NICE National Institute for Health and Care Excellence

Single Technology Appraisal

Nivolumab for previously treated unresectable advanced oesophageal cancer [ID1249]

Committee Papers

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

SINGLE TECHNOLOGY APPRAISAL

Nivolumab for previously treated unresectable advanced oesophageal cancer [ID12149]

Contents:

The following documents are made available to consultees and commentators:

The **final scope and final stakeholder list** are available on the NICE website.

1. **Company submission** from **Bristol-Myers Squibb**

2. **Clarification questions and company responses**

3. **Patient group, professional group and NHS organisation submission** from:
   a. The Royal College of Physicians on behalf of the NCRI-ACP-RCP-RCR

4. **Evidence Review Group report** prepared by the **Peninsula Technology Assessment Group (PenTAG)**
   the ERG report was updated after the factual accuracy check

5. **Evidence Review Group – factual accuracy check**

6. **Technical Report**

7. **Technical engagement response** from **Bristol-Myers Squibb**

8. **Technical engagement responses from experts:**
   a. Dr Elizabeth Smyth – clinical expert, nominated by the NCRI-ACP-RCP-RCR

9. **Evidence Review Group critique of company response to technical engagement** prepared by the **Peninsula Technology Assessment Group (PenTAG)**

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

NATIONAL INSTITUTE FOR HEALTH AND CARE EXELLENCE

Single technology appraisal

Nivolumab for unresectable, advanced oesophageal cancer when standard chemotherapy has failed

ID1249

Document B

Company evidence submission

April 2020

File name	Version	Contains confidential information	Date
		Yes/~~no~~	

Contents

Tables and figures

List of Tables

List of Figures

B.1 Decision problem, description of the technology and clinical care pathway

B.1.1. Decision problem

The submission covers the technology's full marketing authorisation for this indication.

The decision problem is presented in Table 1.

Table 1. The decision problem

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope
Population	Adults with previously treated advanced or recurrent unresectable oesophageal cancer that is refractory or intolerant to chemotherapy	Nivolumab as monotherapy for the treatment of adult patients with unresectable advanced, recurrent or metastatic oesophageal squamous cell carcinoma after prior fluoropyrimidine and platinum-based combination therapy	Not applicable
Intervention	Nivolumab	Nivolumab	As per NICE scope
Comparator(s)	• Chemotherapy including taxanes (docetaxel/paclitaxel) or irinotecan • Best supportive care (including, but not limited to antiemetics, blood transfusions, oesophageal stents)	• Chemotherapy including taxanes (docetaxel/paclitaxel) • Best supportive care (including, but not limited to antiemetics, blood transfusions, oesophageal stents)	The main treatment options in this setting are primarily palliative. However, the majority of patients in this setting will receive taxane monotherapy, based on market research and clinician opinion. Some patients are unable to receive chemotherapy and these patients will receive BSC. Clinicians felt strongly that irinotecan would not be used in the UK setting for treatment of second-line oesophageal squamous cell carcinoma. This view is supported by market research, where irinotecan comprises only 6% of current usage.
Outcomes	The outcome measures to be considered include: • overall survival • progression-free survival • response rate • adverse effects of treatment • health-related quality of life.	The outcome measures to be considered include: • overall survival • progression-free survival • response rate • adverse effects of treatment • health-related quality of life.	As per NICE scope

Economic analysis	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year. The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared. Costs will be considered from an NHS and Personal Social Services perspective.	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year. The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared. Costs will be considered from an NHS and Personal Social Services perspective.	As per NICE scope
Subgroups to be considered	No patient subgroups have been identified.	No patient subgroups have been identified.	As per NICE scope
Special considerations including issues related to equity or equality	No equality issues have been identified.	No equality issues have been identified.	As per NICE scope

B.1.2. Description of the technology being appraised

Details of the technology being appraised in this submission are summarised in Table 2. The Summary of Product Characteristics is attached as Appendix C.1.1. A European public assessment report describing nivolumab for the treatment of oesophageal cancer is not available at time of submission.

Table 2. Technology being appraised

UK approved name and brand name	Nivolumab (Opdivo®)
Mechanism of action	It has been demonstrated that the upregulation of PD-1 and its ligands in oesophageal cancer tissues is correlated with poor prognosis.[1] Through exploitation of the PD-1 immune checkpoint inhibitor pathway, oesophageal cancer cells are able to escape immune surveillance. Nivolumab is a fully human monoclonal antibody that acts as a PD-1 checkpoint-inhibitor that blocks the binding of PD-1 (expressed on effector T cells) with its ligands (PD-L1 and PD-L2, expressed on target cells such as tumour cells). By interrupting this interaction, nivolumab prevents tumour cell evasion from destruction and restores T cell activity. Hence, nivolumab stimulates the patient's own immune system to directly destroy cancer cells (in the same way that it would any other "foreign" cell), resulting in destruction of the tumour through pre-existing, intrinsic processes. Further details are provided in Section B.1.3.5.1
Marketing authorisation/CE mark status	A Marketing Authorisation Application (MAA) was submitted to the European Medicines Agency (EMA) in February 2020 and the product has been submitted for registration via the Centralised Procedure. The earliest point at which an opinion from the Committee for Medicinal Products for Human Use (CHMP) could be anticipated is [redacted]. If CHMP opinion is provided during July[redacted], then regulatory approval would be available during [redacted].

Indications and any restriction(s) as described in the summary of product characteristics (SmPC)	The proposed indication for nivolumab for the treatment of oesophageal cancer is as follows: [redacted]
Method of administration and dosage	Nivolumab will be administered intravenously over 30 minutes at 2-week intervals at a dose of 240mg. Treatment will be continued until disease progression.
Additional tests or investigations	Not applicable.
List price and average cost of a course of treatment	List price: 10mg/ml concentrate for solution for infusion in vial; 4 ml vial: £439.00; 10 ml vial £1,097.00; 24 ml vial £2,633.00[2] Average cost/dose: £2,633
Patient access scheme (if applicable)	A patient access scheme has been agreed with the Department of Health, comprising of a [redacted] to nivolumab costs.

B.1.3. Health condition and position of the technology in the treatment pathway

Summary

- Oesophageal cancer (OC) is a malignant tumour developing from the cells lining the oesophagus (Figure 1).[3]
- In the UK, OC is often diagnosed at a late stage (70-80% diagnosed at stage III or IV) and 37-42% of cases have metastases at the point of diagnosis.[4]
- Squamous cell carcinoma and adenocarcinoma are the two major histology types of OC and account for over 95% of the cases of OC.[5] However, there is notable global variation in the distribution of histological types of OC. In Western countries, such as the USA, the majority (64%) of the OC cases are adenocarcinomas, while less than a third (31%) are squamous cell carcinomas.[6]
- The prognosis for unresectable OC is poor. In England, less than half of patients (42%) remain alive at 12 months, regardless of stage at diagnosis.[7]
- Second-line palliative chemotherapy is recommended for patients who have progressed on the first-line therapy; however, specific chemotherapy regimens are not specified in the NICE clinical guidelines in the second-line setting.[8-10] Guidelines from the European Society for Medical Oncology (ESMO) recommend taxane monotherapy in this setting.[11]
- Nivolumab is a highly innovative therapy that has shown unprecedented single-agent activity in the treatment of advanced or recurrent unresectable oesophageal cancer, with a unique mechanism of action and published data describing the beneficial impact of therapy in terms of efficacy and safety.
- Nivolumab would represent an additional treatment option for patients who are refractory or intolerant to fluoropyrimidine and platinum-based combination therapy or patients considered not fit for chemotherapy.

