has left the previous functionality (wherein only one set of curves may be used) intact, should the ERG wish to check the implementation of the full analysis set. Statistical goodness-of-fit scores are provided in Table 3 and Table 4 for the third-line only and third-line, European only populations, respectively. The model also includes the Kaplan-Meier curves for each of the populations.

**Table 3: Statistical goodness-of-fit scores – third-line only population**

| Model | Independent | | | | | | Dependent | |
|---|---|---|---|---|---|---|---|---|
| | TFT | | BSC | | Combined | | | |
| | AIC | BIC | AIC | BIC | AIC | BIC | AIC | BIC |
| **Overall survival** | | | | | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Exponential** | 1,177.85 | 1,180.68 | 658.80 | 660.96 | 1,836.64 | 1,841.64 | 1,836.64 | 1,843.14 |
| **Weibull** | 1,166.29 | 1,171.97 | 660.20 | 664.51 | 1,826.49 | 1,836.48 | 1,827.92 | 1,837.66 |
| **Gompertz** | 1,170.96 | 1,176.64 | 659.75 | 664.06 | 1,830.71 | 1,840.70 | 1,835.29 | 1,845.03 |
| **Log-logistic** | 1,168.85 | 1,174.52 | 649.92 | 654.24 | 1,818.77 | 1,828.76 | 1,817.16 | 1,826.90 |
| **Lognormal** | 1,168.18 | 1,173.85 | 648.73 | 653.05 | 1,816.91 | 1,826.90 | 1,815.15 | 1,824.89 |
| **Gen gam** | 1,167.17 | 1,175.68 | 647.94 | 654.41 | 1,815.11 | 1,830.10 | 1,817.13 | 1,830.12 |
| **Progression-free survival** | | | | | | | | |
| **Exponential** | 1,228.09 | 1,230.93 | 629.57 | 631.73 | 1,857.66 | 1,862.66 | 1,857.66 | 1,864.16 |
| **Weibull** | 1,212.25 | 1,217.92 | 625.50 | 629.82 | 1,837.75 | 1,847.74 | 1,836.44 | 1,846.18 |
| **Gompertz** | 1,224.51 | 1,230.18 | 631.44 | 635.76 | 1,855.95 | 1,865.94 | 1,856.76 | 1,866.50 |
| **Log-logistic** | 1,198.70 | 1,204.37 | 594.75 | 599.07 | 1,793.45 | 1,803.44 | 1,794.53 | 1,804.27 |
| **Lognormal** | 1,194.67 | 1,200.34 | 601.31 | 605.63 | 1,795.98 | 1,805.97 | 1,794.93 | 1,804.67 |
| **Gen gam** | 1,195.66 | 1,204.17 | 597.69 | 604.17 | 1,793.35 | 1,808.34 | 1,792.55 | 1,805.54 |
| **Time on treatment** | | | | | | | | |
| **Exponential** | 1,303.84 | 1,306.67 | | | | | | |
| **Weibull** | 1,302.49 | 1,308.13 | | | | | | |
| **Gompertz** | 1,304.50 | 1,310.14 | | | | | | |
| **Log-logistic** | 1,305.44 | 1,311.08 | | | | | | |
| **Lognormal** | 1,305.11 | 1,310.75 | | | | | | |
| **Gen gam** | 1,302.39 | 1,310.85 | | | | | | |

**Key:** AIC, Akaike information criterion; BIC, Bayesian information criterion; BSC, best supportive care; Gen gam, generalised gamma; TFT, trifluridine/tipiracil.
**Note:** Lowest scores highlighted in grey.

**Table 4: Statistical goodness-of-fit scores – third-line, European only population**

| Model | Independent | | | | | | Dependent | |
|---|---|---|---|---|---|---|---|---|
| | TFT | | BSC | | Combined | | | |
| | AIC | BIC | AIC | BIC | AIC | BIC | AIC | BIC |
| **Overall survival** | | | | | | | | |
| **Exponential** | 1,016.77 | 1,019.48 | 609.87 | 611.97 | 1,626.64 | 1,631.44 | 1,626.64 | 1,632.92 |
| **Weibull** | 1,008.44 | 1,013.86 | 611.69 | 615.88 | 1,620.13 | 1,629.74 | 1,621.58 | 1,631.01 |
| **Gompertz** | 1,012.58 | 1,018.00 | 609.92 | 614.11 | 1,622.50 | 1,632.11 | 1,627.30 | 1,636.72 |
| **Log-logistic** | 1,009.93 | 1,015.35 | 601.20 | 605.39 | 1,611.13 | 1,620.73 | 1,609.55 | 1,618.97 |
| **Lognormal** | 1,009.20 | 1,014.62 | 600.09 | 604.28 | 1,609.29 | 1,618.90 | 1,607.57 | 1,617.00 |
| **Gen gam** | 1,009.10 | 1,017.23 | 598.34 | 604.63 | 1,607.44 | 1,621.85 | 1,609.50 | 1,622.06 |
| **Progression-free survival** | | | | | | | | |
| **Exponential** | 1,071.30 | 1,074.01 | 589.24 | 591.33 | 1,660.54 | 1,665.34 | 1,660.54 | 1,666.82 |
| **Weibull** | 1,059.42 | 1,064.84 | 586.25 | 590.44 | 1,645.67 | 1,655.28 | 1,644.24 | 1,653.66 |
| **Gompertz** | 1,068.65 | 1,074.06 | 591.06 | 595.25 | 1,659.70 | 1,669.31 | 1,660.35 | 1,669.78 |
| **Log-logistic** | 1,050.25 | 1,055.66 | 559.40 | 563.59 | 1,609.64 | 1,619.25 | 1,610.50 | 1,619.93 |
| **Lognormal** | 1,046.49 | 1,051.90 | 564.73 | 568.92 | 1,611.21 | 1,620.82 | 1,610.24 | 1,619.67 |
| **Gen gam** | 1,048.05 | 1,056.18 | 561.74 | 568.03 | 1,609.80 | 1,624.21 | 1,609.26 | 1,621.83 |
| **Time on treatment** | | | | | | | | |
| **Exponential** | 1,142.71 | 1,145.40 | | | | | | |
| **Weibull** | 1,142.26 | 1,147.64 | | | | | | |
| **Gompertz** | 1,143.42 | 1,148.80 | | | | | | |
| **Log-logistic** | 1,146.54 | 1,151.92 | | | | | | |
| **Lognormal** | 1,146.25 | 1,151.64 | | | | | | |
| **Gen gam** | 1,143.19 | 1,151.27 | | | | | | |

**Key:** AIC, Akaike information criterion; BIC, Bayesian information criterion; BSC, best supportive care; Gen gam, generalised gamma; TFT, trifluridine/tipiracil.
**Note:** Lowest scores highlighted in grey.

4. Provide the clinical reason why the gain in health (either LYG or QALY) is much more pronounced for the TFT arm than the BSC arm when moving from the full population to the 3L subgroups. Are the company suggesting that line of therapy is a treatment effect modifier for TFT but not for BSC? If so, provide the biological basis for this assumption.

Servier understands the rationale behind considering the plausibility for differences in results when comparing subgroup analyses. However, it is important to note that comparing subgroup analyses is especially challenging when groups of patients overlap, if there is a risk of confounding, and patient numbers are limited. As such, Servier re-emphasizes caution when interpreting all subgroup analyses (including the third-line only analyses).

Patients with fewer prior lines of therapy are expected to have an improved capacity to derive benefit from treatment with trifluridine/tipiracil. This may be inferred from the forest plot of the pivotal trial publication by Shitara *et al.* (hazard ratio for overall survival = 0.68 for third-line only patients, versus 0.69 for the intention-to-treat population). In addition, clinical advice provided to NICE as part of the technical engagement process noted that patients with a "better reserve" may be expected to derive more benefit from treatment with trifluridine/tipiracil (Clinical expert statement, Dr Mansoor, noting that this point was originally made with respect to the role of prior ramucirumab).

Furthermore, patients treated in an earlier line are expected to have a greater chance of maintaining their health-related quality of life, which would plausibly be at a higher baseline to begin with. Through maintained health-related quality of life (through delayed progression), an improvement in quality-adjusted life-years would be expected.

5. Given the company's preference for the dependent models, can they provide the ICER for this scenario (and others that may be suggested from the new fits to the data)?

Results using a dependent lognormal model for the outcome of overall survival are provided in Table 5 and Table 6 for the third-line only and third-line, European only populations, respectively (with no change to other model parameters).

**Table 5: Cost-effectiveness results: Third-line only (dependent lognormal for OS)**

| Arm | Total | | | Incremental | | | ICER |
|---|---|---|---|---|---|---|---|
| | Costs | QALYs | LYs | Costs | QALYs | LYs | |
| BSC | [redacted] | 0.351 | 0.512 | | | | |
| T/T | [redacted] | 0.555 | 0.822 | [redacted] | 0.204 | 0.309 | £37,907 |

**Key:** ACD, Appraisal Consultation Document; BSC, best supportive care; ICER, incremental cost-effectiveness ratio; LY, life-year; QALY, quality-adjusted life year; T/T, trifluridine/tipiracil.

**Table 6: Cost-effectiveness results: Third-line, European only (dependent lognormal for OS)**

| Arm | Total | | | Incremental | | | ICER |
|---|---|---|---|---|---|---|---|
| | Costs | QALYs | LYs | Costs | QALYs | LYs | |
| BSC | [redacted] | 0.354 | 0.517 | | | | |
| T/T | [redacted] | 0.556 | 0.822 | [redacted] | 0.202 | 0.305 | £39,908 |

**Key:** ACD, Appraisal Consultation Document; BSC, best supportive care; ICER, incremental cost-effectiveness ratio; LY, life-year; QALY, quality-adjusted life year; T/T, trifluridine/tipiracil.

Servier's choice of base-case curves remains unchanged – that is, the use of a dependent lognormal approach for OS, and independent generalised gamma models for PFS and ToT. However, through providing the updated model, the ERG may wish to explore alternative models.

NHS E would like to make the following statements:

1. NHS England regards the benefits of chemotherapy with trifluridine/tipiracil in this 3rd line gastric cancer indication as being very small (a 2 month increment in median overall survival) with a large part of this additional survival being spent on chemotherapy with its accompanying and significant side effects. NHS England notes that the incremental life year gain and the incremental QALY gain are very small.

2. NHS England is wary of the multiple subgroup analyses (eg inclusion of US and European patients and only European patients eg previous treatment with ramucirumab) employed by the company to achieve with greater QALY gains

3. NHS England sees no reason for the Appraisal Committee to consider the CDF for this appraisal given the maturity of the trial results.

NICE National Institute for Health and Care Excellence

**Trifluridine–tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies [ID1507]**

**Consultation on the appraisal consultation document – deadline for comments: 5pm** on **Friday 24 January 2020. Email:** TACommC@nice.org.uk / NICE DOCS

| | |
|---|---|
| | Please read the checklist for submitting comments at the end of this form. We cannot accept forms that are not filled in correctly.<br><br>The Appraisal Committee is interested in receiving comments on the following:<br>• has all of the relevant evidence been taken into account?<br>• are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence?<br>• are the provisional recommendations sound and a suitable basis for guidance to the NHS?<br><br>NICE is committed to promoting equality of opportunity, eliminating unlawful discrimination and fostering good relations between people with particular protected characteristics and others. Please let us know if you think that the preliminary recommendations may need changing in order to meet these aims. In particular, please tell us if the preliminary recommendations:<br>• could have a different impact on people protected by the equality legislation than on the wider population, for example by making it more difficult in practice for a specific group to access the technology;<br>• could have any adverse impact on people with a particular disability or disabilities.<br><br>Please provide any relevant information or data you have regarding such impacts and how they could be avoided or reduced. |
| **Organisation name – Stakeholder or respondent** (if you are responding as an individual rather than a registered stakeholder please leave blank): | |
| **Disclosure** Please disclose any past or current, direct or indirect links to, or funding from, the tobacco industry. | None |
| **Name of commentator person completing form:** | Elizabeth Smyth |

| **Comment number** | **Comments** |
|---|---|

Please return to: **NICE DOCS**

NICE National Institute for Health and Care Excellence

**Trifluridine–tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies [ID1507]**

**Consultation on the appraisal consultation document – deadline for comments: 5pm** on **Friday 24 January 2020. Email:** TACommC@nice.org.uk / NICE DOCS

| | Insert each comment in a new row.<br>Do not paste other tables into this table, because your comments could get lost – type directly into this table. |
|---|---|
| Example 1 | We are concerned that this recommendation may imply that …….. |
| 1 | **Has all of the relevant evidence been taken into account?**<br><br>No, if all of the evidence is taken into account the committee should consider the overall survival benefit in the trial as a whole, rather than cherrypicking an underpowered and potentially biased subgroup of a subgroup for analysis The TAGS trial power calculation was based on a dataset of 500 patients, and while there may be some scientific value in assessing a pre-specified stratification factor subgroup analysis (e.g. no prior ramicurumab or Japan vs. the rest of the world), a European vs other subgroup analyis was not stratified for and may be subject to significant confounding. As these subgroups of subgroups reduce in size the validityof any conclusion based on these premises disipates due to wide confidence margins. In conclusion, I believe the assessment of the committee in this regard to be incorrect, and ignores the true benefit of the drug in the gastroesophageal cancer population. |
| 2 | **Are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence?**<br><br>No, because the analyses are based on a limited dataset which is not prespecified.<br>The efficacy of trifluridine tipiracil has been underestimated based on valid evidence from a phase III randomised trial. The committee has not considered that the an extension of overall survival of 30% with limited toxicity is meaningful in a group of patients with survial <6 months. |
| 3 | **Are the recommendations sound and a suitable basis for guidance to the NHS?**<br><br>No, gastroeosphageal cancer patients in the NHS are already denied access to ramurcirumab which adds significantly to survival in the second line setting. Now that NICE has declined funding for this trifluridine, fit NHS patients with gastroesophageal adenocarcinoma can expect to live 4-5 months less than patients outside the UK. This is in the context of a disease where the median survival is less than one year. These incremental gains matter. The UK has amongst the poorest outcomes for oesophageal and gastric cancer in Europe and oesophageal cancer has been raised an a concern by the Chief Medical Officer. This funding decision is at odds with government policy which is to improve survival for patients with gastroesophageal cancer. |
| 4 | |
| 5 | |
| 6 | |

Insert extra rows as needed

**Checklist for submitting comments**

- Use this comment form and submit it as a Word document (not a PDF).
- Complete the disclosure about links with, or funding from, the tobacco industry.
- Combine all comments from your organisation into 1 response. We cannot accept more than 1 set of comments from each organisation.
- Do not paste other tables into this table – type directly into the table.
- Please underline all confidential information, and separately highlight information that is submitted under 'commercial in confidence' in turquoise and all information submitted under 'academic in confidence' in yellow. If confidential information is submitted, please also send a 2nd version of your comment with that information replaced with the following text: 'academic / commercial in confidence information removed'. See

Please return to: **NICE DOCS**

**Trifluridine–tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies [ID1507]**

**Consultation on the appraisal consultation document – deadline for comments: 5pm** on **Friday 24 January 2020. Email:** TACommC@nice.org.uk / NICE DOCS

the Guide to the processes of technology appraisal (section 3.1.23 to 3.1.29) for more information.

- Do not include medical information about yourself or another person from which you or the person could be identified.
- Do not use abbreviations
- Do not include attachments such as research articles, letters or leaflets. For copyright reasons, we will have to return comments forms that have attachments without reading them. You can resubmit your comments form without attachments, it must send it by the deadline.
- If you have received agreement from NICE to submit additional evidence with your comments on the appraisal consultation document, please submit these separately.

**Note:** We reserve the right to summarise and edit comments received during consultations, or not to publish them at all, if we consider the comments are too long, or publication would be unlawful or otherwise inappropriate.

Comments received during our consultations are published in the interests of openness and transparency, and to promote understanding of how recommendations are developed. The comments are published as a record of the comments we received, and are not endorsed by NICE, its officers or advisory committees.

NICE National Institute for Health and Care Excellence

**Trifluridine–tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies [ID1507]**

**Consultation on the appraisal consultation document – deadline for comments: 5pm** on **Friday 24 January 2020. Email:** TACommC@nice.org.uk / NICE DOCS

| | |
|---|---|
| | Please read the checklist for submitting comments at the end of this form. We cannot accept forms that are not filled in correctly.<br><br>The Appraisal Committee is interested in receiving comments on the following:<br>• has all of the relevant evidence been taken into account?<br>• are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence?<br>• are the provisional recommendations sound and a suitable basis for guidance to the NHS?<br><br>NICE is committed to promoting equality of opportunity, eliminating unlawful discrimination and fostering good relations between people with particular protected characteristics and others. Please let us know if you think that the preliminary recommendations may need changing in order to meet these aims. In particular, please tell us if the preliminary recommendations:<br>• could have a different impact on people protected by the equality legislation than on the wider population, for example by making it more difficult in practice for a specific group to access the technology;<br>• could have any adverse impact on people with a particular disability or disabilities.<br><br>Please provide any relevant information or data you have regarding such impacts and how they could be avoided or reduced. |
| **Organisation name – Stakeholder or respondent** (if you are responding as an individual rather than a registered stakeholder please leave blank): | |
| **Disclosure** Please disclose any past or current, direct or indirect links to, or funding from, the tobacco industry. | None. |
| **Name of commentator person completing form:** | Wasat Mansoor |

| **Comment number** | **Comments** |
|---|---|

Please return to: **NICE DOCS**

NICE National Institute for Health and Care Excellence

**Trifluridine–tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies [ID1507]**

**Consultation on the appraisal consultation document – deadline for comments: 5pm** on **Friday 24 January 2020. Email:** TACommC@nice.org.uk / NICE DOCS

| | Insert each comment in a new row.<br>Do not paste other tables into this table, because your comments could get lost – type directly into this table. |
|---|---|
| Example 1 | We are concerned that this recommendation may imply that …….. |
| 1 | **Has all of the relevant evidence been taken into account?**<br><br>No, the sub population selected (only the EU population and those that have not had ramucirumab) makes little sense biologically. ramucirumab does NOT affect the biology of this cancer or how it reacts to treatment beyond its use in any way known. So, excluding these patients just makes translation of the trials results more difficult to translate for the purpose of appraisal. |
| 2 | **Are the summaries of clinical and cost effectiveness reasonable interpretations of the evidence?**<br><br>please see below |
| 3 | **Are the recommendations sound and a suitable basis for guidance to the NHS?**<br><br>Problem with End of Life Criteria and Gastric cancer. This cancer will always be disadvantaged by this measure. Post first line treatment for the non resectable or metastatic patient, the global research body has trialled many different class of drugs and many of these drugs have shown efficacy signal. However, the biology of this cancer has not been generous in allowing survival advantages beyond 2.5 months. This seems to be regardless of whether the test agent is tested in 2nd line or 3rd line treatment. It is important to note that a 2.1 month survival advantage offered by a drug for gastric cancer over and above what you get with best supportive care (3 months) is a large relative increase in life (40% increase). Therefore, there may be a survival ceiling for this lethal cancer where each regimen gives a 2-2.5 month survival advantage in favour of the test drug . These increments build up, however, through out the cancer patients life time so that together, the survival increments of the collective treatments becomes much more meaningful .<br><br>What does this mean in real terms for the entire cancer pathway:<br>NICE rejected the use of Ramucirumab in 2nd line treatment for this cancer because the median survival advantage achieved at 2nd line was 2.2 months. So, the rest of the world take advantage of this survival advantage because they use ramucirumab. The UK cannot. At third line (or beyond), Lonsurf offers a further 2.1 months median survival advantage. So, the rest of the world takes advantage of this, but the UK cannot. Patients in the rest of the world now have a much better TOTAL median survival advantage of 4.3 months extra in life compared to the UK population. the UK population therefore falls considerably behind the rest of the world in terms of survival. |
| 4 | |
| 5 | |
| 6 | |

Insert extra rows as needed

**Checklist for submitting comments**

- Use this comment form and submit it as a Word document (not a PDF).
- Complete the disclosure about links with, or funding from, the tobacco industry.
- Combine all comments from your organisation into 1 response. We cannot accept more than 1 set of comments from each organisation.
- Do not paste other tables into this table – type directly into the table.
- Please underline all confidential information, and separately highlight information that is submitted under 'commercial in confidence' in turquoise and all information submitted under 'academic in confidence' in yellow. If confidential information is submitted,

Please return to: **NICE DOCS**

**Trifluridine–tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies [ID1507]**

**Consultation on the appraisal consultation document – deadline for comments: 5pm** on **Friday 24 January 2020. Email:** TACommC@nice.org.uk / NICE DOCS

please also send a 2nd version of your comment with that information replaced with the following text: 'academic / commercial in confidence information removed'. See the Guide to the processes of technology appraisal (section 3.1.23 to 3.1.29) for more information.