B.1.3.1. Disease overview

Oesophageal cancer (OC) is a malignant tumour developing from cells lining the oesophagus (Figure 1).[3]

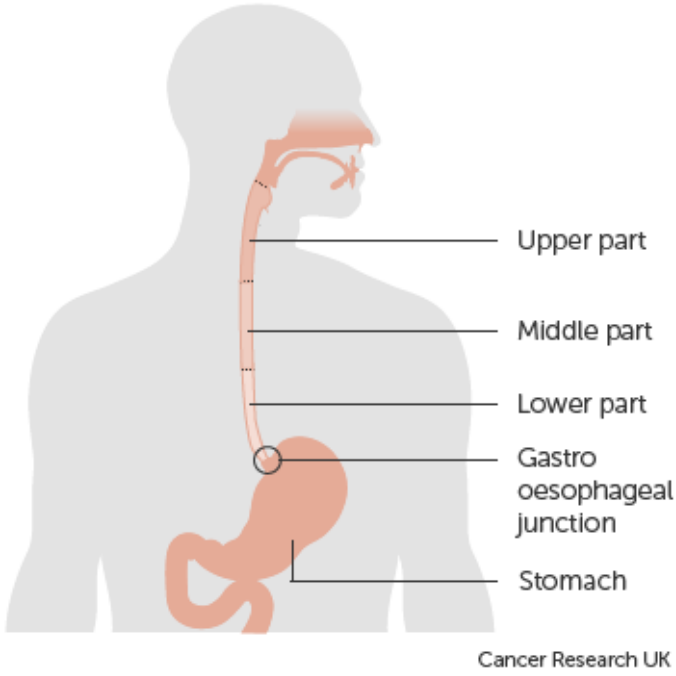


Figure 1. Oesophageal cancer locations[3]

Over 95% of oesophageal cancers present as oesophageal squamous cell carcinoma (OSCC) or adenocarcinoma,[5] which can be considered two epidemiologically and pathologically distinct diseases that share an anatomical site. OSCC develops from the squamous epithelial cells that make up the inner lining of the oesophagus and risk factors include recurrent chemical or physical insults to the oesophageal mucosa, such as tobacco smoking and alcohol consumption,[12] as outlined in Figure 2. By contrast, adenocarcinomas typically arise from Barrett oesophagus and risk factors include excess body weight and gastro-oesophageal reflux. [12, 13] Further, OSCC is more common in the upper and middle third of the oesophagus, while adenocarcinomas are more common the distal (lower) section of the oesophagus.[13]

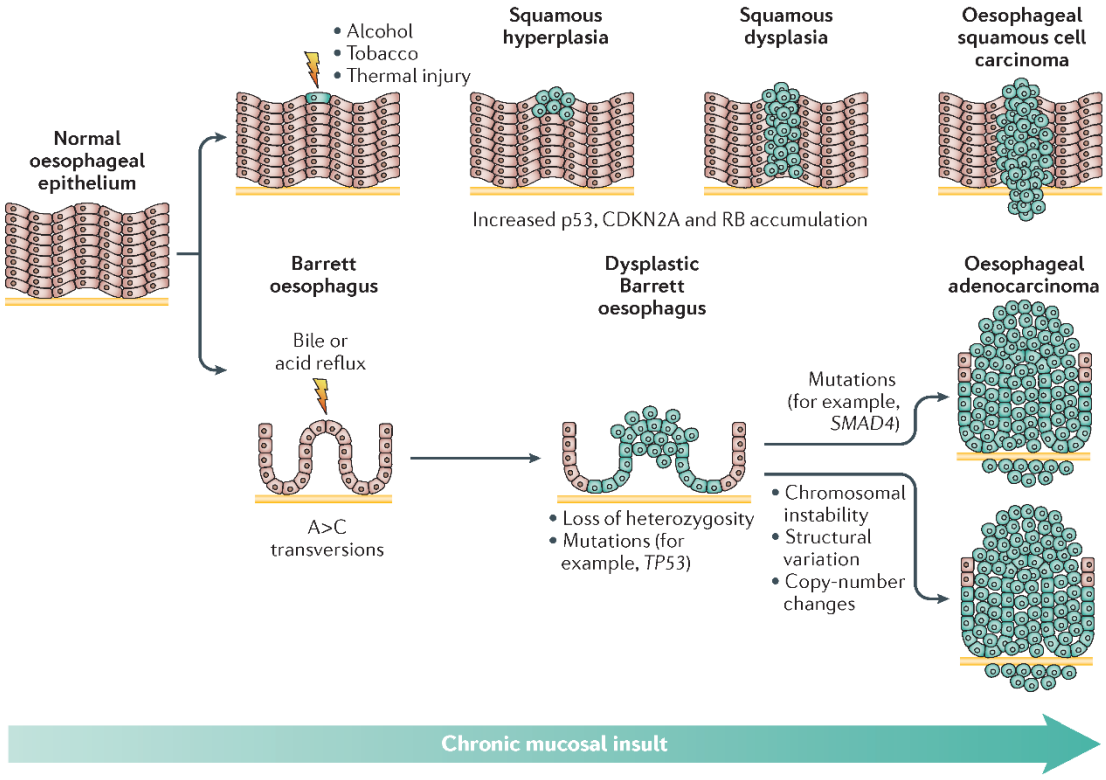


Figure 2 | **The pathogenesis of OSCC and OAC.** The oesophageal mucosa is exposed to repeated insults, which result in changes to the squamous oesophageal mucosa. Molecular changes also accumulate, and this ultimately leads to a malignant phenotype. In oesophageal squamous cell carcinoma (OSCC), squamous hyperplasia precedes low-grade and high-grade squamous dysplasia, which then develops into invasive cancer. In oesophageal adenocarcinoma (OAC), a metaplastic epithelium (Barrett oesophagus) is transformed through low-grade and high-grade dysplasia to invasive cancer. CDKN2A, cyclin-dependent kinase inhibitor 2A; RB, retinoblastoma-associated protein.

Figure 2. Pathogenesis of OC (reproduced from Figure 2 of Smyth et al., 2017[13])

Staging is based on the widely accepted TNM staging system developed by the American Joint Committee on Cancer (AJCC), which classifies according to the amount of tumour invasion (T), involvement of lymph nodes (N), and distant metastasis (M), as outlined in Figure 3 and Table 3.[14] Tumours may be classified by pathological stage (following surgery) or clinical stage (using a physical exam, biopsy and imaging). Patients with cT3-T4 or cN1-3 disease are classed as having locally advanced disease, while those with metastatic disease are classified as advanced or stage IV disease.[11]

Table 3. Clinical (cTNM) staging of oesophageal squamous cell carcinoma

Stage	T category	N category	M category
0	Tis: High-grade dysplasia, defined as malignant cells confined by the basement membrane	N0: No regional lymph node metastasis	M0: No distant metastasis
I	T1: Tumour invades the lamina propria, muscularis mucosae, or submucosa	N0: No regional lymph node metastasis N1: Metastasis in 1–2 regional lymph nodes	M0: No distant metastasis
II	T2: Tumour invades the muscularis propria	N0: No regional lymph node metastasis N1: Metastasis in 1–2 regional lymph nodes	M0: No distant metastasis
	T3: Tumour invades adventitia	N0: No regional lymph node metastasis	
III	T3: Tumour invades adventitia	N1: Metastasis in 1–2 regional lymph nodes	M0: No distant metastasis
	T1: Tumour invades the lamina propria, muscularis mucosae, or submucosa T2: Tumour invades the muscularis propria T3: Tumour invades adventitia	N2: Metastasis in 3–6 regional lymph nodes	
IVA	T4: Tumour invades adjacent structures	N0: No regional lymph node metastasis N1: Metastasis in 1–2 regional lymph nodes N2: Metastasis in 3–6 regional lymph nodes	M0: No distant metastasis
	T1: Tumour invades the lamina propria, muscularis mucosae, or submucosa T2: Tumour invades the muscularis propria T3: Tumour invades adventitia T4: Tumour invades adjacent structures	N3: Metastasis in 7 or more regional lymph nodes	
IVB	T1: Tumour invades the lamina propria, muscularis mucosae, or submucosa T2: Tumour invades the muscularis propria T3: Tumour invades adventitia T4: Tumour invades adjacent structures	N0: No regional lymph node metastasis N1: Metastasis in 1–2 regional lymph nodes N2: Metastasis in 3–6 regional lymph nodes N3: Metastasis in 7 or more regional lymph nodes	M1: Distant metastasis

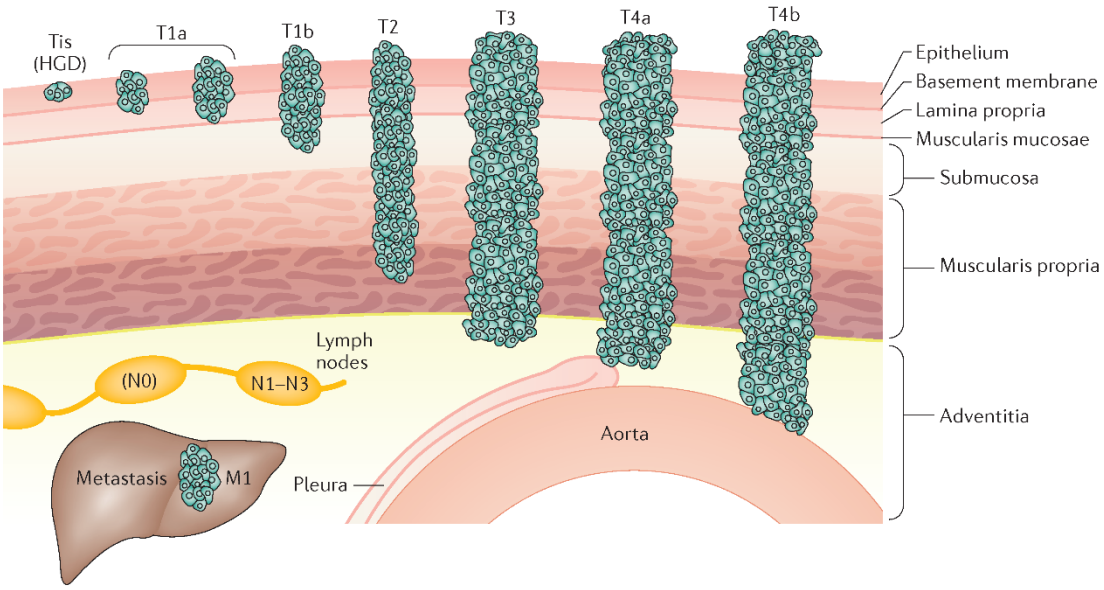


Figure | **Tumour–node–metastasis categories.** Tumour classification according to the categories. T refers to the size of the primary tumour and whether it invades the nascent tissue as shown. N refers to lymph node involvement: N0 describes no regional lymph node metastasis; N1 describes regional lymph node metastases involving one or two nodes; N2 describes regional lymph node metastases involving from three to six nodes; and N3 describes regional lymph node metastases involving seven or more nodes. M refers to distant metastasis and is categorized as M0 (no distant metastasis) or M1 (distant metastasis). HGD, high-grade dysplasia; Tis, cancer *in situ*. Adapted with permission from REF. 247, Elsevier.

Figure 3. TNM staging in oesophageal cancer (reproduced from Figure 4 of Smyth et al., 2017[13])

The most common sites of metastases include liver, distant lymph nodes, lung, bone and brain, with lung metastases more common in patients with OSCC and liver, bone and brain more common in patients with adenocarcinoma. [15-17] Further, survival in patients with metastases varies by metastatic site and histological subtype; in OSCC, lymph node metastases were associated with improved survival.[15, 17]

B.1.3.2. Epidemiology

OC is a significant health issue worldwide. Although this cancer is relatively rare (9,209 new diagnoses in UK in 2017, of which 7,569 cases were in England), it is the seventh most common cause of cancer death in the UK and was responsible for an estimated 7,925 deaths in UK in 2017.[4, 18] This reflects the fact that survival rates are extremely poor: only around 15% of people diagnosed with oesophageal cancer survive for 5 years or more.[4, 19]

The majority of global OC cases present as OSCC; in Western countries, such as the USA, the majority (64%) of the OC cases are adenocarcinomas, while less than a third (31%) are squamous cell carcinomas.[6] Similarly, there are significant differences in OC patient gender by histological

subtype; while the male to female incidence rate ratio for adenocarcinoma is higher (around 52:10), it is more balanced for OSCC (around 11:10).[4]

In the UK, OC is often diagnosed at a late stage (70-80% diagnosed at stage III or IV) and 37-42% of cases have metastases at the point of diagnosis.[4]

B.1.3.3. Life expectancy

The prognosis for unresectable OC is poor. In England, less than half of patients (42%) remain alive at 12 months, regardless of stage at diagnosis.[7] Patients with unresectable, advanced OC have worse outcomes than those diagnosed with localised disease. In OC patients diagnosed with regional and distant disease, five-year survival is 25% and 5%, respectively, and median survival in patients diagnosed with metastatic OC is 10 months.[6, 20] Further, survival is impacted by histological type, with adenocarcinoma patients experiencing reduced life expectancy versus OSCC patients.[15, 17]

In the context of treatment-experienced patients with unresectable OSCC, survival outcomes remain poor. Patients receiving therapy with taxane monotherapy achieve median survival of 5-9 months, while for those receiving best supportive care (BSC), median survival is only around four months.[21, 22]Thus, there is substantial unmet need for effective therapies to improve outcomes in this patient population.[23, 24]

B.1.3.4. Current pathway of care

The stage of the disease is a critical factor for treatment decisions. Patients diagnosed in the early stages of OC are most commonly treated by surgery, which is potentially curative; other treatments, including chemotherapy and radiotherapy, may also be appropriate depending on the extent of disease and the patient's fitness.[19] However, most patients in the UK are diagnosed at an advanced disease stage (70-80% diagnosed at stage III or IV), by which time surgery may no longer be a viable treatment option.[4, 12] In these patients, chemotherapy or radiation can improve symptoms and may prolong survival, but these treatments are not curative. Thus, the aim of treatment for patients with unresectable disease is primarily palliative: to prolong the time to progression, extend survival, and relieve symptoms to improve or maintain quality of life.[11]

A summary of the National Institute for Health and Care Excellence (NICE) guidelines for the treatment of OC is shown in Figure 4 and described below:

- Surgery is considered the treatment of choice for OC patients with localised disease. The pathway for the management of OC from NICE describes interventional procedures comprising oesophagectomy and lymph node dissection, accompanied by neoadjuvant or adjuvant therapies (such as chemotherapy and chemoradiotherapy) for OC patients suitable for radical treatment.[8-10]

- For patients with locally advanced or metastatic disease, palliative chemotherapy with doublet (5-fluorouracil or capecitabine in combination with cisplatin or oxaliplatin) or triplet (5-fluorouracil or capecitabine in combination with cisplatin or oxaliplatin plus epirubicin) regimens is recommended for first-line systemic treatment.[8-10]
- Second-line palliative chemotherapy is recommended for patients who have progressed on the first-line therapy; however, specific chemotherapy regimens are not defined in the NICE clinical guidelines in the second-line setting.[8-10] The guidelines do not differentiate between the two main histology types of OC: OSCC and adenocarcinoma.