- Do not include medical information about yourself or another person from which you or the person could be identified.
- Do not use abbreviations
- Do not include attachments such as research articles, letters or leaflets. For copyright reasons, we will have to return comments forms that have attachments without reading them. You can resubmit your comments form without attachments, it must send it by the deadline.
- If you have received agreement from NICE to submit additional evidence with your comments on the appraisal consultation document, please submit these separately.

**Note:** We reserve the right to summarise and edit comments received during consultations, or not to publish them at all, if we consider the comments are too long, or publication would be unlawful or otherwise inappropriate.

Comments received during our consultations are published in the interests of openness and transparency, and to promote understanding of how recommendations are developed. The comments are published as a record of the comments we received, and are not endorsed by NICE, its officers or advisory committees.

**Trifluridine-tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies: A Single Technology Appraisal. The ERG's critique of the company's response to the ACD.**


| | |
|---|---|
| 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<br><br>Sue Harnan, Senior Research Fellow, ScHARR, University of Sheffield, Sheffield, UK<br><br>Andrew Metry, Research Associate, ScHARR, University of Sheffield, Sheffield, UK<br><br>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 (04/02/2020) |



**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 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, Metry A, Sripadam R. Trifluridine-tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies: A Single Technology Appraisal. The ERG's critique of the company's response to the ACD. School of Health and Related Research (ScHARR), 2020.

Contributions of authors

Matt Stevenson and Andrew Metry critiqued the health economic analysis submitted by the company. Sue Harnan summarised and critiqued the clinical effectiveness data reported within the company's submission. Matt Stevenson critiqued the baseline characteristics. Raj Sripadam provided clinical advice to the ERG. All authors were involved in drafting and commenting on the final report.

# 1 Background

NICE appraised trifluridine/tipiracil (TFT) (Lonsurf®) at an appraisal committee on the 5th of December 2019. This resulted in an Appraisal Consultation Document (ACD) which did not recommend the use of TFT for treating metastatic gastric cancer or gastro-oesophageal junction adenocarcinoma in adults who have had 2 or more systemic treatment regimens.[1]

On the 28th January 2020 the Evidence Review Group (ERG) received comments from the company (Servier Laboratories) which responded to NICE's ACD. In this document the ERG attempts to summarise the main points raised by the company and to provide an ERG critique of these issues.

# 2 Key Points raised by the company in its response to the ACD

The first three subsections in this section relate to the comment number within the company response to the ACD with the fourth presenting new analyses that were contained in an appendix. For brevity, the position of the company has been summarised by the ERG.

## 2.1 Amending the patient population within the decision problem

Within the ACD the committee's preferred population is European-only who had not had prior ramucirumab. The company state that as patients in England would have largely only received two prior lines of treatment then the most appropriate population would be that of patients receiving third-line treatment who have not had prior ramucirumab. The company present two new analyses: one concentrating on third-line patients only, and one focussing on third-line patients only who were also from Europe. The company does not present any analysis in the third-line of patients without prior ramucirumab treatment, explaining that "*In the third-line analyses, both the number of treatment lines and geographical region were explicitly accounted for when determining eligible patients, and so prior ramucirumab use is no longer necessary to consider as a proxy for these parameters*." (p3, appendix to the ACD response).[2]

The company also believes that narrowing of the population to European-only is unjust as previous STAs had not done so, and that because Europe was not a stratification factor (the geographical stratification was on Japan versus the Rest of the World) imbalances could occur in important prognostic factors such as HER-2 positivity, diffuse histology, and peritoneal metastases.

## 2.2 End of Life

The NICE appraisal committee decided that the End of Life criteria were not met as the criterion related to achieving a robust three-month extension to life was not met. The company disagree, restating that the life expectancy of the decision population was low and that the relative extension to life was large.

Furthermore, the company state that in the additional analyses presented which focussed only on patients receiving third-line treatment that the modelled extension to life was in excess of three months.

### 2.3 Survival modelling

Within the ACD the NICE appraisal committee state a preference for the fitting of survival curves independently to each trial arm to extrapolate overall survival, a view that was also held by the ERG. The company continues to support the use of dependent models stating that this is statistically the best fit to the data, that this is aligned with '*clinical advice provided to Servier, the ERG and NICE*', that this makes use of the totality of data from the TAGS study,[3] and can avoid clinically implausible extrapolations. The company '*encourages the Committee to consider analyses using both the dependent and independent modelling approaches to inform its decision making, especially when considering non-pre-specified subgroup analyses (for which data from a relatively small number of placebo patients are available to inform extrapolations).*' (p5, ACD response).[2]

### 2.4 New analyses

The company provided new analyses that considered patient populations of i) all third-line patients, and (ii) third-line patients from Europe only. As shown in Figure 1, these analyses result in considerable discarding of data, as the third-line population represents [redacted]% of the full study population, with third-line and European represents [redacted]% of the full study population.

**Figure 1: Breakdown of patients in the TAGS study by treatment line and region**

[redacted]

The company used the same parametric functions preferred by the NICE appraisal committee which were: independent lognormal functions for overall survival, independent generalised gamma functions for progression-free survival, and a generalised gamma function for time on treatment.

For reference, the NICE-preferred analyses are shown in Figure 1.

**Table 1: Cost-effectiveness results: No prior ramucirumab, European only population (ACD base-case)**

| Arm | Total | | | Incremental | | | ICER |
|---|---|---|---|---|---|---|---|
| | Costs | QALYs | LYs | Costs | QALYs | LYs | |
| BSC | [redacted] | 0.363 | 0.538 | | | | |
| TFT | [redacted] | 0.462 | 0.677 | [redacted] | 0.099 | 0.140 | £68,061 |

**Key:** ACD, Appraisal Consultation Document; BSC, best supportive care; ICER, incremental cost-effectiveness ratio; LY, life-year; QALY, quality-adjusted life year; TFT, trifluridine/tipiracil.

The results produced when using independent lognormal functions for modelling OS of the third-line population only are shown in Table 2, and those for the European-only third-line population are shown in Table 3. Table 4 and Table 5 shows results using dependent lognormal functions for the third-line population only and the European-only third-line population respectively. It is seen that the use of dependent functions rather than independent functions lowers the ICER by approximately £6000.

**Table 2: Cost-effectiveness results: Third-line only population (independent lognormal functions for OS)**

| Arm | Total | | | Incremental | | | ICER |
|---|---|---|---|---|---|---|---|
| | Costs | QALYs | LYs | Costs | QALYs | LYs | |
| BSC | [redacted] | 0.363 | 0.533 | | | | |
| TFT | [redacted] | 0.542 | 0.801 | [redacted] | 0.179 | 0.268 | £43,052 |

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

**Table 3: Cost-effectiveness results: Third-line, European-only population (independent lognormal functions for OS)**

| Arm | Total | | | Incremental | | | ICER |
|---|---|---|---|---|---|---|---|
| | Costs | QALYs | LYs | Costs | QALYs | LYs | |
| BSC | [redacted] | 0.369 | 0.541 | | | | |
| TFT | [redacted] | 0.541 | 0.796 | [redacted] | 0.172 | 0.255 | £46,731 |

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

**Table 4: Cost-effectiveness results: Third-line only population (dependent lognormal for OS)**

| Arm | Total | | | Incremental | | | ICER |
|---|---|---|---|---|---|---|---|
| | Costs | QALYs | LYs | Costs | QALYs | LYs | |
| BSC | [redacted] | 0.351 | 0.512 | | | | |
| TFT | [redacted] | 0.555 | 0.822 | [redacted] | 0.204 | 0.309 | £37,907 |

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

**Table 5: Cost-effectiveness results: Third-line, European-only population (dependent lognormal for OS)**

| Arm | Total | | | Incremental | | | ICER |
|---|---|---|---|---|---|---|---|
| | Costs | QALYs | LYs | Costs | QALYs | LYs | |
| BSC | [redacted] | 0.354 | 0.517 | | | | |
| TFT | [redacted] | 0.556 | 0.822 | [redacted] | 0.202 | 0.305 | £39,908 |

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

The company notes that the ICERs for TFT in the two new analyses are improved compared with the NICE-preferred results, and that the life years gained are now in excess of 0.25 (3 months) and that the End of Life criteria should be met. The company acknowledges that the analyses have limitations, with the majority of data discarded, as more patients had three or more prior lines of treatment, but that there were still [redacted] third-line patients, [redacted] of whom were European.

# 3 ERG critique of the company's response to the ACD

The subsections in this section correspond to those in Section 2.

## 3.1 Amending the patient population within the decision problem

The ERG highlights that the company appeared to have a contrary view relating to breaking of randomisation. In the response to the ERG clarification the company "*urges extreme caution when interpreting the results for this subgroup, which is not based on stratification within the TAGS trial (patients were stratified based on region [Japan versus rest-of-the-world], ECOG performance status [0 or 1], and prior ramucirumab [yes or no]).*" The ERG acknowledged this point but felt that this was not likely to be a large influence as Europe accounted for approximately 95% of the rest of the world group. In the company's new analysis approximately [redacted] of patients have been discarded leaving behind a group that was not based on stratification. It is unclear why the company does not also urge extreme caution (or greater than extreme caution) in interpreting the new results.

The ERG notes that the company's rationale for subgroup selection was initially unclear. In the original submission, the rationale for selecting patients without prior ramucirumab treatment was that "*This is relevant for the UK as ramucirumab is not reimbursed for NHS patients*" (p67 of the company submission (CS))[4], that "*this population is more reflective of the UK patient population who would be eligible to be treated with trifluridine/tipiracil should it be recommended."* (p71 of the CS)[4] and that "*This population reflects the UK treatment pathway, and is therefore more representative of patients who would be eligible for trifluridine/tipiracil in UK NHS practice*." (p134 of the CS).[4] The concept that the no prior ramucirumab population was being selected as a proxy for line of therapy and geographical region was only made explicit during the company's fact check of the ERG's report. The company's assertion that the third-line patient subgroup analyses "*were not previously provided to the Committee as exposure to ramucirumab was intended to serve as a proxy for the number of prior treatment lines*" does not seem consistent with the initial vague rationale given. The ERG agrees with the company that "*prior exposure to ramucirumab is not expected to be a treatment effect modifier or of prognostic importance (in isolation of all other patient characteristics)*" (p3 of the appendix to the company's ACD response[2]) and in this respect do not think restricting the patient population to patients without prior raumcirumab treatment is appropriate. However, the ERG highlights that the population deemed most relevant by the NICE appraisal committee, that of European patients without prior ramucirumab has not been modelled by the company.

When comparing the results of the full population analyses and those of third-line only it is seen that there has been an increase in modelled survival for TFT patients, but contrastingly, a minimal modelled increase for those on placebo.

shows how the Kaplan-Meier data change from the whole TAGS study population[3] (light lines) and third-line patients only (dark lines) with the solid lines representing TFT and dotted lines representing placebo patients.

**Figure 2:** [redacted]

It is plausible that the relative increase in TFT survival is due to an imbalance of prognostic factors between the arms and thus the company were asked to provide the baseline characteristics for third-line patients, both the full data set and Europeans-only. These data are replicated in Table 6 and Table 7.

It is noted that there are marked differences in prognostic factors between the two groups. The ERG sought clinical advice relating to potential imbalances, discussed in order of appearance in Table 6. For brevity, only the full third-line population is discussed, values for European-only third-line patients can be sourced from Table 7.

[redacted]

[redacted]

The overall clinical impression was that the two arms were not well-balanced in some aspects, but whilst some imbalances, [redacted], favour TFT, some imbalances, [redacted] could favour placebo. Clinical advice stated that it is difficult to ascertain how these effects combine.

The view of the company who state that "*Servier does not consider any of the imbalances highlighted to have a large influence on the overall results, given that any differences in the proportions of patients exhibiting a given characteristic are based on a relatively small sample size (compared with the ITT population)*." and also that "[redacted] The ERG does not necessarily agree with the first statement made by the company, and believes the second statement to be irrelevant as changing the composition of the placebo population would also be expected to change the observed Kaplan-Meier curve.

Methods exist to address imbalances between arms although the company state that "*further analysis could be performed to re-weight patients on one treatment arm with the objective of re-balancing patient characteristics*" but that this was not undertaken as the imbalances were not considered to have "*a large influence on the overall results*".

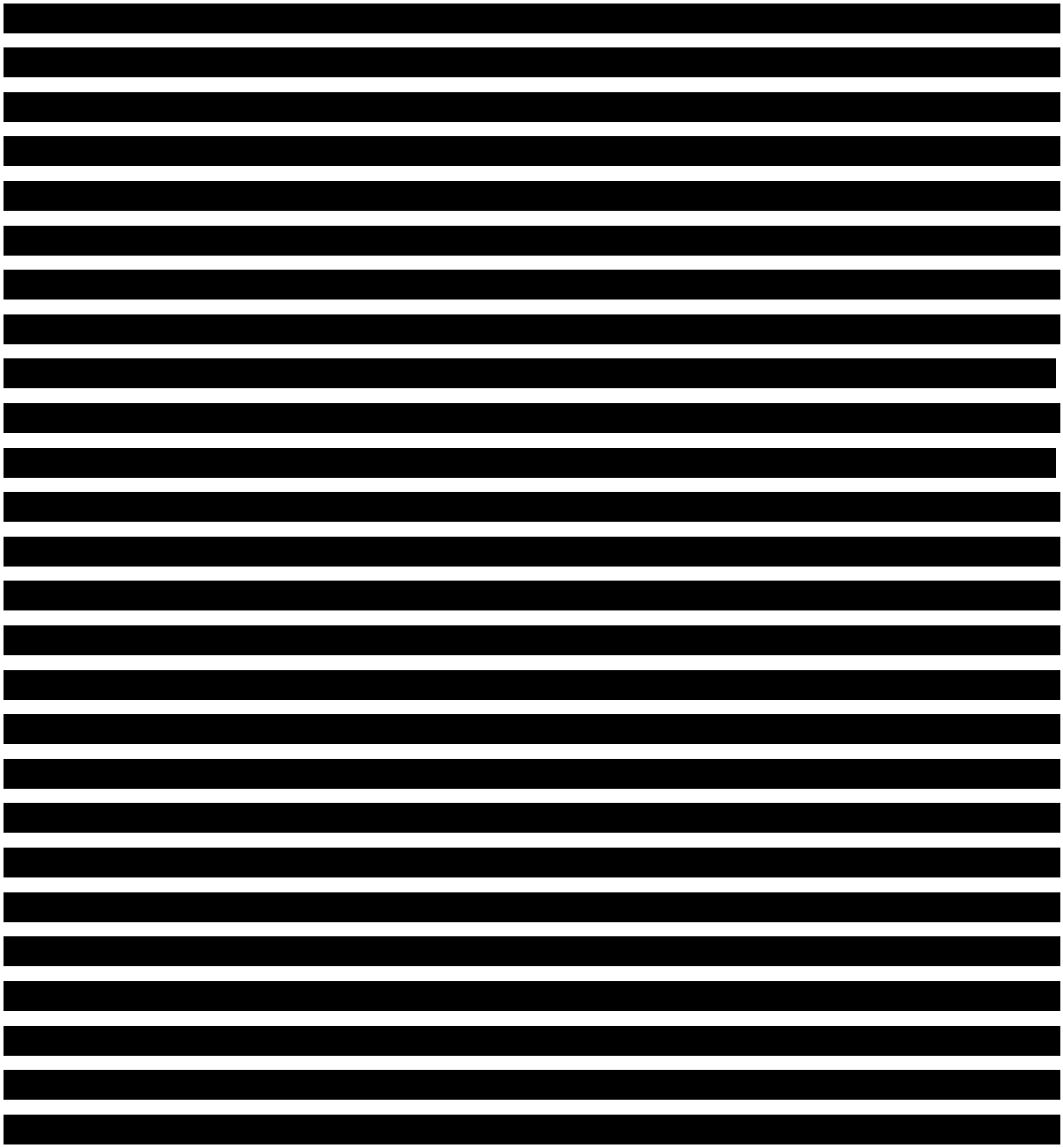
**Table 6: Baseline patient characteristics – third-line only population**

| | | |
|---|---|---|
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

**Table 7: Baseline patient characteristics – third-line, European only population**

| [redacted] | [redacted] | [redacted] |
|---|---|---|
| [redacted] | [redacted] | [redacted] |
| [redacted] | [redacted] | [redacted] |
| [redacted] | [redacted] | [redacted] |
| [redacted] | [redacted] | [redacted] |
| [redacted] | [redacted] | [redacted] |
| [redacted] | [redacted] | [redacted] |
| [redacted] | [redacted] | [redacted] |
| [redacted] | [redacted] | [redacted] |
| [redacted] | [redacted] | [redacted] |
| [redacted] | [redacted] | [redacted] |
| [redacted] | [redacted] | [redacted] |
| [redacted] | [redacted] | [redacted] |
| [redacted] | [redacted] | [redacted] |
| [redacted] | [redacted] | [redacted] |
| [redacted] | [redacted] | [redacted] |
| [redacted] | [redacted] | [redacted] |
| [redacted] | [redacted] | [redacted] |

These comparisons could not fully explain the reason as to why there was a noticeable improvement on survival in the TFT arm, but not within the placebo arm. As such, the ERG has produced a table to provide an insight on how the patient composition changed, for TFT and placebo from the full population to the third-line population only. These data are shown in Table 8. The sources used were the CS (Appendix E)[4] and data provided by the company following the ACD.[2] The ERG has identified changes in the relative balance of variables between the full study population and the third-line only population. These are discussed in the order of appearance in Table 8.