Similar to UK guidance, guidelines from the European Society for Medical Oncology (ESMO) recommend palliative chemotherapy in the management of advanced or metastatic OC.[11] Taxane monotherapy is indicated for the second-line treatment (after failure of first-line treatment with taxane combination therapy) of OC, particularly in patients with adenocarcinoma with a good performance status.[11] Palliative monotherapy is also recommended for patients with OSCC; however, due to a lack of evidence of effectiveness, specific chemotherapy regimens are not specified.[11]

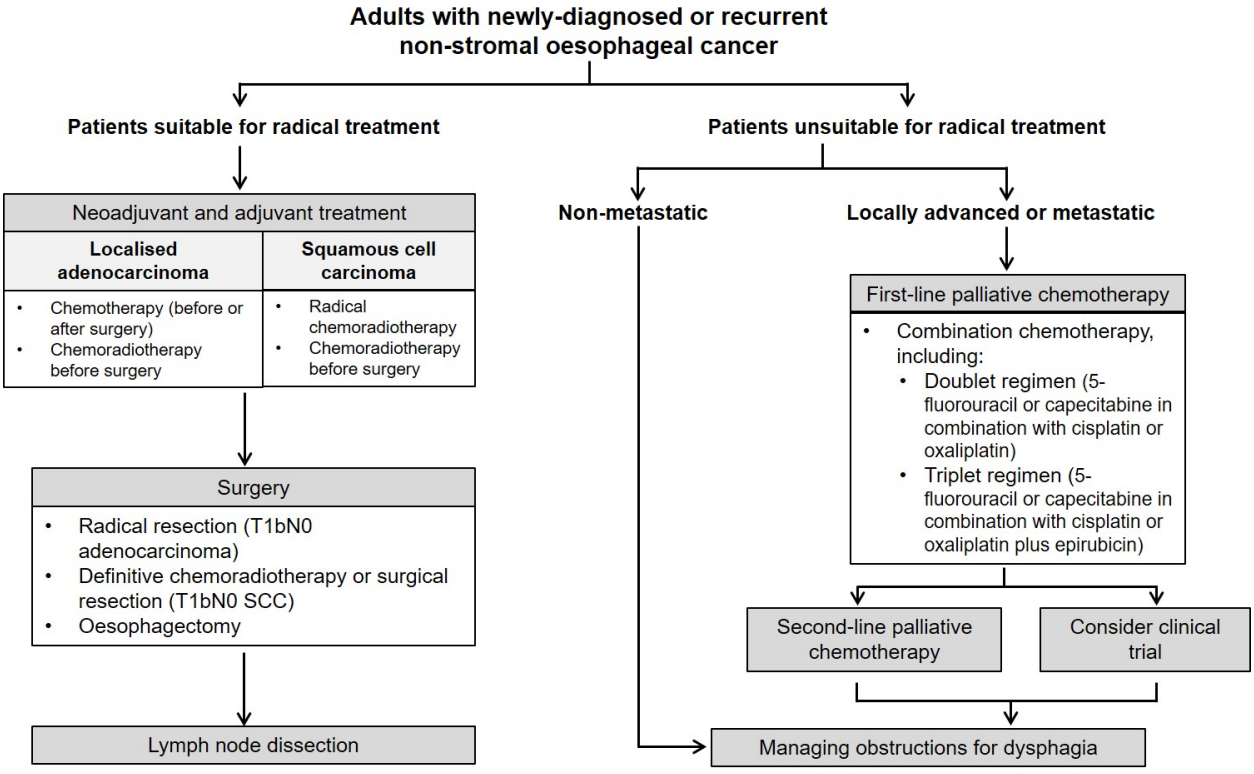


Figure 4. Treatment pathway for oesophageal cancer in UK (derived from NICE NG83)[8]

B.1.3.5. Nivolumab in the treatment of unresectable oesophageal cancer

B.1.3.5.1. *Mechanism of action*

Immunotherapy has been at the forefront of therapeutic development in oncology since the discovery that cancer cells evade destruction by exploiting the signalling pathways that control the immune system. The typical immune response to foreign cells or antigens in the body is the activation of T-cells that can then destroy those foreign cells or antigens. T-cells proliferate and differentiate through various pathways, with T-cell activation regulated through a complex balance of positive and negative signals provided by co-stimulatory and co-inhibitory receptor interactions on the T-cell surface (Figure 5). Healthy, non-foreign cells ('self'-cells) avoid T-cell destruction by stimulating inhibitory receptors, known as checkpoints, to suppress the T-cell response; cancer cells can use these same inhibitory receptors to escape destruction by T-cell activity. Antibodies designed to bind to and block these checkpoints (so called 'checkpoint-inhibitors') can prevent tumour-driven T-cell suppression, as depicted in Figure 5, and increase immune activity against cancer cells.

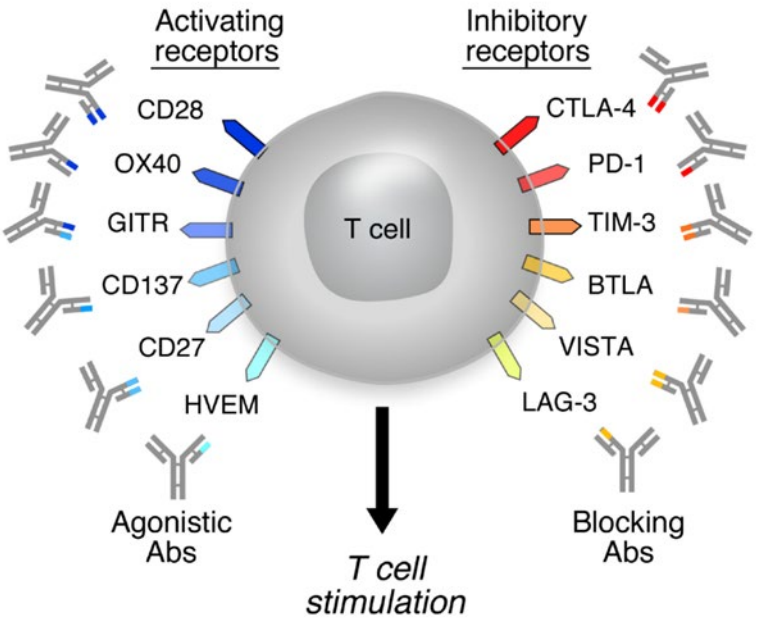


Figure 5. Receptors involved in the regulation of the T-cell immune response (from Mellman 2011[25])

PD-1 is an immune checkpoint protein receptor expressed at high levels on activated T-cells, which has been shown to control the inhibition of T-cell response at the effector stage of the immune response, in the setting of human malignancy.[26-30] Tumour cells can exploit this pathway by up-regulating proteins that engage PD-1 with its ligands (programmed death ligand-1 [PD-L1] and programmed death ligand-2 [PD-L2]) to limit the activity of T-cells at the tumour site.

Expression of PD-L1 and PD-L2 in tumour cells has been detected in approximately 44% of oesophageal cancer patients.[1] Further, patients with PD-L1 and PD-L2 expression are reported have a poorer prognosis compared with those without expression of these ligands.[1] Through exploitation of the PD-1 immune checkpoint inhibitor pathway, oesophageal cancer cells are able to escape immune surveillance (Figure 6). Hence, PD-1 and its ligands may be considered as therapeutic targets for immune-mediated therapies in oesophageal cancer.

Nivolumab is a fully human, monoclonal immunoglobulin G4 antibody (IgG4 HuMAb) that acts as a PD-1 checkpoint-inhibitor, blocking the interaction of PD-1 with PD-L1 and PD-L2 (Figure 6 and Figure 7). Through interruption of PD-1 binding to PD-L1 and PD-L2, nivolumab stops the evasion of immune-mediated tumour destruction and restores T-cell activity by stimulating the patient's own immune system to directly destroy cancer cells (in the same way that it would any other "foreign" cell), resulting in destruction of the tumour through pre-existing, intrinsic processes (Figure 7).

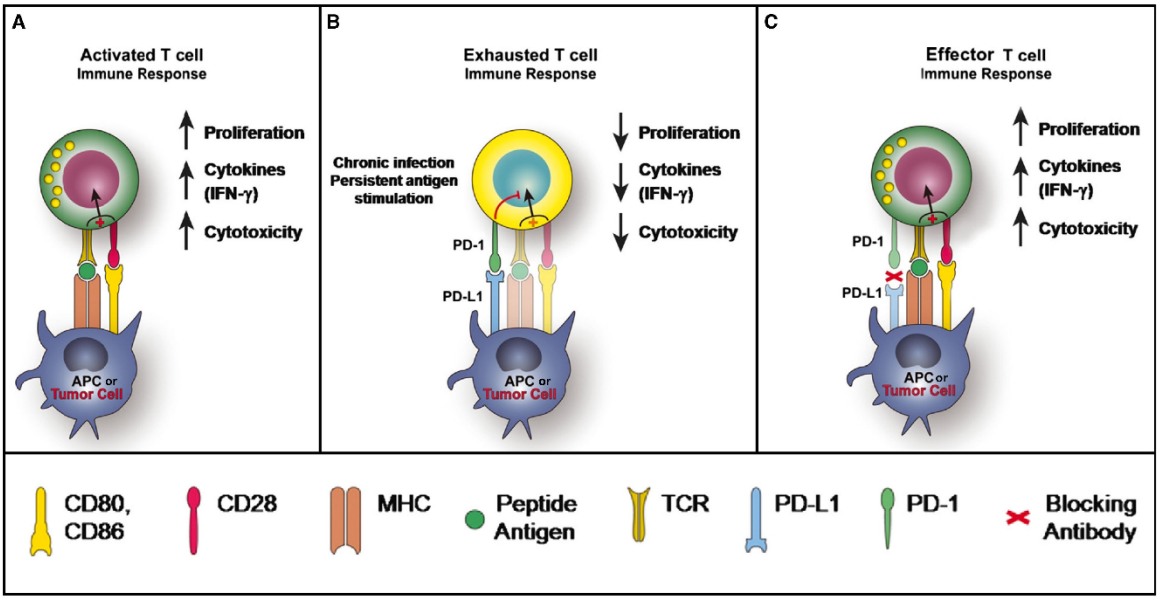


Figure 1. PD-1 in T-cell activation, exhaustion, and effector function. (A) T cells are activated via (1) binding of MHC plus peptide on an APC to the TCR and then (2) binding of APC CD80/86 to T-cell CD28. In patients with cancer, tumor cells can also serve as APCs. Upon T-cell activation, PD-1 expression is induced. (B) In situations of chronic infection or persistent stimulation, PD-L1 signals through T-cell PD-1 to "turn off" T cells in order to minimize damage to healthy tissue. Tumor cells can upregulate PD-L1 in order to "turn off" T cells that might destroy them. (C) Blocking the PD-1/PD-L1 signaling pathway allows T cells to maintain their effector functions. In patients with cancer, activated tumor-specific T cells can kill tumor cells and secrete cytokines that activate/recruit other immune cells to participate in the antitumor response. APC, antigen-presenting cell; IFN-γ, interferon gamma; MHC, major histocompatibility complex; PD-1, programmed death-1; PD-L1, PD ligand 1; TCR, T-cell receptor.

Figure 6. PD-1 pathway and blockade (from McDermott and Atkins 2013[31])

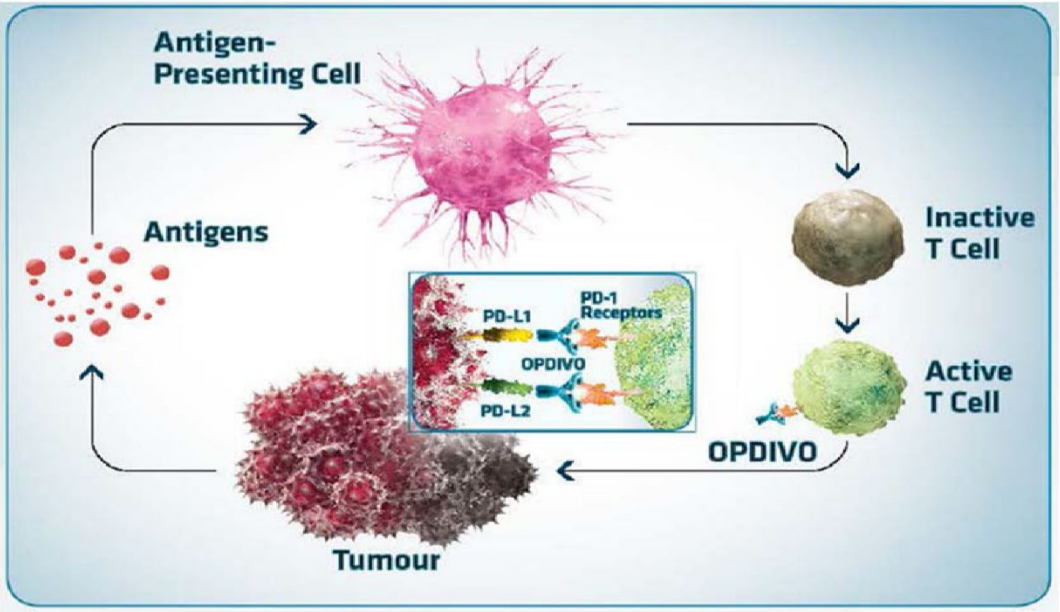


Figure 7. Nivolumab stimulation of immune-mediation destruction

Pseudoprogression in response to checkpoint inhibitor therapy

Conventional anti-cancer therapies typically aim to reduce the tumour burden through direct disruption of tumour cell proliferation or induction of apoptosis. In contrast, there are key differences with immunotherapy agents such as nivolumab, as a result of their novel mechanism of action. One of these differences is the varying patterns of response that can be observed with immunotherapy agents, compared with chemotherapy.