[redacted]

Examining the patient characteristics for placebo it is seen that comparing third-line treatment with the full population there is a [redacted], all of which are associated with a worse prognosis. However, clinical advice to the ERG stated that patients at earlier lines of treatment have a better prognosis. Combining these factors leads to the slight improvement in survival seen in the placebo arm. Contrastingly, the changes between the third-line only population and the full population characteristics for TFT are associated with a better prognosis as whilst there is a [redacted] The ERG believes that the changes in patient composition between the full population and the third-line only population detailed above largely explains the differences in relative survival between TFT and placebo.

**Table 8: Baseline patient characteristics by treatment arm – full study population and third-line only population**

| | TFT | | Placebo | |
|---|---|---|---|---|
| | All lines (n=337) | 3rd line (n=[redacted]) | All lines (n=170) | 3rd line (n=[redacted]) |
| **Age** | | | | |
| <65 | 183 (54%) | [redacted] | 96 (56%) | [redacted] |
| ≥ 65 | 154 (46%) | [redacted] | 74 (44%) | [redacted] |
| **Sex** | | | | |
| Male | 252 (75%) | [redacted] | 117 (69%) | [redacted] |
| Female | 85 (25%) | [redacted] | 53 (31%) | [redacted] |
| **Ethnicity** | | | | |
| White | 244 (72%) | [redacted] | 113 (66%) | [redacted] |
| Asian | 51 (15%) | [redacted] | 29 (17%) | [redacted] |
| Other / Not available | 42 (12%) | [redacted] | 28 (16%) | [redacted] |
| **Region** | | | | |
| USA | 21 (6%) | [redacted] | 5 (3%) | [redacted] |
| Europe | 270 (80%) | [redacted] | 138 (81%) | [redacted] |
| Japan | 46 (14%) | [redacted] | 27 (16%) | [redacted] |
| **ECOG performance status** | | | | |
| 0 | 123 (36%) | [redacted] | 68 (40%) | [redacted] |
| 1 | 214 (64%) | [redacted] | 102 (60%) | [redacted] |
| **Primary Site** | | | | |
| Gastroesophageal | 98 (29%) | [redacted] | 47 (28%) | [redacted] |
| junction | 239 (71%) | [redacted] | 121 (71%) | [redacted] |
| Gastric | 0 (0%) | [redacted] | 2 (1%) | [redacted] |
| Both | | | | |
| **Measurable Disease** | | | | |
| Yes | | | | |
| No | 306 (91%) | [redacted] | 150 (88%) | [redacted] |
| | 31 (9%) | [redacted] | 20 (12%) | [redacted] |
| **Histology** | | | | |
| Diffuse | 53 (16%) | [redacted] | 21 (12%) | [redacted] |
| Intestinal | 103 (31%) | [redacted] | 52 (31%) | [redacted] |
| Mixed | 14 (4%) | [redacted] | 8 (5%) | [redacted] |
| Unknown / Not available | 167 (50%) | [redacted] | 89 (52%) | [redacted] |

| | | | | |
|---|---|---|---|---|
| **HER2 status** | | | | |
| Positive | 67 (20%) | [redacted] | 27 (16%) | [redacted] |
| Negative | 207 (61%) | [redacted] | 106 (62%) | [redacted] |
| Unknown / Not assessed | 63 (19%) | [redacted] | 37 (22%) | [redacted] |
| **No. of metastatic sites** | | | | |
| 1-2 | 155 (46%) | [redacted] | 72 (42%) | [redacted] |
| ≥ 3 | 182 (54%) | [redacted] | 98 (58%) | [redacted] |
| **Peritoneal metastases** | | | | |
| Yes | 87 (26%) | [redacted] | 53 (31%) | [redacted] |
| No | 250 (74%) | [redacted] | 117 (69%) | [redacted] |
| **Previous gastrectomy** | | | | |
| Yes | 147 (44%) | [redacted] | 74 (44%) | [redacted] |
| No | 190 (56%) | [redacted] | 96 (56%) | [redacted] |
| **Number of previous regimens** | | | | |
| 2 | 126 (37%) | [redacted] | 64 (38%) | [redacted] |
| 3 | 134 (40%) | [redacted] | 60 (35%) | [redacted] |
| ≥4 | 77 (23%) | [redacted] | 46 (27%) | [redacted] |
| **Previous ramucirumab** | | | | |
| Yes | 114 (34%) | [redacted] | 55 (32%) | [redacted] |
| No | 223 (66%) | [redacted] | 115 (67%) | [redacted] |
| **Previous irinotecan** | | | | |
| Yes | | | | |
| No | 183 (54%) | [redacted] | 98 (58%) | [redacted] |
| | 154 (46%) | [redacted] | 72 (42%) | [redacted] |
| **Previous taxane** | | | | |
| Yes | 311 (92%) | [redacted] | 148 (87%) | [redacted] |
| No | 26 (8%) | [redacted] | 22 (13%) | [redacted] |

### 3.2 End of Life

As previously, the ERG does not believe it should provide an opinion on whether the NICE committee should allow an intervention to be considered to have met the end of life criteria when the modelled extension to life is below 3 months.

In the new analyses presented by the company the extension to life has been estimated at a value in excess of 3 months (0.25 years) being 0.268 years in the full third-line population and 0.255 years in the European-only third-line population. The ERG comments that based on the descriptions of the limitations, and believed confounding of the data, detailed in Section 3.1 that the extension to life presented in the new analyses is an over-estimate and in the opinion of the ERG it is likely that the true extension to life remains below 3 months.

### 3.3 Survival modelling

Following a request from the ERG the company undertook full curve fitting to the third-line data rather than assume that the previous curves were the most appropriate. Akaike information criterion (AIC) and Bayesian information criterion (BIC) were provided by the company with these values reproduced, with minor corrections, in Table 9 for the full third-line population and in Table 10 for the European-only third-line population.

The ERG notes that for overall survival there is only a slight difference between independent models and dependent models in the combined AIC and BIC values, with the dependent models having a score typically under two points lower than the independent curves. Kass *et al*. state that differences in BIC of two or less is '*Not worth more than a bare mention*'.[5] Without a clearly better statistical fit to the data from the dependent models the ERG maintains that it is preferable not to force an assumption of proportional hazards or acceleration factors and that the independent curves are the most appropriate to use. Furthermore, the ERG refutes the suggestion that clinical advice to it suggested that a dependent model was more appropriate than an independent model as suggested by the company.

Based purely on the data the ERG believes that independent lognormal functions are most suitable for modelling overall survival, that independent log-logistic functions are most suitable for modelling progression-free survival and that an exponential function is most suitable for modelling time on treatment. However, the ERG believes for the reasons stated in Section 3.1 that the data are potentially confounded and favourable to TFT.

**Table 9: Statistical goodness-of-fit scores – third-line only population**

| Model | Independent | | | | | | Dependent | |
|---|---|---|---|---|---|---|---|---|
| | TFT | | Placebo | | Combined | | | |
| | AIC | BIC | AIC | BIC | AIC | BIC | AIC | BIC |
| **Overall survival** | | | | | | | | |
| **Exponential** | 1,177.85 | 1,180.68 | 658.80 | 660.96 | 1,836.64 | 1,841.64 | 1,836.64 | 1,843.14 |
| **Weibull** | 1,166.29 | 1,171.97 | 660.20 | 664.51 | 1,826.49 | 1,836.48 | 1,827.92 | 1,837.66 |
| **Gompertz** | 1,170.96 | 1,176.64 | 659.75 | 664.06 | 1,830.71 | 1,840.70 | 1,835.29 | 1,845.03 |
| **Log-logistic** | 1,168.85 | 1,174.52 | 649.92 | 654.24 | 1,818.77 | 1,828.76 | 1,817.16 | 1,826.90 |
| **Lognormal** | 1,168.18 | 1,173.85 | 648.73 | 653.05 | 1,816.91 | 1,826.90 | 1,815.15 | 1,824.89 |
| **Gen gam** | 1,167.17 | 1,175.68 | 647.94 | 654.41 | 1,815.11 | 1,830.10 | 1,817.13 | 1,830.12 |
| **Progression-free survival** | | | | | | | | |
| **Exponential** | 1,228.09 | 1,230.93 | 629.57 | 631.73 | 1,857.66 | 1,862.66 | 1,857.66 | 1,864.16 |
| **Weibull** | 1,212.25 | 1,217.92 | 625.50 | 629.82 | 1,837.75 | 1,847.74 | 1,836.44 | 1,846.18 |
| **Gompertz** | 1,224.51 | 1,230.18 | 631.44 | 635.76 | 1,855.95 | 1,865.94 | 1,856.76 | 1,866.50 |
| **Log-logistic** | 1,198.70 | 1,204.37 | 594.75 | 599.07 | 1,793.45 | 1,803.44 | 1,794.53 | 1,804.27 |
| **Lognormal** | 1,194.67 | 1,200.34 | 601.31 | 605.63 | 1,795.98 | 1,805.97 | 1,794.93 | 1,804.67 |
| **Gen gam** | 1,195.66 | 1,204.17 | 597.69 | 604.17 | 1,793.35 | 1,808.34 | 1,792.55 | 1,805.54 |
| **Time on treatment** | | | | | | | | |
| **Exponential** | 1,303.84 | 1,306.67 | | | | | | |
| **Weibull** | 1,302.49 | 1,308.13 | | | | | | |
| **Gompertz** | 1,304.50 | 1,310.14 | | | | | | |
| **Log-logistic** | 1,305.44 | 1,311.08 | | | | | | |
| **Lognormal** | 1,305.11 | 1,310.75 | | | | | | |
| **Gen gam** | 1,302.39 | 1,310.85 | | | | | | |

**Key:** AIC, Akaike information criterion; BIC, Bayesian information criterion; Gen gam, generalised gamma; TFT, trifluridine/tipiracil.

**Note:** Lowest scores highlighted in grey.

**Table 10: Statistical goodness-of-fit scores – third-line, European only population**

| Model | Independent | | | | | | Dependent | |
|---|---|---|---|---|---|---|---|---|
| | TFT | | Placebo | | Combined | | | |
| | AIC | BIC | AIC | BIC | AIC | BIC | AIC | BIC |
| **Overall survival** | | | | | | | | |
| **Exponential** | 1,016.77 | 1,019.48 | 609.87 | 611.97 | 1,626.64 | 1,631.44 | 1,626.64 | 1,632.92 |
| **Weibull** | 1,008.44 | 1,013.86 | 611.69 | 615.88 | 1,620.13 | 1,629.74 | 1,621.58 | 1,631.01 |
| **Gompertz** | 1,012.58 | 1,018.00 | 609.92 | 614.11 | 1,622.50 | 1,632.11 | 1,627.30 | 1,636.72 |
| **Log-logistic** | 1,009.93 | 1,015.35 | 601.20 | 605.39 | 1,611.13 | 1,620.73 | 1,609.55 | 1,618.97 |
| **Lognormal** | 1,009.20 | 1,014.62 | 600.09 | 604.28 | 1,609.29 | 1,618.90 | 1,607.57 | 1,617.00 |
| **Gen gam** | 1,009.10 | 1,017.23 | 598.34 | 604.63 | 1,607.44 | 1,621.85 | 1,609.50 | 1,622.06 |
| **Progression-free survival** | | | | | | | | |
| **Exponential** | 1,071.30 | 1,074.01 | 589.24 | 591.33 | 1,660.54 | 1,665.34 | 1,660.54 | 1,666.82 |
| **Weibull** | 1,059.42 | 1,064.84 | 586.25 | 590.44 | 1,645.67 | 1,655.28 | 1,644.24 | 1,653.66 |
| **Gompertz** | 1,068.65 | 1,074.06 | 591.06 | 595.25 | 1,659.70 | 1,669.31 | 1,660.35 | 1,669.78 |
| **Log-logistic** | 1,050.25 | 1,055.66 | 559.40 | 563.59 | 1,609.64 | 1,619.25 | 1,610.50 | 1,619.93 |
| **Lognormal** | 1,046.49 | 1,051.90 | 564.73 | 568.92 | 1,611.21 | 1,620.82 | 1,610.24 | 1,619.67 |
| **Gen gam** | 1,048.05 | 1,056.18 | 561.74 | 568.03 | 1,609.80 | 1,624.21 | 1,609.26 | 1,621.83 |
| **Time on treatment** | | | | | | | | |
| **Exponential** | 1,142.71 | 1,145.40 | | | | | | |
| **Weibull** | 1,142.26 | 1,147.64 | | | | | | |
| **Gompertz** | 1,143.42 | 1,148.80 | | | | | | |
| **Log-logistic** | 1,146.54 | 1,151.92 | | | | | | |
| **Lognormal** | 1,146.25 | 1,151.64 | | | | | | |
| **Gen gam** | 1,143.19 | 1,151.27 | | | | | | |

**Key:** AIC, Akaike information criterion; BIC, Bayesian information criterion; Gen gam, generalised gamma; TFT, trifluridine/tipiracil.
**Note:** Lowest scores highlighted in grey.

### 3.3 ERG exploratory analyses

The ERG can confirm that the results presented by the company in the new analyses match those withint he model. Analyses using the survival functions reported at the end of Section 3.2 result in an ICER of £42,283 per QALY gained which is similar to the company's base case estimate. However, as stated, the data are potentially confounded with the difference in the percentage of patients between the placebo group and the TFT group increasing when the analysis is restricted solely to third-line. Additionally, patients in the placebo group have a [redacted] in third-line compared with the TFT group, a feature that was reversed in the full population. Further, there was a [redacted] in the placebo group in the full population but a lower percentage in third-line alone.

The ERG comments that restriction of the data to third-line patients only has broken randomisation and has led to imbalances in characteristics which may be prognostic of outcome or treatment effect modifiers. Given the potential for confounding both in observed and unobserved variables, the larger numbers of patients in the full population compared with third-line only (507 vs [redacted] respectively) and the fact that median survivals did not have a trend across different prior number of therapies, the ERG prefers the full European population data when estimating the clinical and cost-effectiveness of TFT. Whilst the ERG does not have the data to produced an estimate of the ICER, it is anticipated that this would be in excess of £68,000 (ERG report p 85).[6] The impact of adjusting for these factors is unknown and the ERG would caution against assuming that the impact would not affect outcomes.

As a separate point, the ERG notes that prescribing patterns can be distorted when interventions have been recommended only at one line in the treatment pathway. Recent discussions held by an ERG member in relation to an ongoing STA in multiple myeloma have indicated that clinically preferred options may be withheld until later lines in order to increase the number of lines of treatment, due to prevalent NICE guidance recommending some drugs only at specified lines. For example, daratumumab is only recommended as a fourth-line treatment therapy within the Cancer Drugs Fund whereas pomalidomide and dexamethasone is recommended at fourth and later lines. Clinical advisors to the ERG in the multiple myeloma STA, have stated there are instances when clinically they would prefer to use pomalidomide and dexamethasone at fourth-line, but use daratumumab with pomalidomide and dexamethasone used as fifth-line in order to allow the patient to have two treatment options rather than one if pomalidomide and dexamethasone was used at fourth-line. Therefore, it is possible that the recommendation of TFT only at third-line could cause similar distortions in the future should newer drugs for gastric or gastro-oesophageal junction cancer emerge.

# 4 REFERENCES


1. National Institute for Health and Care Excellence. Appraisal consultation document. Trifluridine–tipiracil for treating metastatic gastric cancer or gastro-oesophageal junction adenocarcinoma after 2 or more therapies. 2019.
2. Servier Laboraties. Trifluridine–tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies [ID1507]. Clarification Response. 2019.
3. 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.
4. Servier Laboraties. Trifluridine–tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies [ID1507]. Company evidence submission. 2019.
5. Kass RE, Raftery AE. "Bayes Factors". *Journal of the American Statistical Association* 1995;90:773-95.
6. Stevenson M, Harnan S, Ren S, Metry A, Orr M, Campbell F*, et al.* 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.

# Additional weighting analysis based on the third-line population enrolled within the TAGS clinical trial

## Contents

Summary results table ..... 1
1. Introduction ..... 2
2. Methods ..... 2
3. Results ..... 4
3.1. 3L only ..... 4
3.2. 3L EU only ..... 10
4. Discussion ..... 15
Appendix: Weighting by ethnicity (instead of geographical region) ..... 16

Figure 1: Histogram of propensity score values – matching on region ..... 4
Figure 2: Absolute difference between arms per covariate – matching on region ..... 5
Figure 3: Kaplan-Meier curves, naïve versus weighted – matching on region ..... 5
Figure 4: Parametric survival models – matching on region ..... 8
Figure 5: Histogram of propensity scores – 3L EU matching ..... 10
Figure 6: Absolute difference between arms per covariate – 3L EU matching ..... 10
Figure 7: Kaplan-Meier curves, naïve versus weighted – 3L EU matching ..... 11
Figure 8: Parametric survival models – 3L EU matching ..... 13
Figure 9: Histogram of propensity score values – matching on ethnicity ..... 16
Figure 10: Absolute difference between arms per covariate – matching on ethnicity ..... 16
Figure 11: Kaplan-Meier curves, naïve versus weighted – matching on ethnicity ..... 17
Figure 12: Parametric survival models – matching on ethnicity ..... 19

Table 1: Summary of cost-effectiveness results ..... 1
Table 2: Statistical goodness-of-fit scores – matching on region ..... 6
Table 3: Cost-effectiveness results: 3L only – matching on region (independent lognormal) ..... 9
Table 4: Cost-effectiveness results: 3L only – matching on region (dependent lognormal) ..... 9
Table 5: Statistical goodness-of-fit scores – 3L EU matching ..... 12
Table 6: Cost-effectiveness results: 3L EU only (independent lognormal) ..... 14
Table 7: Cost-effectiveness results: 3L EU only (dependent lognormal) ..... 14
Table 8: Statistical goodness-of-fit scores – matching on ethnicity ..... 18
Table 9: Cost-effectiveness results: 3L only – matching on ethnicity (independent lognormal) ..... 20
Table 10: Cost-effectiveness results: 3L only – matching on ethnicity (dependent lognormal) ..... 20

## Summary results table

**Table 1: Summary of cost-effectiveness results**

| Analysis description | ICER | Absolute survival gain | % survival gain | Source |
|---|---|---|---|---|
| ACD base-case – ind lognormal | £66,523 | 1.68 months | 25.97% | ACD |
| Naïve 3L – ind lognormal | £42,086 | 3.21 months | 50.23% | Response to ACD |
| Naïve 3L – dep lognormal | £37,058 | 3.71 months | 60.37% | |
| Naïve 3L EU – ind lognormal | £45,681 | 3.05 months | 47.03% | |
| Naïve 3L EU – dep lognormal | £39,013 | 3.66 months | 58.97% | |
| Weighted 3L (region) – ind lognormal | £45,662 | 2.89 months | 44.46% | Section 3.1 |
| Weighted 3L (region) – dep lognormal | £39,925 | 3.37 months | 53.29% | |
| Weighted 3L EU – ind lognormal | £49,771 | 2.72 months | 41.41% | Section 3.2 |
| Weighted 3L EU – dep lognormal | £41,858 | 3.33 months | 52.37% | |
| Weighted 3L (ethnicity) – ind lognormal | £45,611 | 2.89 months | 44.51% | Appendix |
| Weighted 3L (ethnicity) – dep lognormal | £40,724 | 3.29 months | 51.85% | |

**Key:** 3L, third-line; ACD, Appraisal Consultation Document; dep, dependent; EU, Europe; ICER, incremental cost-effectiveness ratio; ind, independent.