- Due to the indirect anti-tumour mechanism associated with immunotherapies, where host immune cells are recruited to the tumour site, the initial effect of immunotherapy may present as growth of existing lesions or formation of new lesions that result from the infiltration of tumour-specific immune cells and other inflammatory cells ("pseudo-progression", Figure 8).[32-34] This brief initial enlargement of the tumour may be followed by tumour shrinkage or eradication.[32, 33]
- Hence, due to the delayed clinical responses observed in immunotherapies, the "time to response" from immunotherapy treatment may differ from that seen after conventional chemotherapy. [34]
- In addition, these differences in response patterns after immunotherapy may potentially be prematurely misclassified as disease progression under the WHO or RECIST criteria.[33, 34] For the same reasons, PFS may not be an adequate endpoint in immunotherapy trials and may not be considered a surrogate for OS for the achievement of clinical efficacy.

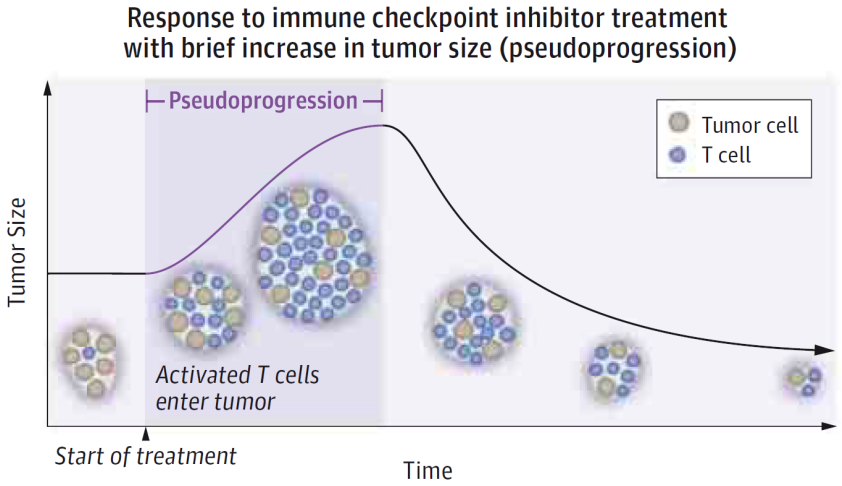


Figure 8. Pseudo-progression response to immune checkpoint inhibitor treatment[32]

B.1.3.5.2. *Nivolumab within the current clinical pathway*

In the second line setting, the treatment pathway is highly variable for unresectable, advanced or metastatic OC patients, as decisions on treatment options are individualised for each patient, as per clinical expert opinion obtained through advisory board. Currently, no treatment strategy is defined as best practice for patients who are refractory or intolerant to the standard therapies or who are considered to be unable to tolerate further chemotherapy. Chemotherapy agents such as paclitaxel or docetaxel may be used in this setting. However, these agents are associated with poor efficacy and tolerability, and patients often exhibit or develop resistance to them within a relatively short period. Patients who are not fit to receive further chemotherapy are left with best supportive care (BSC) and while BSC is able to relieve symptoms with minimal adverse effects, there is limited impact on progression or survival.

Nivolumab would represent an additional treatment option for patients with unresectable, advanced or metastatic oesophageal squamous-cell carcinoma who are either refractory or intolerant to fluoropyrimidine and platinum-based combination therapy or not eligible to receive further chemotherapy. The introduction of nivolumab would change the treatment paradigm for these patients and thus represents a “step-change” in the management of OC following failure of prior chemotherapy regimens.

B.1.4. Equality considerations

The incidence of oesophageal cancer is strongly correlated to age, where around 41% of new cases in the UK between 2014 to 2015 were diagnosed in those over 75 years old.[4] In addition, the five-year net survival of oesophageal cancer patients aged 70 years and over is notably poorer compared with younger patients, particularly in female patients (Table 4).

Palliative chemotherapy is recommended for patients who have progressed on the first-line therapy; however, specific chemotherapy regimens are currently not recommended in the NICE clinical guidelines in the second-line setting.[8-10] Guidelines from the European Society for Medical Oncology (ESMO) recommend taxane monotherapy in this setting (see below).[11] Therefore, patients currently have very limited treatment options in the second-line setting.

Treatment options for these older patients may be extremely limited due to their reduced ability to tolerate chemotherapy. Hence, older patients may be more likely to receive BSC than the overall cohort of patients with OC. As BSC has limited or no impact on symptoms, quality of life, progression and survival, this subgroup of older patients has a high unmet need, and an efficacious therapy that is well-tolerated would address this need. Nivolumab provides a treatment option with proven efficacy and tolerability, with the potential to impact on symptoms, progression and survival. Ageing well and tackling premature mortality is a priority for NHS England.[35] However, certain services and system rules in the UK are skewed in favour of the young, with far worse access and quality for older people in service.[36] Providing nivolumab will be in line with addressing the issue of ageism.

Table 4. Oesophageal cancer five-year net survival (2009-2013); by age group[4]

	Five-year Net Survival, Oesophageal Cancer England, all cancer stages	
Age group (years)	**Men**	**Women**
15–49	17.0%	23.5%
50–59	18.1%	27.7%
60–69	17.8%	22.2%
70–79	15.0%	16.5%
80–99	5.7%	4.2%

B.2 Clinical effectiveness

Key points

- Nivolumab therapy has significant benefits in terms of patient-relevant outcomes, including overall survival (OS), safety profile and quality-of-life endpoints.
- During ATTRACTION-3, nivolumab-treated patients achieved significantly improved survival over patients receiving taxanes (median OS: 10.91 months versus 8.38 months).
- At 12 months, 46.9% of patients in the nivolumab arm remained alive versus 34.4% in the taxane arm.
- The safety profile of nivolumab was improved over that for the taxanes: 65.6% of patients in the nivolumab arm reported a drug-related AE (grade 3-5: 18.2%) versus 95.2% for patients receiving paclitaxel or docetaxel (grade 3-5: 64.0%)
- Quality of life remained relatively stable in the nivolumab arm, as determined by EQ-5D and EQ-VAS; however, patients receiving taxanes frequently reported worsened quality of life outcomes during the trial period.
- Although ATTRACTION-1 was a single arm study, results were comparable to the nivolumab arm of ATTRACTION-3.

B.2.1. Identification and selection of relevant studies

A systematic literature review (SLR) was undertaken to identify the clinical effectiveness evidence (efficacy and safety) of interventions for the treatment of unresectable advanced oesophageal cancer where standard chemotherapy has failed. Full details of the process and methods to identify and select the relevant clinical evidence are summarised in Appendix D.

B.2.2. List of relevant clinical effectiveness evidence

Evidence to support the effectiveness of nivolumab for the treatment of advanced or recurrent unresectable squamous cell oesophageal cancer that is refractory or intolerant to chemotherapy, the indication described in the regulatory application, is derived primarily from ATTRACTION-3 (ONO-4538-24) and ATTRACTION-1 (ONO-4538-07), shown in Table 5 and Table 6, respectively.

Table 5. Clinical effectiveness evidence: ATTRACTION-3[37, 38]

Study	**ATTRACTION-3**				
Study design	Phase III, multicentre, randomised, docetaxel- or paclitaxel-controlled, open-label study				
Population	Adult patients with histologically confirmed unresectable advanced or recurrent oesophageal cancer, refractory or intolerant to combination therapy with fluoropyrimidine and platinum-based drugs.				
Intervention(s)	Nivolumab (referred to as ONO-4538) monotherapy, at a dose of 240mg administered intravenously over 30 minutes at 2-week intervals				
Comparator(s)	Docetaxel (intravenous administration of 75mg/m^2 every 3 weeks) or paclitaxel (intravenous administration of 100mg/m^2 weekly for 6 weeks followed by a 2-week drug holiday)				
Indicate if trial supports application for marketing authorisation	Yes	✓	Indicate if trial used in the economic model	Yes	✓
	No			No	
Rationale for use/non-use in the model	Source of direct comparative evidence evaluating the efficacy of nivolumab versus taxanes in the correct patient population				
Reported outcomes specified in the decision problem	• **Overall survival** • **Progression-free survival** • **Objective response rate** • **Adverse events and safety outcomes** • **Patient reported outcomes**				
All other reported outcomes	• Disease control rate • Duration of response • Time to response • Maximum percent change from baseline in the sum of diameters of the target lesion				
ECG: electrocardiogram; ECOG: Eastern Co-operative Oncology Group					

Table 6. Clinical effectiveness evidence: ATTRACTION-1[39]

Study	ATTRACTION-1				
Study design	Phase II, multicentre, open-label, uncontrolled single-arm study				
Population	Oesophageal cancer patients refractory or intolerant to standard therapies				
Intervention(s)	Nivolumab (referred to as ONO-4538) monotherapy, at a dose of 3mg/kg administered intravenously at 2-week intervals				
Comparator(s)	None.				
Indicate if trial supports application for marketing authorisation	Yes	✓	Indicate if trial used in the economic model	Yes	
	No			No	✓
Rationale for use/non-use in the model	Does not provide direct comparative evidence evaluating the efficacy of nivolumab versus taxanes				
Reported outcomes specified in the decision problem	• **Objective tumour response** • **Overall survival** • **Progression-free survival** • **Adverse events**				
All other reported outcomes	• Disease control rate and immune-related disease control rate • Time to progression • Duration of response • Time to response • Best overall response and immune-related best overall response • Change in tumour size • Effects on primary tumour in the oesophagus				
ECG: electrocardiogram; ECOG: Eastern Co-operative Oncology Group					

B.2.3. Summary of methodology of the relevant clinical effectiveness evidence

A summary of methodology for ATTRACTION-3 and ATTRACTION-1 is provided in Table 7. Baseline characteristics of patients enrolled in these studies are summarised in Table 8. Full details of design and methodology for each trial are provided in Section B.2.6, together with the trial results.

Table 7. Comparative summary of trial methodology

Trial number (acronym)	ATTRACTION-3[37, 38]	ATTRACTION-1[39, 40]
Location	USA, Europe (Denmark, Germany, Italy, UK), Japan, Korea, Taiwan	Japan
Trial design	Phase III, multicentre, randomised, docetaxel- or paclitaxel-controlled, open-label study	Phase II, multicentre, open-label, uncontrolled single-arm study
Eligibility criteria for participants	Adult patients with histologically confirmed unresectable advanced or recurrent oesophageal cancer, refractory to or intolerant to standard therapy, ECOG PS of 0 or 1, and life expectancy of at least three months	Adult patients with histologically proven oesophageal cancer that was refractory or intolerant to fluoropyrimidine-based, platinum-based, and taxane-based chemotherapy, ECPG PS of 0 or 1, life expectancy or at least 90 days
Settings and locations where data were collected	The study was conducted in 90 study locations across USA, Europe and Asia.	The study was conducted in eight academic centres and hospitals in Japan.
Trial drugs	Nivolumab (n = 195) Patients were administered nivolumab at 240mg every 2 weeks by intravenous infusion. Docetaxel or Paclitaxel (n = 195) Patients were administered docetaxel intravenously at 75mg/m^2 every 3 weeks or paclitaxel administered intravenously at 100mg/m^2 weekly for 6 weeks followed by a 2-week drug holiday.	Nivolumab (n = 65) Nivolumab (was administered at a dose of 3mg/kg at 2-week intervals by intravenous infusion.
Primary outcomes	OS, defined as the time from randomisation until death from any cause.	ORR, determined by central assessment.
Other outcomes used in the economic model/specified in the scope	ORR, PFS, Adverse events, Patient-reported outcomes (QoL)	ORR, assessed by investigator, OS, PFS, Adverse events
Pre-planned subgroups	Location (Japan vs rest of the world), No, of organs with metastases at randomisation (≤1 or ≥2), PD-L1 expression (≥1% vs <1% or indeterminate) Age Sex Race ECOG PS at baseline Lesion sites Histological classification Metastatic site Past treatments History of smoking	Histological classification ECOG PS at baseline Past treatments History of smoking History of alcohol consumption

ECOG PS: Eastern Corporative Oncology Group Performance Score; GOJ: gastro-oesophageal junction; ORR: objective response rate; OS: overall survival; PFS: progression-free survival, QoL: quality of life

Table 8. Characteristics of participants in the studies across treatment groups

Baseline characteristic		ATTRACTION-1[41]	ATTRACTION-3[38]			
		Nivolumab	Nivolumab	Control	Docetaxel	Paclitaxel
Cohort size (N)		65	210	209	65	144
Age	Median (range), years	62 (49-80)	64 (37 - 82)	67 (33 - 87)	[redacted]	[redacted]
	<65 years, n (%)		112 (53%)	85 (41%)	[redacted]	[redacted]
Sex, n (%)	Female	11 (17%)	31 (15%)	24 (12%)	[redacted]	[redacted]
	Male	54 (83%)	179 (85%)	185 (88.5)	[redacted]	[redacted]
Race, n (%)	White	0	9 (4%)	9 (4%)	[redacted]	[redacted]
	Black / African American	0	0	0	[redacted]	[redacted]
	Asian	65 (100%)	201 (96%)	200 (96%)	[redacted]	[redacted]
	Native Hawaiian or other Pacific Islander	0	0	0	[redacted]	[redacted]
Geographic location, n (%)	Japan	65 (100%)	136 (65%)	138 (66%)	[redacted]	[redacted]
	Rest of the world	0	74 (35%)	71 (34%)	[redacted]	[redacted]
ECOG PS, n (%)	0	29 (45%)	101 (48%)	107 (51%)	[redacted]	[redacted]
	1	36 (55%)	109 (52%)	102 (49%)	[redacted]	[redacted]
Histological type, n (%)	Squamous-cell carcinoma	65 (100%)	210 (100%)	209 (100%)	[redacted]	[redacted]
	Adenocarcinoma	0	0	0	[redacted]	[redacted]
	Other	0	0	0	[redacted]	[redacted]
Disease stage TNM classification*, n (%)	Stage I	9 (14%)	1 (0.5%)	0	[redacted]	[redacted]
	Stage II	11 (17%)	2 (1.0%)	5 (2.4%)	[redacted]	[redacted]
	Stage III	24 (37%)	8 (3.8%)	13 (6.2%)	[redacted]	[redacted]
	Stage IV	20 (31%)	172 (81.9%)	168 80.4%)	[redacted]	[redacted]
	Unknown / not evaluated	1 (1%)	27 (12.9%)	23 (11.0%)	[redacted]	[redacted]
Sites of metastases, n (%)	Lymph node	21 (32%)	159 (76%)	163 (78%)	[redacted]	[redacted]
	Peritoneum	24 (37%)	5 (2%)	11 (5%)	[redacted]	[redacted]
	Liver	43 (66%)	57 (27%)	54 (26%)	[redacted]	[redacted]
	Lung	1 (2%)	98 (47%)	92 (44%)	[redacted]	[redacted]
	Pleural tissue	4 (6%)	22 (11%)	13 (6%)	[redacted]	[redacted]
	Adrenal gland	2 (3%)	6 (3%)	7 (3%)	[redacted]	[redacted]
	Brain	1 (2%)	5 (2%)	1 (0.5%)	[redacted]	[redacted]
	Bone	21 (32%)	23 (11%)	25 (12%)	[redacted]	[redacted]
	Other	24 (37%)	1 (0.5%)	1 (0.5%)	[redacted]	[redacted]

ECOG PS: Eastern Corporative Oncology Group Performance Score; NR: not reported

B.2.4. Statistical analysis and definition of study groups in the relevant clinical effectiveness evidence

A summary of methodology for ATTRACTION-3 and ATTRACTION-1 is provided in Section B.2.6.