## 1. Introduction

On 3 March 2020, a teleconference was held between the National Institute for Health and Care Excellence (NICE) and the company (Servier) concerning the use of trifluridine/tipiracil (T/T) for patients with metastatic gastric cancer or gastro-oesophageal junction adenocarcinoma after 2 or more therapies. During this call, NICE invited the company to provide additional analysis concerning the third-line only population enrolled within the TAGS trial (i.e. patients with only 2 prior therapies), accounting for any potential imbalances in patient characteristics.

As part of this teleconference, NICE highlighted five potentially-important characteristics that would need to be acknowledged as part of the weighting analysis. These were:

1. **Peritoneal metastases:** Patients with an absence or presence of peritoneal metastases (also known as peritoneal involvement)
2. **Eastern Cooperative Oncology Group Performance Status (ECOG PS):** Patients with an ECOG PS of 0 versus 1
3. **Histology:** Patients with intestinal* versus non-intestinal histology
4. **Ethnicity or Region:** Patients residing in Japan or the rest of the world ("region") or patients who are Asian versus non-Asian ("ethnicity")
5. **Prior irinotecan:** Patients with previous exposure to irinotecan versus no previous exposure to irinotecan

This report details the findings of additional analyses focusing on the population of patients that were treated in the third-line setting only, accounting for potential imbalances in the five aforementioned characteristics. The provision of these analyses is intended to demonstrate the cost-effectiveness of T/T in the most relevant population of patients for decision making, and highlight the difference in results when matching on potentially-important variables versus the previously-provided naïve (i.e. unadjusted) analysis (response to ACD). Two sub-populations of patients are considered within this analysis: (1) all third-line patients ("3L only"), and (2) third-line patients residing in Europe ("3L EU only").

[redacted]

## 2. Methods

In order to address the potential bias in the estimation of treatment effects that can arise with imbalances in baseline patient characteristics, statistical methods may be used to obtain a balanced dataset with comparable groups. One possible method that may be used to achieve this is through the use of a propensity score analysis.

Let $Y$ denote treatment assignment, where $Y = 1$ corresponds to patients receiving T/T, and $Y = 0$ corresponds to patients receiving placebo. The propensity score ($PS$) is the probability to receive T/T given a set of n observed baseline covariates $X = (X_1, \dots, X_n)$.

$$PS = P(Y = 1 \,/\, X)$$

This probability can be estimated via a logistic regression model where the dependent variable is the allocated treatment $Y$ and the independent variables are the baseline covariates under consideration $= (X_1, \dots, X_n)$ . The logistic regression model is specified as:

* *Ideally, matching would have been performed based on each of the histological categories: Diffuse, Intestinal, and Mixed. However, matching on each of the three categories was not considered to be possible to robustly perform, within the context of the sample size avaialble for the third-line only population. Therefore, matching was based on intestinal versus non-intestinal.*

$$logit(PS) = \ln\left(\frac{PS}{1-PS}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n$$

… where $\beta_0, \ldots, \beta_n$ are the parameters of the regression model.

The estimation of the $PS$ for the $i^{th}$ patient is therefore:

$$\widehat{PS_i} = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_n x_n}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_n x_n}}$$

… where $x_1, \ldots, x_n$ are the covariate observations for the $i^{th}$ patient, $\hat{\beta}_0, \ldots, \hat{\beta}_n$ are the coefficients obtained from maximum likelihood estimation, and $e$ denotes the exponential constant.

Once the $PS$ is estimated for each patient, it can be used to create comparable groups. With the inverse probability of treatment weighting (IPTW) method, each patient is assigned a weight (a general weight) such that for each combination of selected baseline characteristics, which corresponds to a certain $PS$ value, the sums of contributions of all treated and control patients are equal, resulting in a balanced dataset between treatment groups.

In particular, a given $PS$ leads to a $PS^{-1}$ weight for treated (i.e. T/T) patients and a $(1-PS)^{-1}$ weight for control (i.e. placebo) patients. Thus, a direct consequence of the IPTW implementation is the creation of a pseudo-population in which each combination of covariates results in a more balanced comparison between treatment and control. In this analysis, general weights are used (as opposed to stabilized weights).

IPTW aims at giving more importance (i.e. more "weight") to those patients that have unexpected $PS$ values:

- **T/T patients with low $PS$:** given their covariates, these patients should be part of the placebo arm and are therefore assigned a higher weight;
- **Placebo patients with high $PS$:** given their characteristics, they should have received the T/T and are also therefore assigned a higher weight.

In practice, patients with unexpected $PS$ values are counted more than once in the pseudo-population. The analysis was carried out with SAS, using the PROC LOGISTIC to derive the propensity scores.

In the TAGS trial, a total of n=337 T/T patients and n=170 placebo patients were randomised to receive treatment. Of the n=507 intention-to-treat population, [redacted] patients were third-line (i.e. had 2 prior lines) and therefore comprised the "3L only" group. In the third-line population, [redacted] patients were non-European and so a sensitivity analysis was conducted wherein these patients were excluded (the "3L EU only" population).

The following variables were considered in each of the $PS$ analyses: peritoneal metastases, ECOG PS, histology, region/ethnicity, and prior irinotecan. For the "3L only" population, two different sets of $PS$ values could be produced using this list of variables – i.e. one set based on geographical region, versus another set based on ethnicity. Both variables were captured within the TAGS clinical trial, though geographical region is considered to be the most appropriate variable for inclusion within weighting analysis. This is because a combination of the location in which treatment is provided (and the corresponding health care system) and differences in lifestyles across regions (e.g. local diets) is expected to explain differences in prognosis, as opposed to biological differences in patients according to their racial characteristics (which could be inferred via their self-reported ethnicity). Results for both sets of weights are reported, but for brevity the weights using geographical region are considered the 'main analysis' for the "3L only" population. For the 3L EU only population, one set of weights was derived using the remaining four variables.

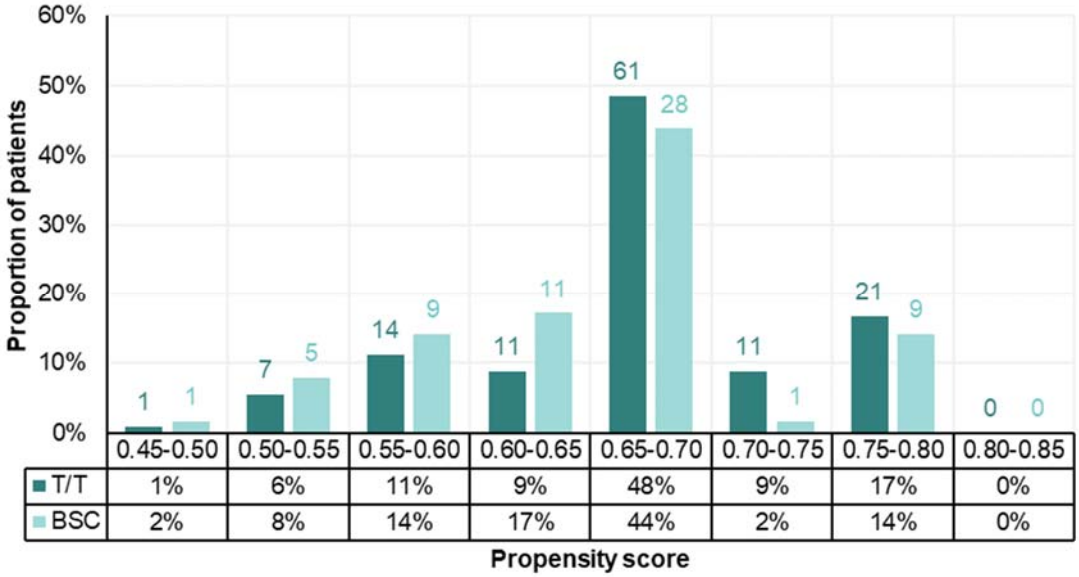
The obtained weights for each patient (based on the $PS$) were then used to inform the re-estimation of parametric survival models for overall survival (OS), progression-free survival (PFS), and time-to-treatment-discontinuation (TTD). Parametric survival models were fitted in the statistical software *R*, using the 'flexsurv' package. The fitted models were incorporated into the cost-effectiveness model previously submitted to NICE, and the impact on the cost-effectiveness results was recorded. All other model settings are aligned with the Committee's preferred assumptions used to inform the base-case ICER presented in the ACD of £66,523.

## 3. Results

### 3.1. 3L only

A histogram demonstrating the distribution of $PS$ values across both treatment arms is presented in Figure 1. The distribution of $PS$ values is reasonably consistent across both arms, with a peak observed in the interval of 0.65-0.70. The mean $PS$ value was slightly lower for the placebo group (0.65) versus the T/T group (0.67).

**Figure 1: Histogram of propensity score values – matching on region**

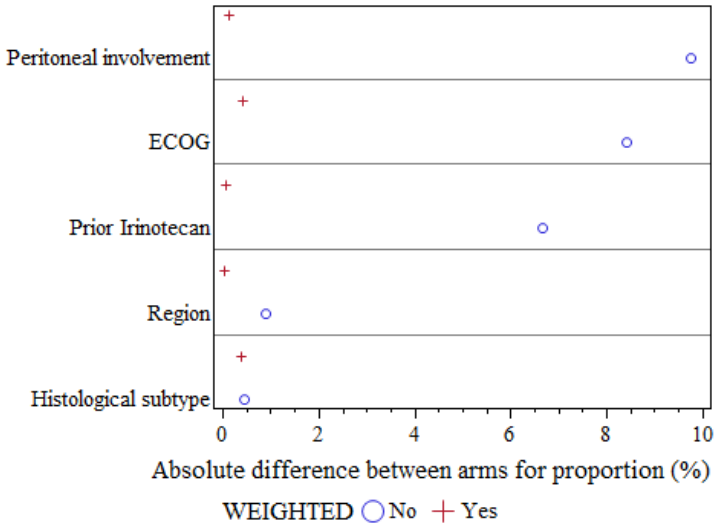



| | 0.45-0.50 | 0.50-0.55 | 0.55-0.60 | 0.60-0.65 | 0.65-0.70 | 0.70-0.75 | 0.75-0.80 | 0.80-0.85 |
|---|---|---|---|---|---|---|---|---|
| T/T | 1% | 6% | 11% | 9% | 48% | 9% | 17% | 0% |
| BSC | 2% | 8% | 14% | 17% | 44% | 2% | 14% | 0% |

**Key:** BSC, best supportive care; T/T, trifluridine/tipiracil.

An overview of the differences between arms for each of the covariates included within the matching analysis is provided in Figure 2. The largest differences were noted for peritoneal involvement and ECOG PS.

**Figure 2: Absolute difference between arms per covariate – matching on region**




**Key:** ECOG, Eastern Cooperative Oncology Group (Performance Score).

Using the $PS$ values as a basis for estimating patient weights, the corresponding Kaplan-Meier curves for each outcome (OS, PFS, and TTD) may be produced (presented in Figure 3). The dashed lines denote unweighted curves, whereas the solid lines denote the adjusted curves.

**Figure 3: Kaplan-Meier curves, naïve versus weighted – matching on region**

| | |
|---|---|
| OS | [redacted] |

| | |
|---|---|
| **PFS** |  |
| **TTD** |  |

**Key:** OS, overall survival; PFS, progression-free survival; TTD, time-to-treatment-discontinuation.

Using these weights, parametric survival models were re-fitted. The statistical goodness-of-fit scores are provided in Table 2. Based on these scores, there is relatively limited evidence to overtly reject the previously selected base-case analysis models, and so these were carried forward to inform updated model results (though alternative models may be selected to inform the cost-effectiveness results).

**Table 2: Statistical goodness-of-fit scores – matching on region**

| Model | Independent | | | | | | Dependent | |
|---|---|---|---|---|---|---|---|---|
| | T/T | | BSC | | Combined | | | |
| | AIC | BIC | AIC | BIC | AIC | BIC | AIC | BIC |
| ***Overall survival*** | | | | | | | | |
| Exponential | 1,784.99 | 1,787.83 | 1,938.66 | 1,940.82 | 3,723.65 | 3,728.64 | 3,723.65 | 3,730.14 |
| Weibull | 1,765.63 | 1,771.30 | 1,937.84 | 1,942.16 | 3,703.47 | 3,713.46 | 3,707.67 | 3,717.42 |
| Gompertz | 1,772.77 | 1,778.44 | 1,939.02 | 1,943.34 | 3,711.79 | 3,721.78 | 3,722.43 | 3,732.17 |
| Log-logistic | 1,770.02 | 1,775.70 | 1,910.11 | 1,914.43 | 3,680.14 | 3,690.13 | 3,678.97 | 3,688.71 |
| Lognormal | 1,768.71 | 1,774.38 | 1,905.49 | 1,909.81 | 3,674.20 | 3,684.19 | 3,672.64 | 3,682.38 |
| Gen gam | 1,766.05 | 1,774.56 | 1,899.23 | 1,905.70 | 3,665.28 | 3,680.26 | 3,674.01 | 3,687.00 |
| ***Progression-free survival*** | | | | | | | | |
| Exponential | 1,858.65 | 1,861.49 | 1,873.49 | 1,875.65 | 3,732.15 | 3,737.14 | 3,732.15 | 3,738.64 |
| Weibull | 1,833.47 | 1,839.15 | 1,847.74 | 1,852.06 | 3,681.21 | 3,691.20 | 3,679.43 | 3,689.17 |
| Gompertz | 1,851.94 | 1,857.61 | 1,875.33 | 1,879.65 | 3,727.27 | 3,737.26 | 3,728.29 | 3,738.03 |

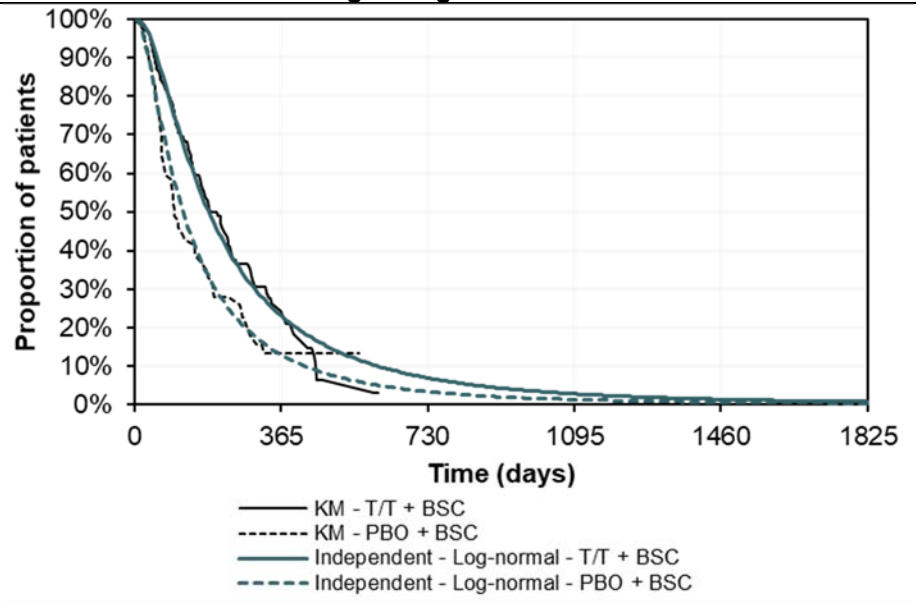
| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Log-logistic | 1,813.22 | 1,818.89 | 1,754.59 | 1,758.91 | 3,567.81 | 3,577.80 | 3,577.78 | 3,587.52 |
| Lognormal | 1,806.75 | 1,812.43 | 1,774.83 | 1,779.14 | 3,581.58 | 3,591.57 | 3,584.95 | 3,594.69 |
| Gen gam | 1,807.15 | 1,815.66 | 1,762.33 | 1,768.81 | 3,569.48 | 3,584.47 | 3,576.31 | 3,589.30 |
| ***Time to treatment discontinuation*** | | | | | | | | |
| Exponential | 1,969.39 | 1,972.21 | | | | | | |
| Weibull | 1,966.19 | 1,971.83 | | | | | | |
| Gompertz | 1,969.23 | 1,974.87 | | | | | | |
| Log-logistic | 1,970.79 | 1,976.43 | | | | | | |
| Lognormal | 1,970.02 | 1,975.66 | | | | | | |
| Gen gam | 1,964.96 | 1,973.42 | | | | | | |

**Key:** AIC, Akaike information criterion; BIC, Bayesian information criterion; BSC, best supportive care; Gen gam, generalised gamma; T/T, trifluridine/tipiracil.
**Note:** Lowest scores highlighted in grey.

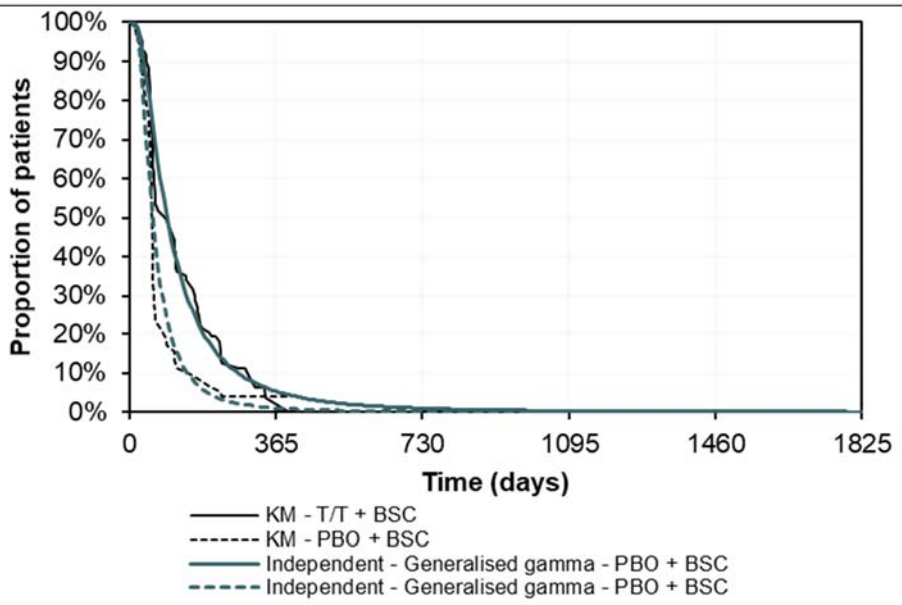
The corresponding survival models for OS (independent lognormal), PFS (independent generalised gamma), and TTD (generalised gamma) are provided in Figure 4.

**Figure 4: Parametric survival models – matching on region**

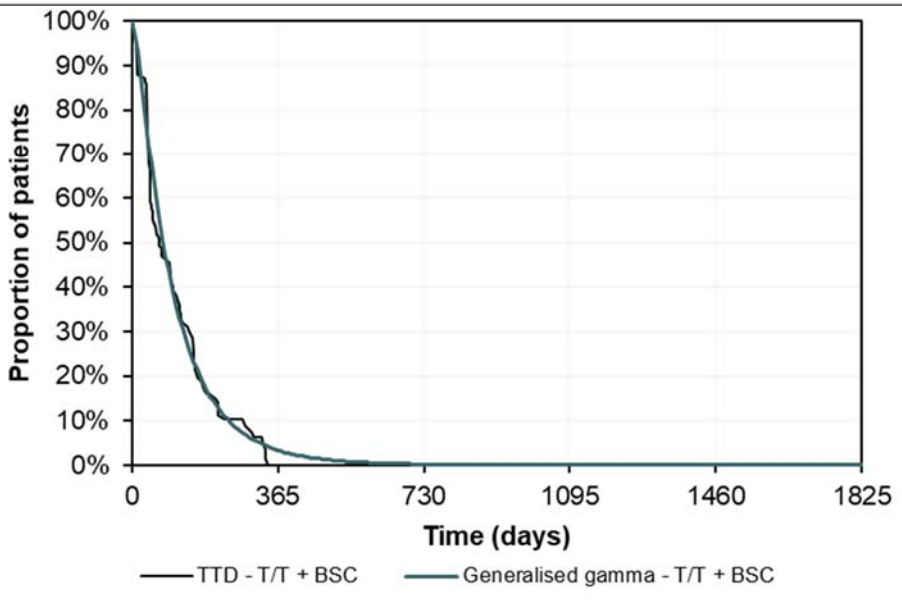
OS




PFS




TTD




**Key:** BSC, best supportive care; KM, Kaplan-Meier; OS, overall survival; PBO, placebo; PFS, progression-free survival; T/T, trifluridine/tipiracil; TTD, time-to-treatment-discontinuation.

The weighted curves were then used to inform the cost-effectiveness model, and the impact on results was recorded (presented in Table 3). After re-weighting, the survival benefit attributable to T/T increases to 0.241 life-years (LYs) – equivalent to a benefit of approximately 2.89 months. The improvement in survival translates to an incremental quality-adjusted life year (QALY) gain of 0.164, leading to an incremental cost-effectiveness ratio (ICER) of £45,662 (including a [redacted] simple PAS discount). As a sensitivity analysis, the corresponding results using a dependent lognormal model for OS are provided in Table 4.

**Table 3: Cost-effectiveness results: 3L only – matching on region (independent lognormal)**

| Arm | Total | | | Incremental | | | ICER |
|---|---|---|---|---|---|---|---|
| | Costs | QALYs | LYs | Costs | QALYs | LYs | |
| BSC | [redacted] | 0.367 | 0.541 | | | | |
| T/T | [redacted] | 0.531 | 0.782 | [redacted] | 0.164 | 0.241 | £45,662 |

**Key:** BSC, best supportive care; ICER, incremental cost-effectiveness ratio; LY, life-year; QALY, quality-adjusted life year; T/T, trifluridine/tipiracil.

**Table 4: Cost-effectiveness results: 3L only – matching on region (dependent lognormal)**

| Arm | Total | | | Incremental | | | ICER |
|---|---|---|---|---|---|---|---|
| | Costs | QALYs | LYs | Costs | QALYs | LYs | |
| BSC | [redacted] | 0.359 | 0.527 | | | | |
| T/T | [redacted] | 0.546 | 0.808 | [redacted] | 0.188 | 0.281 | £39,925 |

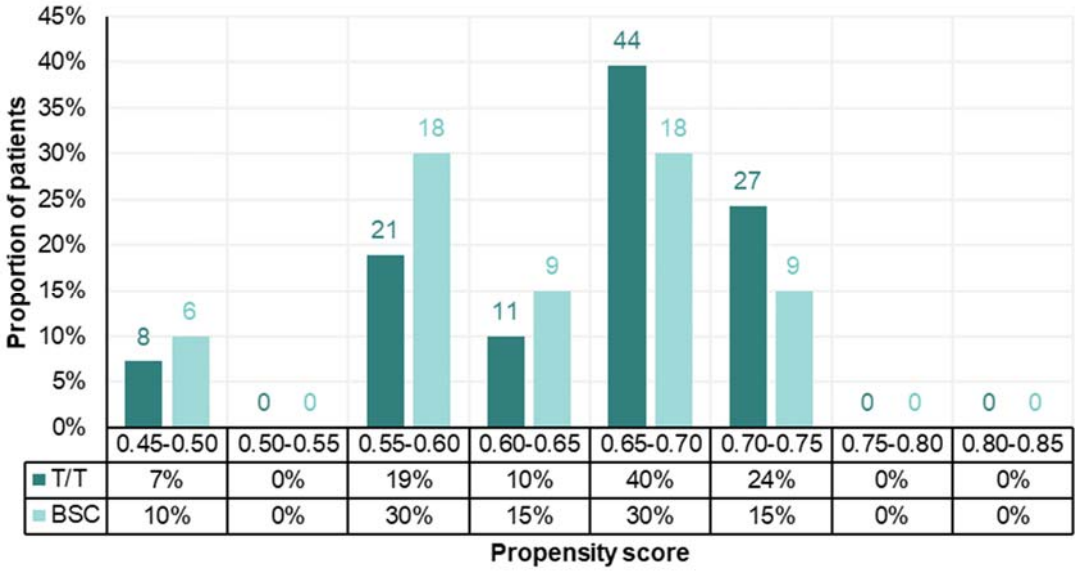
**Key:** BSC, best supportive care; ICER, incremental cost-effectiveness ratio; LY, life-year; QALY, quality-adjusted life year; T/T, trifluridine/tipiracil.

As described previously, this analysis was repeated using weights based on ethnicity instead of geographical region. The overall results are very similar, and for completeness are provided in the Appendix.

### 3.2. 3L EU only

A histogram demonstrating the distribution of $PS$ values across both treatment arms is presented in Figure 5. The distribution of $PS$ values is reasonably consistent across both arms, with a peak observed in the interval of 0.65-0.70, with an additional peak for the palcebo arm seen in the interval 0.55-0.60. The mean $PS$ values for each group were 0.64 (placebo) versus 0.66 (T/T).

**Figure 5: Histogram of propensity scores – 3L EU matching**

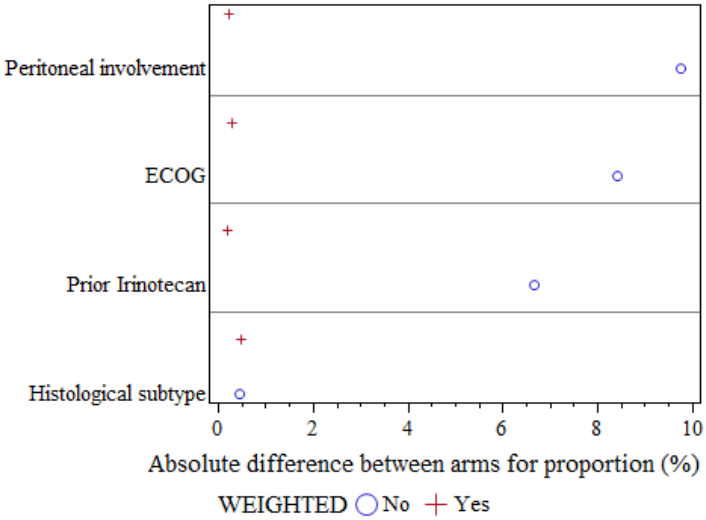



| | 0.45-0.50 | 0.50-0.55 | 0.55-0.60 | 0.60-0.65 | 0.65-0.70 | 0.70-0.75 | 0.75-0.80 | 0.80-0.85 |
|---|---|---|---|---|---|---|---|---|
| T/T | 7% | 0% | 19% | 10% | 40% | 24% | 0% | 0% |
| BSC | 10% | 0% | 30% | 15% | 30% | 15% | 0% | 0% |

**Key:** BSC, best supportive care; T/T, trifluridine/tipiracil.

An overview of the differences between arms for each of the covariates included within the matching analysis is provided in Figure 6. The largest differences were noted for peritoneal involvement and ECOG PS.

**Figure 6: Absolute difference between arms per covariate – 3L EU matching**




**Key:** ECOG, Eastern Cooperative Oncology Group (Performance Score).

**Figure 7: Kaplan-Meier curves, naïve versus weighted – 3L EU matching**

| | | |
|---|---|---|
| **OS** | | [redacted] |
| **PFS** | | [redacted] |
| **TTD** | | [redacted] |

**Key:** OS, overall survival; PFS, progression-free survival; TTD, time-to-treatment-discontinuation.

Using these weights, parametric survival models were re-fitted. The statistical goodness-of-fit scores are provided in Table 5. The previously selected base-case analysis models were carried forward to inform updated model results (though alternative models may be selected to inform the cost-effectiveness results).

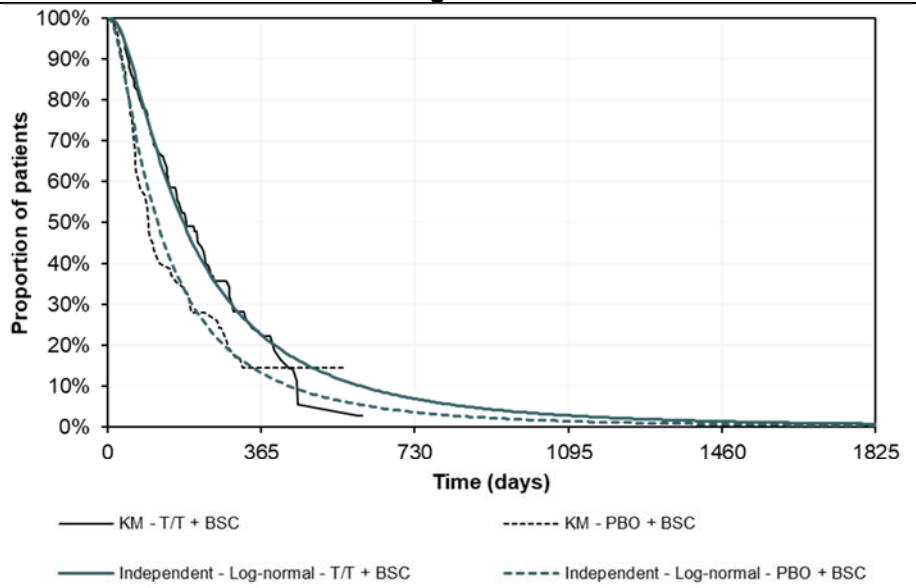
**Table 5: Statistical goodness-of-fit scores – 3L EU matching**

| Model | Independent | | | | | | Dependent | |
|---|---|---|---|---|---|---|---|---|
| | T/T | | BSC | | Combined | | | |
| | AIC | BIC | AIC | BIC | AIC | BIC | AIC | BIC |
| ***Overall survival*** | | | | | | | | |
| Exponential | 1,577.21 | 1,579.92 | 1,719.03 | 1,721.13 | 3,296.24 | 3,301.05 | 3,296.24 | 3,302.53 |
| Weibull | 1,562.44 | 1,567.86 | 1,720.05 | 1,724.24 | 3,282.50 | 3,292.10 | 3,286.94 | 3,296.36 |
| Gompertz | 1,568.84 | 1,574.26 | 1,717.12 | 1,721.31 | 3,285.97 | 3,295.57 | 3,297.46 | 3,306.88 |
| Log-logistic | 1,565.37 | 1,570.79 | 1,692.00 | 1,696.19 | 3,257.37 | 3,266.98 | 3,256.27 | 3,265.70 |
| Lognormal | 1,563.80 | 1,569.22 | 1,687.92 | 1,692.11 | 3,251.72 | 3,261.33 | 3,250.27 | 3,259.69 |
| Gen gam | 1,562.50 | 1,570.63 | 1,678.60 | 1,684.88 | 3,241.10 | 3,255.51 | 3,250.37 | 3,262.94 |
| ***Progression-free survival*** | | | | | | | | |
| Exponential | 1,659.16 | 1,661.87 | 1,684.16 | 1,686.26 | 3,343.32 | 3,348.13 | 3,343.32 | 3,349.61 |
| Weibull | 1,639.39 | 1,644.81 | 1,663.71 | 1,667.90 | 3,303.11 | 3,312.72 | 3,301.28 | 3,310.70 |
| Gompertz | 1,653.65 | 1,659.07 | 1,686.10 | 1,690.29 | 3,339.76 | 3,349.36 | 3,340.65 | 3,350.07 |
| Log-logistic | 1,625.29 | 1,630.71 | 1,585.75 | 1,589.93 | 3,211.04 | 3,220.65 | 3,219.65 | 3,229.07 |
| Lognormal | 1,619.11 | 1,624.53 | 1,601.47 | 1,605.66 | 3,220.58 | 3,230.18 | 3,223.79 | 3,233.22 |
| Gen gam | 1,620.37 | 1,628.50 | 1,591.01 | 1,597.30 | 3,211.39 | 3,225.80 | 3,218.29 | 3,230.86 |
| ***Time to treatment discontinuation*** | | | | | | | | |
| Exponential | 1,764.48 | 1,767.17 | | | | | | |
| Weibull | 1,762.66 | 1,768.04 | | | | | | |
| Gompertz | 1,764.42 | 1,769.80 | | | | | | |
| Log-logistic | 1,769.30 | 1,774.69 | | | | | | |
| Lognormal | 1,768.46 | 1,773.84 | | | | | | |
| Gen gam | 1,762.90 | 1,770.98 | | | | | | |

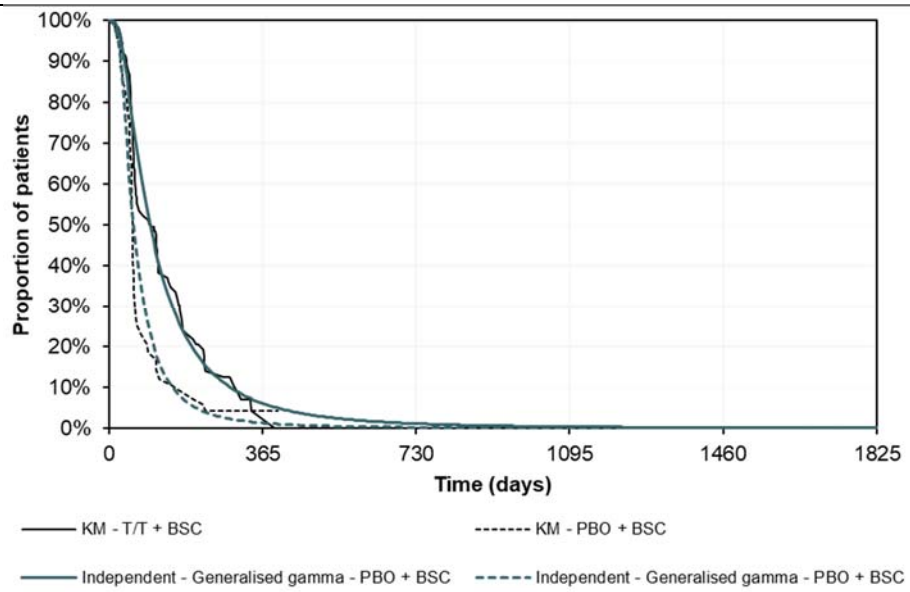
**Key:** AIC, Akaike information criterion; BIC, Bayesian information criterion; BSC, best supportive care; Gen gam, generalised gamma; T/T, trifluridine/tipiracil.
**Note:** Lowest scores highlighted in grey.

**Figure 8: Parametric survival models – 3L EU matching**

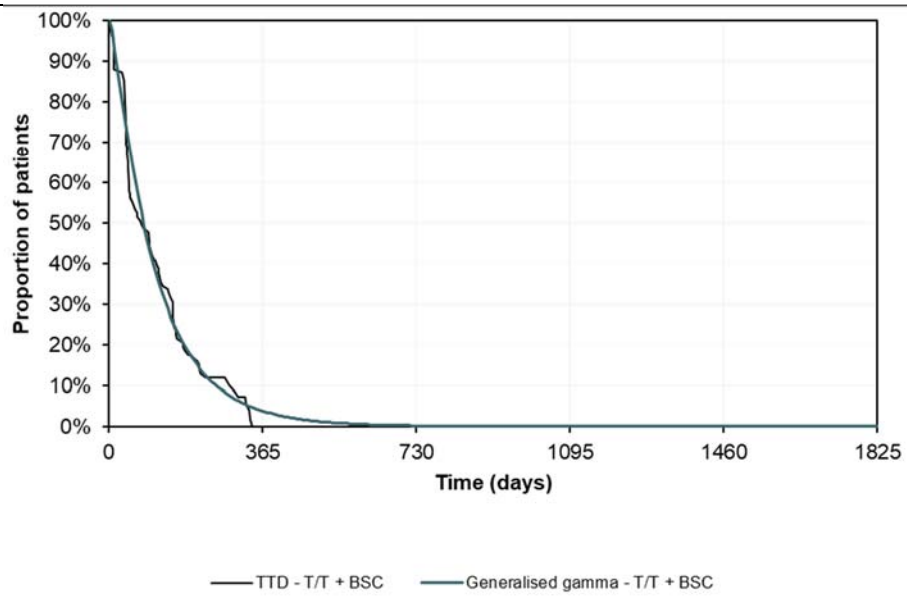
OS




PFS




TTD




**Key:** BSC, best supportive care; KM, Kaplan-Meier; OS, overall survival; PBO, placebo; PFS, progression-free survival; T/T, trifluridine/tipiracil; TTD, time-to-treatment-discontinuation.

The weighted curves were then used to inform the cost-effectiveness model, and the impact on results was recorded (presented in Table 6). After re-weighting, the survival benefit attributable to T/T increases to 0.227 LYs – equivalent to a benefit of approximately 2.72 months. The improvement in survival translates to a QALY gain of 0.156, leading to an ICER of £49,771 (including a [redacted] simple PAS discount). As a sensitivity analysis, the corresponding results using a dependent lognormal model for OS are provided in Table 7.

**Table 6: Cost-effectiveness results: 3L EU only (independent lognormal)**

| Arm | Total | | | Incremental | | | ICER |
|---|---|---|---|---|---|---|---|
| | Costs | QALYs | LYs | Costs | QALYs | LYs | |
| BSC | [redacted] | 0.371 | 0.547 | | | | |
| T/T | [redacted] | 0.527 | 0.774 | [redacted] | 0.156 | 0.227 | £49,771 |

**Key:** BSC, best supportive care; ICER, incremental cost-effectiveness ratio; LY, life-year; QALY, quality-adjusted life year; T/T, trifluridine/tipiracil.

**Table 7: Cost-effectiveness results: 3L EU only (dependent lognormal)**

| Arm | Total | | | Incremental | | | ICER |
|---|---|---|---|---|---|---|---|
| | Costs | QALYs | LYs | Costs | QALYs | LYs | |
| BSC | [redacted] | 0.360 | 0.529 | | | | |
| T/T | [redacted] | 0.547 | 0.806 | [redacted] | 0.186 | 0.277 | £41,858 |

**Key:** BSC, best supportive care; ICER, incremental cost-effectiveness ratio; LY, life-year; QALY, quality-adjusted life year; T/T, trifluridine/tipiracil.

## 4. Discussion

This analysis presents cost-effectiveness results for the third-line population studied within the pivotal TAGS clinical trial, adjusting for differences in five potentially-important baseline variables (see Methods section). The results of the analyses presented in this document are broadly consistent with the findings of the naïve (i.e. unadjusted) analysis previously provided to NICE (see Table 1), with an ICER of between £45,611 and £45,662 (depending on whether matching is performed on region or ethnicity, 3L only population), versus £42,086 (unadjusted analysis, 3L only population). The results of this matched analysis yield a survival improvement of approximately 2.89 months using independent lognormal models for OS, which increased to as much as 3.37 months when using a dependent lognormal model for OS within sensitivity analysis.

The 3-month criterion in NICE's end-of-life criteria is equivalent to a survival gain of approximately 91.3 days (based on 365.25 days per year). The analysis presented in this document demonstrates that after matching adjustment, T/T yields a benefit of 2.89 months, equivalent to 87.9 days (approximately 3.5 days under the 3-month [91.3 day] criterion). While beneath the 3-month value, treatment with T/T offers a survival improvement of 2.89 months, equivalent to a relative survival gain of 44.51% compared to BSC. A benefit of this magnitude is extremely important within the context of this patient population, who have a very poor prognosis.

The TAGS trial recruited patients in the third-line and beyond treatment setting, and the majority of these patients had 3 or more prior lines (i.e. were fourth-line and beyond). Consequently, the sample size available to inform this analysis is smaller than ideal. However, there are still [redacted] patients treated in this setting, [redacted] of which were European. This analysis re-weights patients in both arms to minimise the difference in potentially-important variables at baseline, though it should be noted that any re-weighting approach is subject to limitations owing to the number of patients available to inform the analysis.

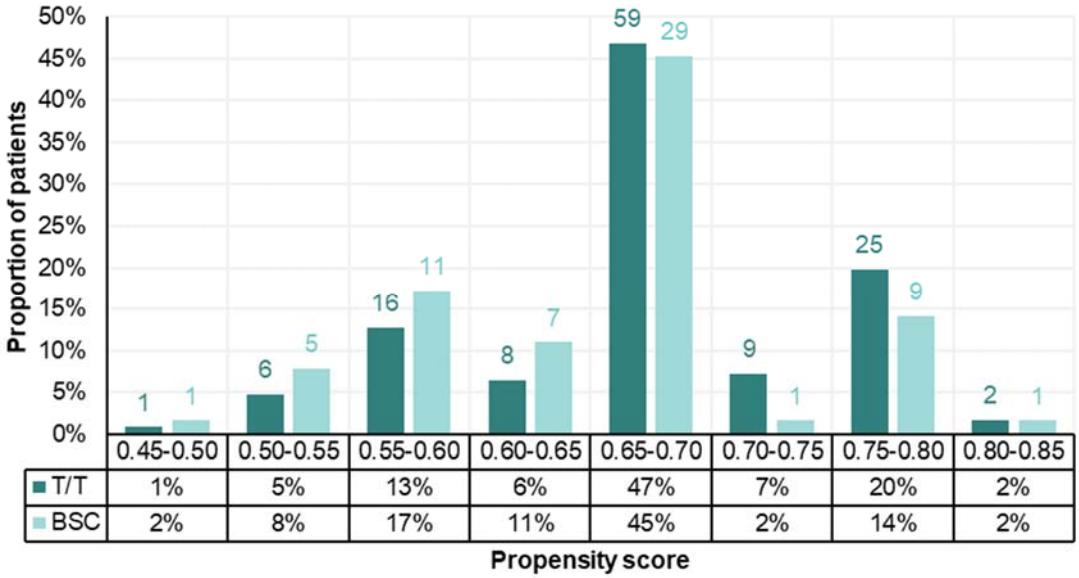
Analyses considering a third-line only population have been provided as this population is expected to most closely resemble the population of patients eligible for treatment with T/T in NHS practice. This is because in NHS practice, there is no other active treatment option that is routinely considered for use at this line. The improved outcomes associated with T/T in a third-line population (versus a third-line *and beyond* population) is aligned with clinical expectation – that is, treatment at later lines of therapy is associated with a poorer prognosis, and hence reduced capacity to derive benefit from active treatment. Thus the cost-effectiveness of T/T is expected to have been underestimated by both Servier's original base-case analysis, and the Committee's preferred base-case analysis provided within the ACD.

In light of the evidence provided as part of this document (and in previous communications with NICE), Servier kindly requests the Committee to again reconsider its positions relating to both the population most suitable for decision making and T/T meeting the life extension criterion in order to be considered a life-extending end-of-life treatment. Servier appreciates the opportunity to provide this matching analysis to address the concerns raised at ACM2, and believes this analysis demonstrates the robustness of the survival gain associated with T/T in the population most closely aligned with those expected to be treated in NHS practice.

## Appendix: Weighting by ethnicity (instead of geographical region)

A histogram demonstrating the distribution of $PS$ values across both treatment arms is presented in Figure 9. The distribution of $PS$ values is reasonably consistent across both arms, with a peak observed in the interval of 0.65-0.70 (as per the analysis based on geographical region). The mean $PS$ values for each group were identical (to two decimal places) to those obtained from the analysis based on geographical region (0.65 [placebo] versus 0.67 [T/T]).

**Figure 9: Histogram of propensity score values – matching on ethnicity**

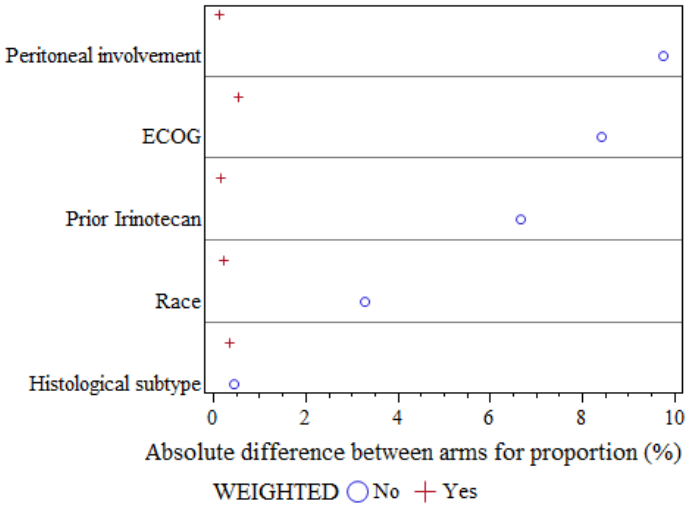



| | 0.45-0.50 | 0.50-0.55 | 0.55-0.60 | 0.60-0.65 | 0.65-0.70 | 0.70-0.75 | 0.75-0.80 | 0.80-0.85 |
|---|---|---|---|---|---|---|---|---|
| T/T | 1% | 5% | 13% | 6% | 47% | 7% | 20% | 2% |
| BSC | 2% | 8% | 17% | 11% | 45% | 2% | 14% | 2% |

**Key:** BSC, best supportive care; T/T, trifluridine/tipiracil.

An overview of the differences between arms for each of the covariates included within the matching analysis is provided in Figure 10. The largest differences were noted for peritoneal involvement and ECOG PS.

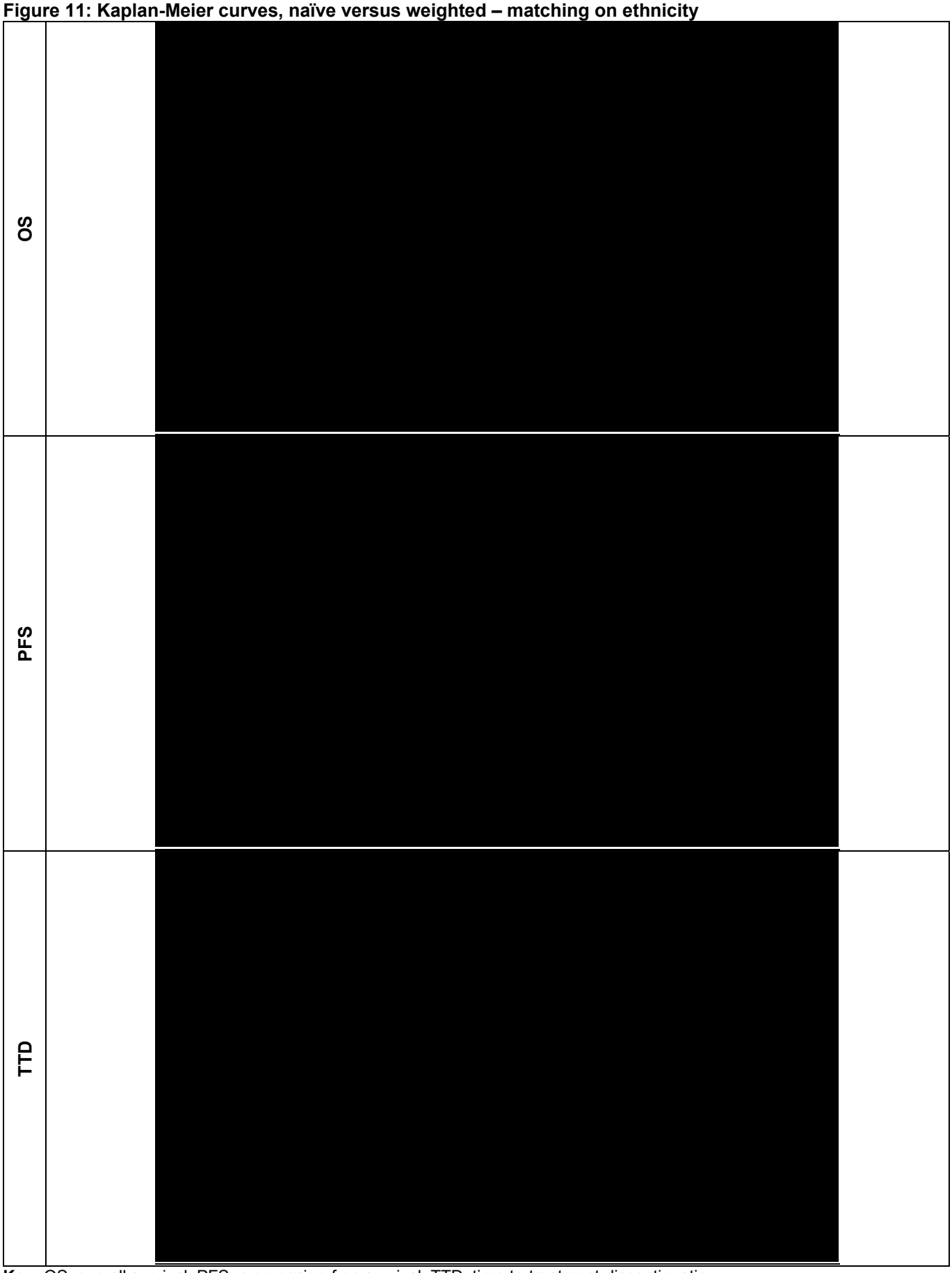
**Figure 10: Absolute difference between arms per covariate – matching on ethnicity**




**Key:** ECOG, Eastern Cooperative Oncology Group (Performance Score).
**Note:** For the context of using this diagram, the terms “race” and “ethnicity” may be considered interchangeable.

Using the $PS$ values as a basis for estimating patient weights, the corresponding Kaplan-Meier curves for each outcome (OS, PFS, and TTD) may be produced (presented in Figure 11).

**Figure 11: Kaplan-Meier curves, naïve versus weighted – matching on ethnicity**




**Key:** OS, overall survival; PFS, progression-free survival; TTD, time-to-treatment-discontinuation.

Using these weights, parametric survival models were re-fitted. The statistical goodness-of-fit scores are provided in Table 8. The previously selected base-case analysis models were carried forward to inform updated model results (though alternative models may be selected to inform the cost-effectiveness results).

**Table 8: Statistical goodness-of-fit scores – matching on ethnicity**

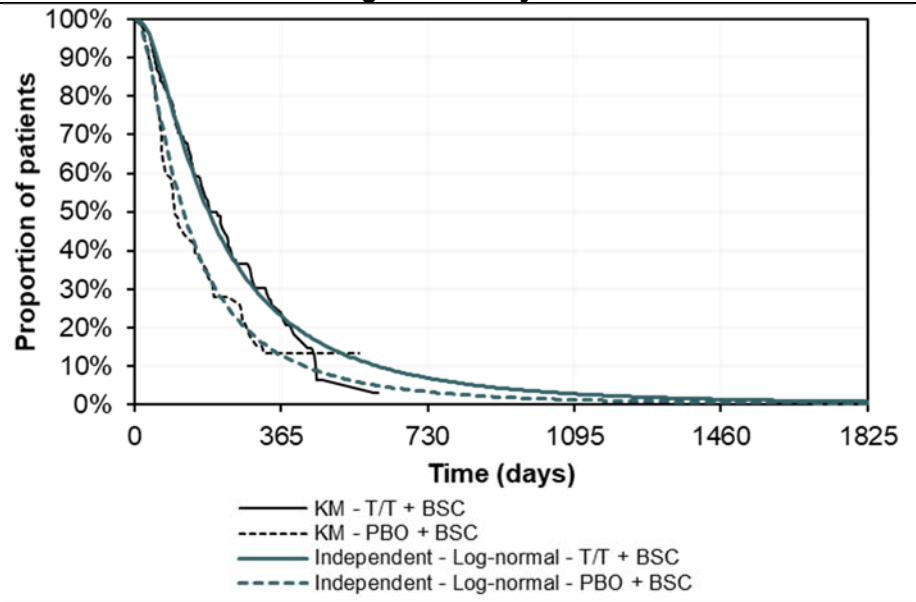
| Model | Independent | | | | | | Dependent | |
|---|---|---|---|---|---|---|---|---|
| | T/T | | BSC | | Combined | | | |
| | AIC | BIC | AIC | BIC | AIC | BIC | AIC | BIC |
| ***Overall survival*** | | | | | | | | |
| Exponential | 1,780.34 | 1,783.17 | 1,948.30 | 1,950.46 | 3,728.64 | 3,733.63 | 3,728.64 | 3,735.13 |
| Weibull | 1,761.14 | 1,766.82 | 1,946.85 | 1,951.17 | 3,708.00 | 3,717.99 | 3,711.55 | 3,721.29 |
| Gompertz | 1,768.33 | 1,774.00 | 1,949.20 | 1,953.51 | 3,717.52 | 3,727.51 | 3,726.91 | 3,736.65 |
| Log-logistic | 1,765.42 | 1,771.09 | 1,920.18 | 1,924.49 | 3,685.60 | 3,695.59 | 3,684.26 | 3,694.01 |
| Lognormal | 1,764.11 | 1,769.78 | 1,915.53 | 1,919.85 | 3,679.64 | 3,689.63 | 3,677.95 | 3,687.69 |
| Gen gam | 1,761.52 | 1,770.03 | 1,910.35 | 1,916.83 | 3,671.87 | 3,686.86 | 3,679.55 | 3,692.53 |
| ***Progression-free survival*** | | | | | | | | |
| Exponential | 1,855.48 | 1,858.31 | 1,876.42 | 1,878.58 | 3,731.90 | 3,736.89 | 3,731.90 | 3,738.39 |
| Weibull | 1,829.96 | 1,835.63 | 1,849.81 | 1,854.13 | 3,679.77 | 3,689.76 | 3,677.97 | 3,687.71 |
| Gompertz | 1,848.55 | 1,854.22 | 1,878.23 | 1,882.54 | 3,726.77 | 3,736.76 | 3,727.82 | 3,737.56 |
| Log-logistic | 1,809.85 | 1,815.52 | 1,754.77 | 1,759.09 | 3,564.62 | 3,574.61 | 3,575.41 | 3,585.15 |
| Lognormal | 1,803.47 | 1,809.14 | 1,775.83 | 1,780.15 | 3,579.30 | 3,589.29 | 3,583.09 | 3,592.83 |
| Gen gam | 1,803.93 | 1,812.44 | 1,763.22 | 1,769.70 | 3,567.15 | 3,582.14 | 3,574.64 | 3,587.63 |
| ***Time to treatment discontinuation*** | | | | | | | | |
| Exponential | 1,966.19 | 1,969.01 | | | | | | |
| Weibull | 1,962.94 | 1,968.58 | | | | | | |
| Gompertz | 1,966.02 | 1,971.66 | | | | | | |
| Log-logistic | 1,967.52 | 1,973.16 | | | | | | |
| Lognormal | 1,966.85 | 1,972.49 | | | | | | |
| Gen gam | 1,961.74 | 1,970.20 | | | | | | |

**Key:** AIC, Akaike information criterion; BIC, Bayesian information criterion; BSC, best supportive care; Gen gam, generalised gamma; T/T, trifluridine/tipiracil.
**Note:** Lowest scores highlighted in grey.

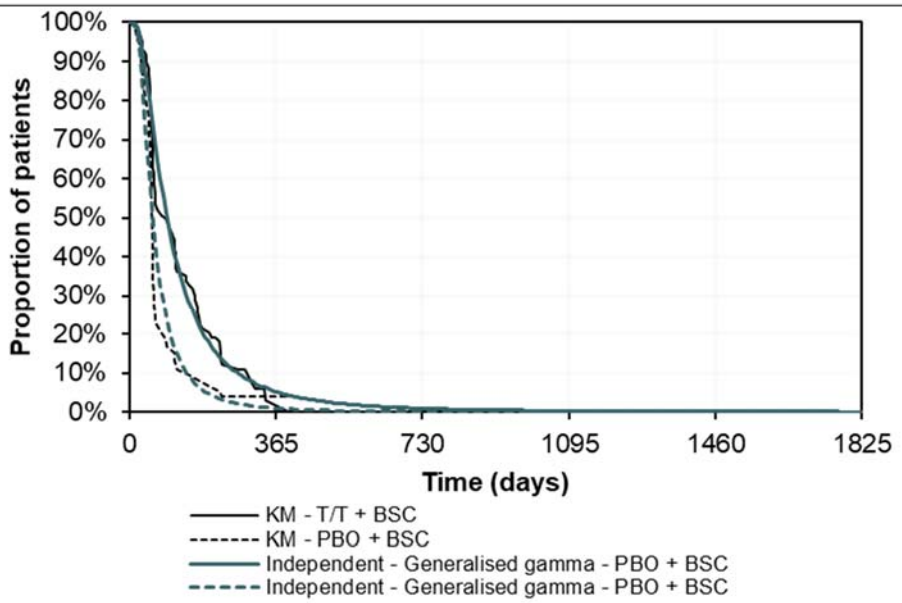
The corresponding survival models for OS (independent lognormal), PFS (independent generalised gamma), and TTD (generalised gamma) are provided in Figure 12.

**Figure 12: Parametric survival models – matching on ethnicity**

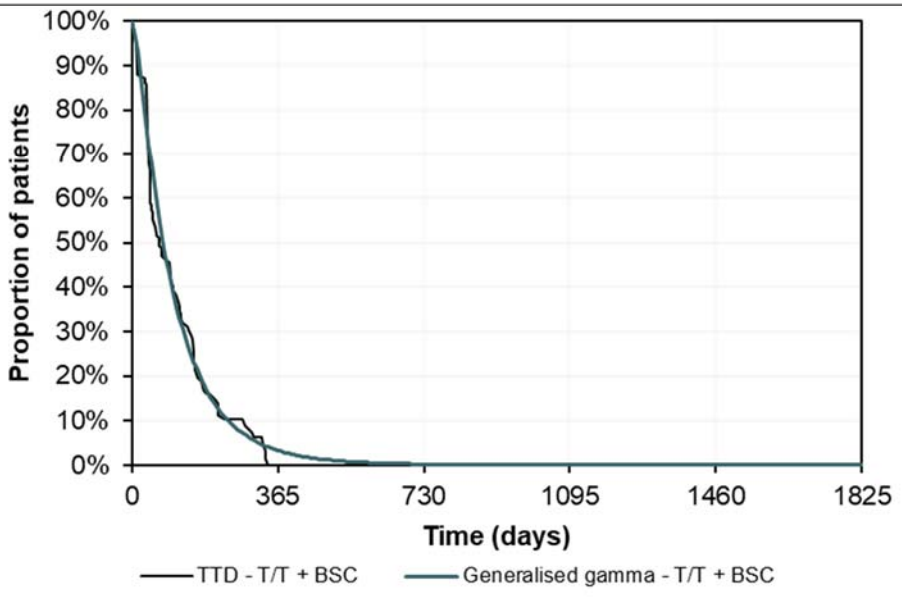
OS




PFS




TTD




**Key:** BSC, best supportive care; KM, Kaplan-Meier; OS, overall survival; PBO, placebo; PFS, progression-free survival; T/T, trifluridine/tipiracil; TTD, time-to-treatment-discontinuation.

The weighted curves were then used to inform the cost-effectiveness model, and the impact on results was recorded (presented in Table 9). After re-weighting, the survival benefit attributable to T/T increases to 0.241 LYs – equivalent to a benefit of approximately 2.89 months. The improvement in survival translates to a QALY gain of 0.164, leading to an ICER of £45,611 (including a [redacted] simple PAS discount). As a sensitivity analysis, the corresponding results using a dependent lognormal model for OS are provided in Table 10.

**Table 9: Cost-effectiveness results: 3L only – matching on ethnicity (independent lognormal)**

| Arm | Total | | | Incremental | | | ICER |
|---|---|---|---|---|---|---|---|
| | Costs | QALYs | LYs | Costs | QALYs | LYs | |
| BSC | [redacted] | 0.366 | 0.540 | | | | |
| T/T | [redacted] | 0.530 | 0.781 | [redacted] | 0.164 | 0.241 | £45,611 |

**Key:** BSC, best supportive care; ICER, incremental cost-effectiveness ratio; LY, life-year; QALY, quality-adjusted life year; T/T, trifluridine/tipiracil.

**Table 10: Cost-effectiveness results: 3L only – matching on ethnicity (dependent lognormal)**

| Arm | Total | | | Incremental | | | ICER |
|---|---|---|---|---|---|---|---|
| | Costs | QALYs | LYs | Costs | QALYs | LYs | |
| BSC | [redacted] | 0.359 | 0.529 | | | | |
| T/T | [redacted] | 0.543 | 0.803 | [redacted] | 0.184 | 0.274 | £40,724 |

**Key:** BSC, best supportive care; ICER, incremental cost-effectiveness ratio; LY, life-year; QALY, quality-adjusted life year; T/T, trifluridine/tipiracil.

# Additional clarification request based on the third-line population enrolled within the TAGS clinical trial

## 1. Introduction

On 6 May 2020, a clarification request was sent by the National Institute for Health and Care Excellence (NICE) to the company (Servier), which asked: *"Please provide an explanation as to why the use of propensity score weights have had such limited impact on the time to event data"*.

This document contains an explanation of the analyses conducted to date, the corresponding results, and interpretation of the findings. Where applicable, the corresponding documentation from which specific statements have been extracted are noted within this document.

## 2. Analyses conducted

Prior to the second Appraisal Committee meeting (ACM2), Servier presented the results of an unadjusted (henceforth termed "naïve") comparison of the third-line (3L) only population enrolled within the TAGS trial (i.e. patients with only 2 prior therapies). Analyses considering a 3L only population have been provided as this population is expected to most closely resemble the population of patients eligible for treatment with T/T in NHS practice. This is because in NHS practice, there is no other active treatment option that is routinely considered for use at this line.

After ACM2, and following an invitation from NICE, Servier provided additional analysis accounting for potential imbalances in the following patient characteristics:

1. **Peritoneal metastases:** Patients with an absence or presence of peritoneal metastases (also known as peritoneal involvement)
2. **Eastern Cooperative Oncology Group Performance Status (ECOG PS):** Patients with an ECOG PS of 0 versus 1
3. **Histology:** Patients with intestinal versus non-intestinal histology
4. **Ethnicity or Region:** Patients residing in Japan or the rest of the world ("region") or patients who are Asian versus non-Asian ("ethnicity")
5. **Prior irinotecan:** Patients with previous exposure to irinotecan versus no previous exposure to irinotecan

The variables included within the weighting analysis were requested by NICE, and were based on the key characteristics deemed to be of potential importance in terms of estimating the difference in time-to-event outcomes observed in TAGS. The statistical methods used to conduct the weighting analysis are discussed in a standalone document which was provided to NICE in March 2020 (titled: *"Additional weighting analysis based on the third-line population enrolled within the TAGS clinical trial"*).

## 3. Analysis results

The mean incremental survival benefit in the 3L only and 3L EU subgroups (based on the committee's preferred model settings and assumptions) were +3.21 months and +3.05 months, respectively. As described previously, an improved survival benefit in the 3L subgroups compared to the more heavily pre-treated intention-to-treat (ITT) population from TAGS is aligned with clinical expectation – that is, treatment at later lines of therapy is associated with a poorer prognosis, and hence reduced capacity to derive benefit from active treatment.

The corresponding results for the propensity weighting adjusted analyses were equivalent to a benefit of approximately +2.89 months (3L only) and +2.72 months (3L EU). This result was shown to be largely unaffected by the specification of the propensity score according to either ethnicity or geographical region (given that these characteristics are expected to be very similar), yet as described previously it is geographical region that has been previously described to have an effect on outcomes seen in mGC.

From these results, it may be inferred that after adjusting for the potential imbalances in the five characteristics, a small limited reduction in the survival benefit is seen for both the 3L only and 3L EU populations (i.e. +3.21 months decreasing to +2.89 months, and +3.05 months decreasing to +2.72 months).

## 4. Interpretation of findings

Prior to the analyses being conducted by Servier, it was expected that any adjustments to account for potential imbalances between the two treatment arms would not be expected to have a large impact on the overall results. In Servier's response to the ACD (wherein 3L only results were first presented to NICE), it was stated that *"… [The 3L only] subgroup was not pre-specified within the TAGS trial protocol, and so the results are unavoidably also at risk of confounding. Nevertheless, the number of treatment lines is associated with all three stratification factors considered within the TAGS trial, mitigating this risk to an extent."*

To further explain this point, the TAGS trial included three stratification factors:

1. **Geographical region:** Patients residing in Japan or the rest of the world
2. **Eastern Cooperative Oncology Group Performance Status (ECOG PS):** Patients with an ECOG PS of 0 versus 1
3. **Prior treatment with ramucirumab:** Patients with previous exposure to ramucirumab versus no previous exposure to ramucirumab

It may be observed that each of these three stratification factors is associated with the number of previous treatment lines a given patient may be exposed to:

- Patients in Japan are expected to be treated with more treatment lines, as a greater number of treatments are both available and recommended in treatment guidelines, including (but not limited to) ramucirumab and irinotecan
- Patients with an ECOG PS of 0 are more likely to have been less heavily pre-treated with mGC compared to those with an ECOG PS of 1 (based on the general principle that as patients progress through multiple lines of treatment, their overall health is expected to decline)
- Patients that have previously been treated with ramucirumab are expected to have (on average) received more lines of treatment compared to those that have not

Based on the above, it may be noted that the risk of any imbalances between treatment arms that may arise from considering a 3L only subgroup is (to an extent) mitigated by the fact that line of therapy is linked to each of the stratification factors in the study. Servier highlights that this observation is by no means a guarantee that there will be no imbalances between arms. However, the specification of these stratification factors in TAGS means that the 3L subgroup is unlikely to have any major imbalances as these would be expected to be addressed via stratification.

There are several other reasons for why the weighting analysis would be expected to have only a small impact on outcomes (versus the unadjusted analysis), which are also described below.

Servier considers that any potential imbalances are not universally biased in favour of either treatment arm. Below a brief description of the likely effect of each of the individual characteristics adjusted for within the propensity weighting analysis is provided.

Region – Region was a pre-specified stratification factor. People residing in Asian countries have been identified as having an improved overall survival compared to western countries due to a combination of differences in the treatment pathway, corresponding health care systems and differing exposure to environmental risk factors across the regions (e.g. exposure to Helicobacter pylori [H.pylori] and local diets). In the 3L only population group 7% of patients were from Japan in the T/T arm versus 6% in the placebo arm. This factor is fairly well balanced between arms and as

a result (although slightly favouring the T/T arm) re-weighting between arms is expected to have a minimal impact on the survival benefit seen in the T/T arm (and this variable is not applicable to the 3L EU group, by definition of this group considering only European patients)

ECOG PS – ECOG PS is a well-documented prognostic factor, where an increase in the proportion of patients with a poorer ECOG status (i.e. a greater ECOG PS) is associated with a worse survival outcome. Within the 3L only subgroup, 40% of patients in the T/T arm compared with 31% in the placebo arm have an ECOG PS of 0 with the remainder of patients in each arm having an ECOG PS of 1. As stated on page 12 of the ERG addendum report, clinical advice provided to the ERG suggested that differences between ECOG PS of 0 and 1 are minor and can be prone to subjectivity – therefore it is difficult to ascertain the impact the imbalances of ECOG PS 0 vs 1 between the two arms could have on the survival benefit seen and may account for why reweighting of this imbalance resulting in minimal impact. For completeness, the equivalent proportions for the 3L EU subgroup are 41% T/T vs 32% placebo for ECOG PS 0.

Prior Irinotecan – In the 3L only subgroup, 23% of patients in the T/T arm vs 30% in the placebo arm previously received irinotecan. However, all patients had received the same *number* of prior lines of treatment and there is no clinical evidence to confirm receiving irinotecan vs another regimen at a previous line of therapy has an impact on either overall prognosis or as a treatment modifier for T/T. Therefore, reweighting to account for a potential imbalance in prior irinotecan use between arms is not expected to have an independent impact on results. In TAGS, through multiple subgroup testing, the forest plot suggested that T/T compared to placebo had relatively less efficacy in patients with prior irinotecan use – however, this was only in a subgroup of patients where a wide confidence interval was seen, and the point estimate still favoured T/T – no other clinical data exist where prior irinotecan use affected the efficacy of T/T or the baseline prognosis of patients. For completeness, the equivalent proportions for the 3L EU subgroup are 23% T/T vs 32% placebo for prior irinotecan use.

Presence of peritoneal metastases – In the 3L only subgroup, 26% vs 36% of patients treated with T/T vs placebo had peritoneal metastases. The presence of peritoneal metastasis is associated with poor prognosis and the imbalance could disfavour the placebo arm, thus is may be expected to affect the survival benefit of T/T negatively when this factor was rebalanced.

Histology – Intestinal histological type gastric cancer has a favourable prognostic outcome compared to gastric cancer with a diffuse histology. In the 3L only subgroup, 21% in the T/T group and 22% in the placebo group had intestinal type gastric cancer which is deemed to be fairly well balanced and reweighting is expected to have little impact. Diffuse histology type gastric cancer is seen in 17% of the T/T arm and 11% of the placebo arm which could disfavour the T/T arm. Reweighting of this factor may be expected to affect the survival benefit of T/T positively when this factor was rebalanced.

Thus, taking into account the above considerations on the prognostic value of each of the five factors, the scale of the imbalance addressed in the reweighting and the differing direction they exert any verified prognostic effect (e.g. presence of peritoneal metastases vs histological subtype) may explain why the propensity score analysis of these five factors has had limited impact on survival outcomes.

In addition to the points noted above, it is important to highlight that any adjustment to account for potential imbalances is (unavoidably) based on a small sample size for the 3L only subgroup in TAGS. This means any imbalances that may appear large when presented as percentages correspond to a relatively small number of patients.

For example, the potential imbalance in ECOG PS would be theoretically corrected if approximately n=5 placebo patients were ECOG PS 0 instead of 1. If the same imbalance existed in the ITT population, the equivalent number of placebo patients would be approximately n=15 (i.e. three times as many patients, based on the number of patients in the 3L only versus ITT populations). Servier therefore considers it

important to note (for context) that some imbalances seen in the 3L only subgroup would be expected through random chance (especially noting that TAGS was randomised 2:1).

## 5. Conclusion

The explanations provided within this document to describe the impact of reweighting on outcomes have been previously presented either in written responses to the ERG/NICE, or discussed at the appraisal committee meetings.

Based on the information provided above, the increase in T/T benefit seen in the 3L only subgroup may reasonably be expected to be due to an expected increase in the efficacy of T/T when used at an earlier line of therapy; and that this effect supersedes any overall potential imbalances in characteristics that prior to adjustment may affect the prognosis and survival benefit.

# Additional clarification request based on the third-line population enrolled within the TAGS clinical trial

## 1. Introduction

On 15 May 2020, a clarification request was sent by the National Institute for Health and Care Excellence (NICE) to the company (Servier), which asked: *"The ERG thanks the company for responding to its additional clarification request. In addition to the supporting information provided by the company, the ERG would like the company to clarify exactly how the weights have been applied and why the naive and adjusted Kaplan-Meier survival functions in Figure 7 of the document, "Additional weighting analysis based on the third-line population enrolled within the TAGS clinical trial" are not more different. The ERG's concern is that the distributions of propensity scores (Figure 5) differ between treatment groups and are not centred on 0.5 as they would be in a completely randomised study, but do not seem to affect the Kaplan-Meier survival functions."*

This document contains Servier's response to this request for clarification.

## 2. Response

The mean propensity score ($PS$) is not centred on 0.5 because the TAGS study had a 2:1 randomisation ratio (i.e. ~66.6% chance to be randomised to receive T/T). A complete balance across groups would lead to a mean $PS$ of approximately two-thirds for both groups. The difference in terms of $PS$ between arms is low (mean values within 0.03 of two-thirds across the range of scenarios presented), and so the global imbalance between arms is small. If the imbalance between arms had been greater, the $PS$ difference between arms would reflect this.

Of note, in the third-line (3L) only base-case analysis, the $PS$ is slightly higher in T/T group, suggesting that patients from the 3L only subgroup randomised to T/T had a slightly higher chance to be randomised in that group. Based on their baseline characteristics, the 3L patients from the BSC group had a 65% chance to be randomised to T/T. However, on average, these values are aligned with the 2:1 randomisation ratio.

The distribution of $PS$ is not identical across the two groups, but is similar. The main difference is that the BSC group (lighter columns in Figure 5 within Servier's previous response document) comprises slightly more patients in the 0.55-0.65 bands versus the T/T group (darker columns) which includes more in the 0.65-0.75 buckets. This is reflected in the difference of the mean scores (0.67 versus 0.65). However, nearly all patients have a $PS$ of between 0.55 and 0.75, which is aligned with the 2:1 randomisation.

The weights were calculated using the approach described in Servier's previous response, using the inverse probability of treatment weighting (IPTW) method. The weights were then incorporated within the estimation of the Kaplan-Meier curves (and later, the parametric survival models) using the *weight* argument in *survfit* (Kaplan-Meier curves) and the *weights* argument in *flexsurvreg* (parametric survival models).

**Trifluridine-tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies: A Single Technology Appraisal. The ERG's critique of the company's response provided after the second appraisal committee.**


| | |
|---|---|
| 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<br><br>John Stevens, Reader in Decision Science, ScHARR, University of Sheffield, Sheffield, UK<br><br>Andrew Metry, Research Associate, ScHARR, University of Sheffield, Sheffield, UK |
| Correspondence Author | Matt Stevenson, Professor of Health Technology Assessment, ScHARR, University of Sheffield, Sheffield, UK |
| Date completed | Date completed (21/04/2020) |



**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 also like to thank 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, Stevens J, Metry A. Trifluridine-tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies: A Single Technology Appraisal. The ERG's critique of the company's response provided after the second appraisal committee. School of Health and Related Research (ScHARR), 2020.

Contributions of authors

Matt Stevenson and Andrew Metry critiqued the health economic analysis submitted by the company. John Stevens critiqued the statistical analyses provided by the company to account for potential imbalances in the study populations. All authors were involved in drafting and commenting on the final report.

# 1 Background

NICE appraised trifluridine/tipiracil (TFT) (Lonsurf®) at an appraisal committee on the 5th of December 2019. This resulted in an Appraisal Consultation Document (ACD) which did not recommend the use of TFT for treating metastatic gastric cancer or gastro-oesophageal junction adenocarcinoma in adults who have had 2 or more systemic treatment regimens.[1] A second appraisal committee meeting was held on the 18th of February 2020; however, no further public documentation was released.

On the 9th April 2020, the Evidence Review Group (ERG) received documents, via NICE, from the company (Servier Laboratories) which provided additional analyses and a revised Patient Access Scheme, with the simple discount increased from [redacted]% to [redacted]%.[2] In this document the ERG summarises the main points raised by the company and, where appropriate, to provide a critique of these issues. The bulk of the ERG critique relates to the new analysis provided by the company to account for imbalances between patient characteristics in third-line patients treated with TFT and BSC (hereafter called TFT) and those receiving placebo and BSC only (hereafter called BSC) in the TAGS study.[3]

# 2 Key Points raised by the company in its response to the ACD

For brevity, the position of the company has been summarised by the ERG.

## 2.1 A teleconference between NICE and the company

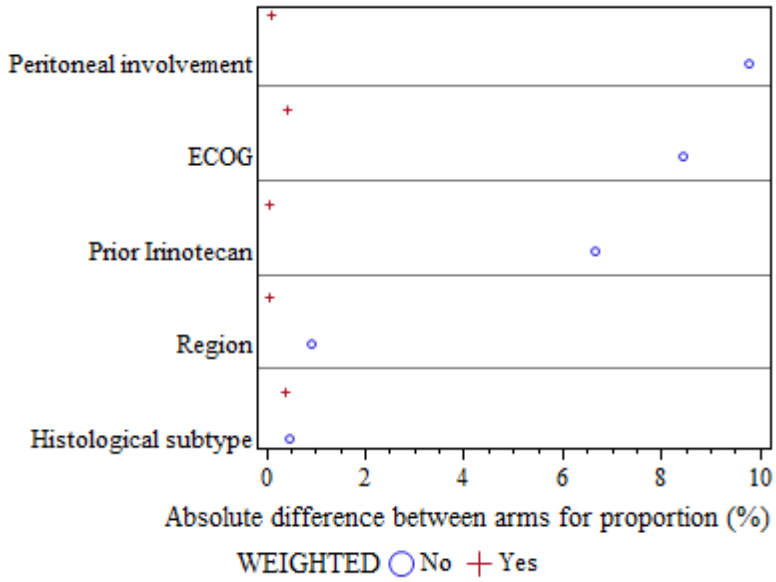
In the previous ERG report which critiqued the company's response to the ACD [4] the ERG had concerns that the analyses focussed on patients receiving third-line treatment only which broke the randomisation and discarded data from [redacted] of the patients recruited to the RCT. The ERG believes that the NICE appraisal committee shared these concerns. The company states that there was a teleconference between NICE and the company on the 3rd of March 2020 when it was asked by NICE to consider five potentially important characteristics that may be imbalanced between the TFT and BSC arms in patients receiving third-line treatment. These were: peritoneal metastases (absence of presence); Eastern Cooperative Oncology Group Performance Status (ECOG PS) (0 vs 1); histology (intestinal vs non-intestinal); ethnicity or region (residing in Japan vs the rest of the world (ROW) or Asian vs Non-Asian); and prior irinotecan (previous exposure vs no previous exposure). The ERG notes that all measured covariates that are predictive of response should be included in a regression analysis irrespective of whether the characteristic is balanced across treatment arms. In non-linear models, omitted measured predictors result in biased estimates of treatment effect and incorrect estimates of standard error. Further detail is provided in Section 2.2.

## 2.2 The weighting analyses undertaken to account for imbalances in potential prognostic factors between the TFT and BSC arms

The company provide details relating to a propensity score analysis which aimed '*to address the potential bias in the estimation of treatment effects that can arise with imbalances in baseline patient characteristics'*. The company used the inverse probability of treatment weighting (IPTW) method which assigns weights based on patients' propensity scores. It is assumed that there are no unmeasured confounders so that given the propensity score the distribution of covariates is the same in the TFT and BSC arms.

Figure 1 reproduces the company's response that shows the absolute difference between the TFT and BSC arms prior to re-weighting of patients and after matching for the five characteristics (using region rather than ethnicity). Additional figures are provided in the company response. It is seen that after weighting, the difference in the mean values of the covariates between the two groups has been reduced.

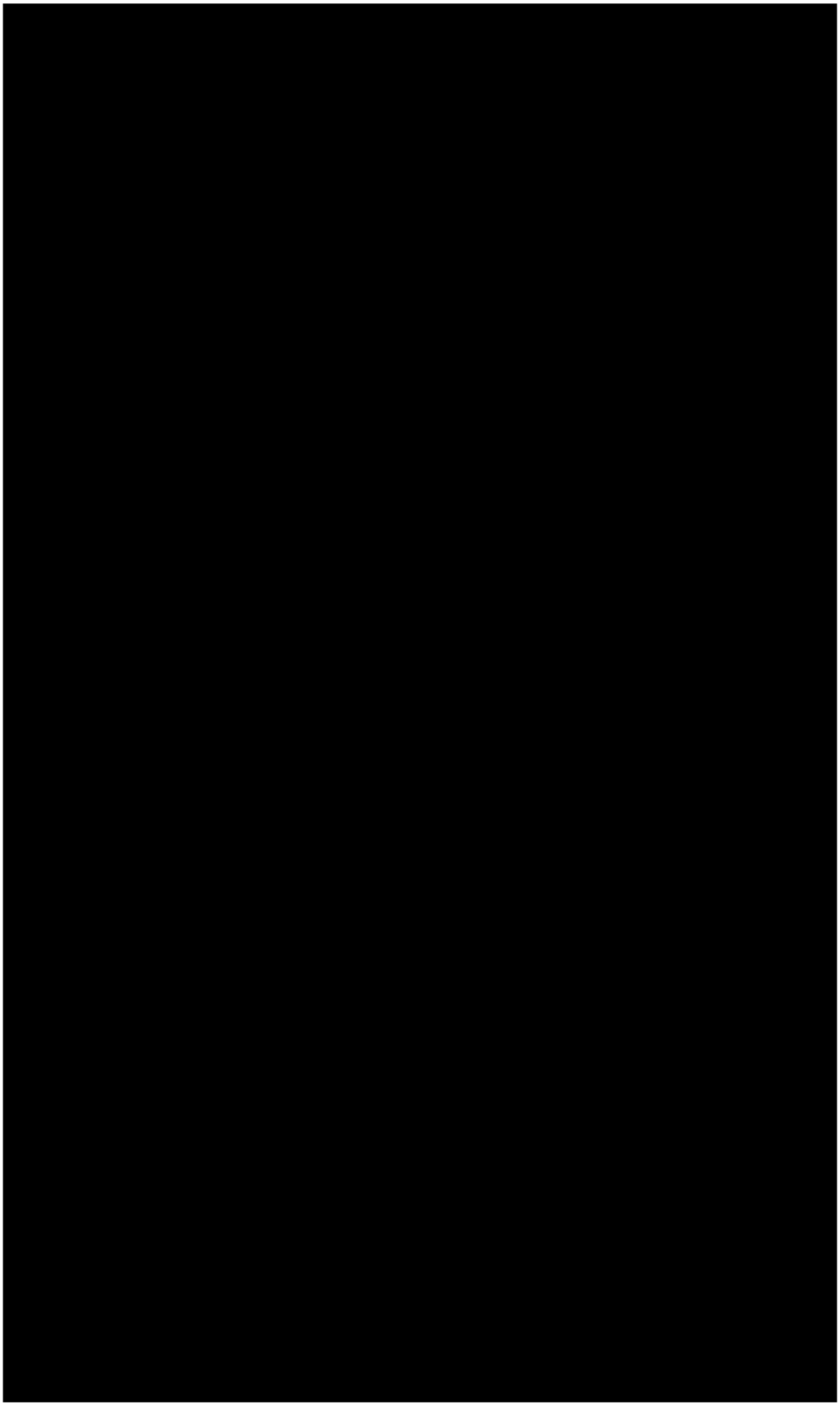
**Figure 1: Absolute difference between arms per covariate – matching on region (reproduced from company's additional analyses, Figure 2)**




The impact of the weighting on the Kaplan-Meier (KM) survival functions for overall survival (OS), progression-free survival (PFS) and time to treatment discontinuation (TTD) is shown in Figure 2.

It is observed that following the weighting analysis [redacted]

**Figure 2: The comparison of naïve KM survival functions and those after weighting of the population (reproduced from company's additional analyses, Figure 3)**



### 2.3 Survival modelling using the weighted survival functions

Relative goodness-of-fit statistics (the Akaike Information Criterion [AIC] and the Bayesian Information Criterion [BIC]) were provided for probability distributions fitted to the weighted time-to-event data. The company chose to use the survival functions preferred by the appraisal committee after the first committee meeting, which were lognormal functions fitted to TFT and BSC separately for OS, generalised gamma functions fitted to TFT and BSC separately for PFS and a generalised gamma function fitted to TTD for TFT as '*there is relatively limited evidence to overtly reject the previously selected base-case analysis models*'. The base case incremental cost-effectiveness ratio (ICER) within the company response document was £45,662 per quality-adjusted life year (QALY) gained when one of the five characteristics was region. When one of the five characteristics was ethnicity, the ICER decreased slightly to £45,611 per QALY gained; this indicates that the results may not be sensitive to the choice of weighting by region or ethnicity for one of the characteristics. Henceforth, all ICERs are reported in terms of cost per QALY gained.

### 2.4 The results from the new analyses

The results of the company's new analyses are shown in Table 1. These analyses assumed all patients would receive TFT as a third-line treatment and covered: (1) the ACD base case, (2) naïve (unweighted) analyses and (3) the weighted results and for those in the EU only. Separate cost-effectiveness analyses have been conducted by the company for each of these analyses based on independent and dependent survival models fitted to the data from each treatment arm.

**Table 1: Cost-effectiveness results produced by the company**

| Scenario | Inc. cost | Inc. QALYs | ICER | Inc. survival (months) | % increase in survival (TFT versus BSC) |
|---|---|---|---|---|---|
| ACD base case (independent models) | £6,590 | 0.099 | £66,523 | 1.68 | 26 |
| Naïve Analyses – independent models | | | | | |
| Naïve 3L | £7,540 | 0.179 | £42,086 | 3.21 | 50 |
| Naïve 3L EU | £7,853 | 0.172 | £45,681 | 3.05 | 47 |
| Naïve Analyses – dependent models | | | | | |
| Naïve 3L | £7,556 | 0.204 | £37,058 | 3.71 | 60 |
| Naïve 3L EU | £7,873 | 0.202 | £39,013 | 3.66 | 59 |
| Weighted Analyses – independent models | | | | | |
| Weighted 3L (region included) † | £7,481 | 0.164 | £45,662 | 2.89 | 44 |
| Weighted 3L (ethnicity included) | £7,465 | 0.164 | £45,611 | 2.89 | 45 |
| Weighted 3L EU | £7,777 | 0.156 | £49,771 | 2.72 | 41 |
| Weighted Analyses – independent models | | | | | |
| Weighted 3L (region included) | £7,497 | 0.188 | £39,925 | 3.37 | 53 |
| Weighted 3L (ethnicity included) | £7,478 | 0.184 | £40,724 | 3.29 | 52 |
| Weighted 3L EU | £7,796 | 0.186 | £41,858 | 3.33 | 52 |

† It is believed that this scenario represents the company's updated base case.

### 2.5 NICE's end of life criteria

In the discussion section of the company's response,[2] it is stated that in the company's updated base case the absolute survival gain is predicted to be 2.89 months which is approximately 3.5 days under the 3 month life-extension criterion. The company further state that "*While beneath the 3-month value, treatment with T/T offers a survival improvement of 2.89 months, equivalent to a relative survival gain of 44.51% compared to BSC. A benefit of this magnitude is extremely important within the context of this patient population, who have a very poor prognosis.*" In its original submission,[5] the company highlighted a precedent (TA476[6]) where NICE accepted that an intervention met the end of life criteria despite the life-extension being less than 3 months; this flexibility was due to the very poor prognosis for patients with metastatic pancreatic cancer.

The criterion associated with short life expectancy has been accepted by the NICE appraisal committee within the ACD.[1]

# 3 ERG critique of the company's response to the ACD

The subsections in this section correspond to those in Section 2.

## 3.1 A teleconference between NICE and the company

As the ERG was not involved in this teleconference it cannot comment on specific details. The rationale for selecting the five characteristics deemed '*potentially important*' and not others has not been made explicit in the company's new analyses. As stated in the initial ERG report[7] '*There were mixed views from clinical advisors to the ERG and NICE about whether prior ramucirumab treatment would alter prognosis*' exploring whether adjusting for prior ramucirumab use may have been beneficial. The ERG comments that the TAGS RCT[3] was stratified based on prior ramucirumab use and the company chose the no prior ramucirumab group for the base case analysis in its original submission.[5] This is discussed further in Section 3.2.

## 3.2 The weighting analyses undertaken to account for imbalances in potential prognostic factors between the TFT and BSC arms

A key feature of the propensity score approach is the assumption that there are no unmeasured confounders and that all measured potential confounders have been included in the propensity score model. As originally derived, the propensity score model "*may have been perceived as advocating the use of covariates merely related to assignment, but the advice to include all covariates predictive of outcome in the construction of empirical propensity scores has been around [a long time]*" (Senn, 2007[8]).

The company, in discussion with NICE, included five potential confounders These were: peritoneal metastases (absence of presence); ECOG PS (0 vs 1); histology (intestinal vs non-intestinal); ethnicity or region (residing in Japan vs the rest of the world (ROW) or Asian vs Non-Asian); prior irinotecan (previous exposure vs no previous exposure. It is unclear whether some measured confounders have been omitted with other potential confounders defined in the original company submission[5] including: age (<65 years vs ≥65 years); sex (male vs female); previous taxane therapy (yes vs no); previous gastrectomy (yes vs no); gastro-oesophageal junction cancer involvement (yes vs no); peritoneal, liver, or lung metastases (yes vs no); number of metastatic sites (one or two vs three or more); measurable disease (yes vs no); and Human Epidermal Growth Factor Receptor 2 status (negative vs positive or not assessed). Consequently, it is not clear whether the propensity score model has adjusted for all measured confounders and it is noted that the number of covariates that were included in the propensity score model were limited by the decision to restrict the population to those receiving third-line treatment. The ERG notes that the company provided an insufficient discussion on the issues of bias and precision

associated with the application of the propensity score model in the context of non-linear models, including supporting references in this context.

The ERG prefers an approach that would involve conducting a regression analysis of the complete dataset with relevant prognostic factors and the interaction between treatment and lines of prior treatment explicitly included in the model. The impact and relevance of the covariates could then be assessed by looking at their coefficients and standard errors, noting that for a non-linear model the effect of covariates is on the coefficients and standard errors, and may not lead to an increase in precision. The ERG believes it is not possible to quantify the extent of the bias or the impact on the standard errors.

### 3.3 Survival modelling using the weighted survival functions

As stated in Section 3.2, the ERG is uncertain whether the propensity score model has adjusted for all measured confounders. Furthermore, it would prefer a regression-based analysis that predicts response to treatment rather than allocation to treatment. Additionally, the extent of any bias in the ICERs is unknown. The following text is based on the presumption that the propensity score model is appropriate.

The ERG would have preferred for additional analyses to be provided by the company exploring the impact of using the survival functions that fit the reweighted time-to-event data best based in BIC, but are reasonably satisfied by the survival functions used in the base case.

For OS, the best-fitting function, based on summed BIC values from independent models, was the generalised gamma; however, when using this model those receiving BSC were, on average, predicted to survive longer than those receiving TFT, due to the long tail associated with the OS of BSC. These results were deemed implausible by clinicians at previous appraisal committees. The next best-fitting function was the lognormal selected in the original company base case which had a BIC that was at least five points lower than the log-logistic model, with other candidate models having a much worse fit. Use of a log-logistic model reduced the base case ICER from £45,662 to £42,195.

Alternative models used for PFS did not materially impact the ICER with the use of independent log-logistic models, which has better combined BIC values than the chosen generalised gamma model reducing the base case ICER from £45,662 to £45,100.

In terms of TTD, the exponential and Weibull models had lower BIC values than the chosen generalised gamma model although use of these survival functions only had a small impact on the base case ICER with values of £45,700 and £45,257, respectively. The lognormal model, however, could not be ruled out as its BIC value was only 2.24 more than the generalised gamma, increased the company base case

to £48,336. The ERG comments that the generalised gamma distribution requires three parameters and that more parsimonious models such as the exponential may be preferred when the difference in BIC values are less two.

The company also present an analysis where only those patients receiving third-line treatment who are in the EU are included and where neither region nor ethnicity was included as a characteristic for adjustment. This analysis may be pertinent as in the ACD[1] it was stated that '*The committee concluded that the TAGS subgroup of people from Europe who had not had ramucirumab was the most relevant to the NHS in England*.' The ERG notes that no adjustment according to ramucirumab use has been provided by the company, although it appears that this was not requested by NICE.

When considering only those patients in the EU receiving third-line treatment the company's base case ICERs increased to £49,771. For information, further assuming a lognormal model for TTD increases this ICER to £52,902; however, this was the fifth best fitting model, with the best-fitting model, the exponential, increasing the ICER to £49,866.

### 3.4 The results from the company's new analyses

The ERG comments that uncertainty associated with the propensity score model cannot be assessed. The ERG neither had the data nor the time to explore the impact of alternative methods for adjusting for the imbalances between treatment arms.

Assuming that the propensity score model is appropriate, the company did not report the impact of using alternative survival models. However, the ERG undertook these analyses with the results reported in Section 3.4; only the use of the lognormal distribution for TTD substantially increased the ICER. Based on clinical advice provided to the ERG in writing its initial report, the ERG prefers the European population to the full trial population, with the analyses presented by the company indicating that the ICER is marginally lower than £50,000 (an ICER of £49,771 when using a generalised gamma for TTD).

### 3.4 NICE end of life criteria

The ERG comments that the results presented by the company can be reproduced although there is inevitable residual uncertainty in the extension to life related to the propensity score model that may have omitted relevant confounders. The ERG believes that whether or not TFT meets the extension to life criterion is a decision for the appraisal committee depending on its views about the most relevant analysis, the likely impact of potentially relevant prognostic factors which have not been included in the propensity score model and the flexibility it believes should be given to the extension of life criterion in patients with poor prognoses. The criterion related to short life-expectancy appears to be met.

# 4 Discussion

The ERG was able to reproduce the results reported in the company's response. However, the ERG highlights two potential factors that cause uncertainty in the ICER.

The first is whether there are any unmeasured confounders in the propensity score model and the extent of residual uncertainty. The ERG would prefer to see the results of a regression-based analysis that includes all measured prognostic factors and the interaction between treatment and lines of prior treatment. The impact of this alternative approach to adjust for imbalances between the TFT and BSC arms is not known.

The second factor relates to the population chosen for the analysis and the models used to fit to the time to event data. Use of a European population increases the company's base case ICER from £45,662 to £49,771. Further, the lognormal model for TTD could not be ruled out based on BIC scores and when combined with the European population increases the ICER to over £52,000, however, the lognormal model was the fifth-best model to TTD.

# 5 REFERENCES


1. National Institute for Health and Care Excellence. Appraisal consultation document. Trifluridine–tipiracil for treating metastatic gastric cancer or gastro-oesophageal junction adenocarcinoma after 2 or more therapies. 2019.
2. Servier Laboratories. Additional weighting analysis based on the third-line population enrolled within the TAGS clinical trial. 2020.
3. 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.
4. Stevenson M, Harnan S, Metry A, Sripadam R. Trifluridine-tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies: A Single Technology Appraisal. The ERG's critique of the company's response to the ACD.: School of Health and Related Research (ScHARR); 2020.
5. Servier Laboraties. Trifluridine–tipiracil for treating metastatic gastric or gastro-oesophageal junction cancer after 2 or more therapies [ID1507]. Company evidence submission. 2019.
6. 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).
7. Stevenson M, Harnan S, Ren S, Metry A, Orr M, Campbell F, *et al.* 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.
8. Senn S, Graf E, Caputo A. Stratification for the propensity score compared with linear regression techniques to assess the effect of treatment or exposure. *Statistics in Medicine* 2007;26:5529-44.