B.2.5. Quality assessment of the relevant clinical effectiveness evidence

There were no notable quality issues in relation to ATTRACTION-3 as well as for the single-arm nivolumab ATTRACTION-1. The complete quality assessment is available in Appendix D.

B.2.6. Clinical effectiveness results of the relevant trials

Evidence for the clinical efficacy of nivolumab is derived from the ATTRACTION-3 study, a Phase III active-controlled study, and ATTRACTION-1, a Phase II non-comparative study. The design, methodology and results for ATTRACTION-3 are described in Section B.2.6.1, followed by the design, methodology and results for ATTRACTION-1 in Section B.2.6.2.

B.2.6.1. ATTRACTION-3

B.2.6.1.1. *Study design*

The ATTRACTION-3 study is an ongoing Phase III, open-label, multi-centre, docetaxel/ paclitaxel-controlled, pivotal registration trial sponsored by Ono Pharmaceutical Co. Ltd to support the filing of nivolumab with the health authorities in Japan, United States and Europe (CA209-473, Clinical Trials Identifier NCT02569242).[42] The objective of the study was to evaluate the efficacy and safety of nivolumab in patients with OC refractory or intolerant to combination therapy with fluoropyrimidine and platinum-based drugs. The trial was initiated in December 2015 and was conducted in USA, Denmark, Germany, Italy, UK, Japan, Korea and Taiwan.[37]

Patients were randomised in a 1:1 ratio to treatment with nivolumab (240mg every 2 weeks IV) or to the control group (docetaxel or paclitaxel; see Section B.2.6.1.3 for regimens). Randomisation was stratified by region (Japan vs. the rest of the world), number of organs with metastases (≤1 vs. ≥2), and expression of PD-L1 (≥1% vs. <1% or intermediate).[37] The study design of ATTRACTION-3 is provided in Figure 9.

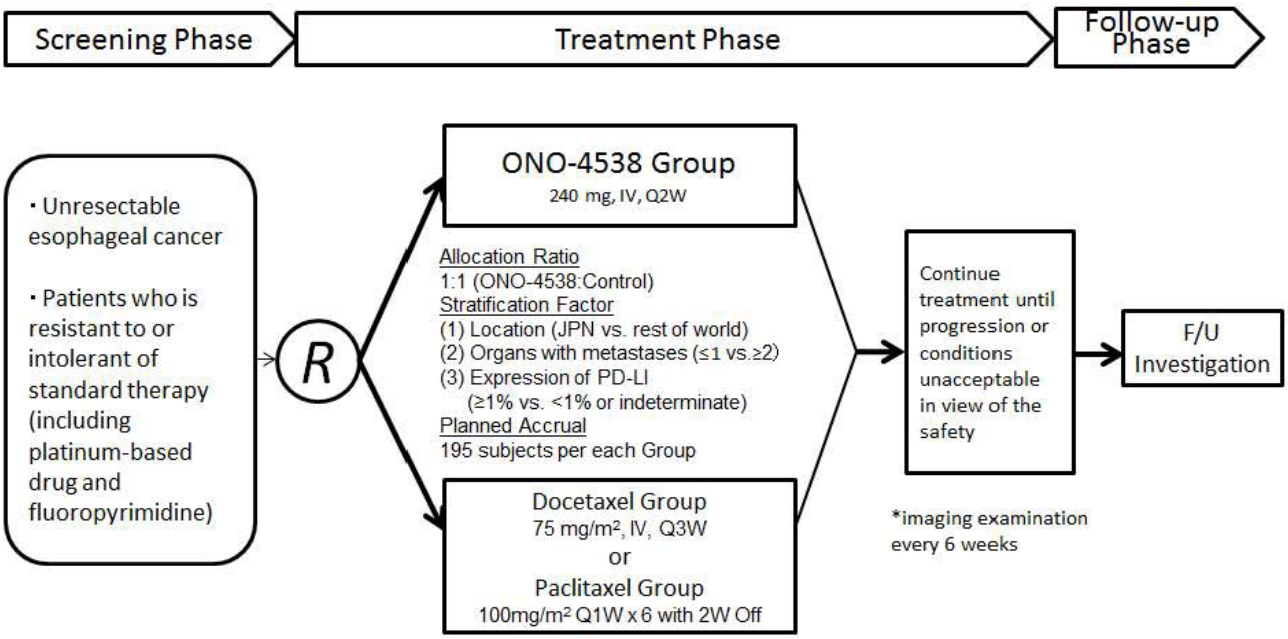


Figure 9. Study design of ATTRACTION-3[37]

Data presented in this submission is derived from published data based on a database lock on 12 November 2018.

B.2.6.1.2. *Eligibility criteria*

Patients with oesophageal cancer who are refractory or intolerant to combination therapy with fluoropyrimidine and platinum-based drugs were enrolled. The main eligibility criteria are listed in Table 9.[37]

Table 9. Inclusion and exclusion criteria for ATTRACTION-3[37]

Key inclusion criteria	Key exclusion criteria
• Men or women of at least 20 years of age • Oesophageal cancer with the major lesion in the oesophagus, histological type of major lesion was pathologically proven squamous cell carcinoma or adenosquamous cell carcinoma • Refractory or intolerant to combination therapy with fluoropyrimidine and platinum-based drugs for oesophageal cancer, previously received with one treatment regimen, and not indicated for a radical resection • At least one measurable or non-measurable lesion per the RECIST Guideline Version1.1 as confirmed by imaging within 28 days before randomisation • ECOG Performance Status score 0 or 1 • A life expectancy of at least 3 months.	• Significant malnutrition • Apparent tumour invasion on organs located adjacent to the oesophageal disease • History of, or current severe hypersensitivity to any other antibody products • Multiple primary cancers • Any metastasis in the brain or meninx that is symptomatic or requires treatment • Active, known or suspected autoimmune disease • Previously received taxane agents to treat oesophageal cancer; not proven refractory or intolerant to taxane-based combination therapy and subsequently received fluoropyrimidine and platinum-based combination therapy, and then proven refractory or intolerant may be randomised • Contraindicated to docetaxel and paclitaxel • Previously received nivolumab (ONO-4538 or MDX-1106 or BMS-936558), anti-PD-1 antibody, anti-PD-L1 antibody, anti-PD-L2 antibody, anti-CD137 antibody, anti-CTLA-4 antibody or other therapeutic antibodies or pharmacotherapies for regulation of T-cells.

B.2.6.1.3. *Study medications*

Patients will be randomized in a 1:1 ratio to the nivolumab group or control group (docetaxel or paclitaxel). After randomisation, the nivolumab group will receive nivolumab treatment (240 mg at 2-week intervals) and the control group will receive docetaxel (75 mg/m^2 at 3-week intervals) or paclitaxel (100 mg/m^2 weekly for 6 weeks in succession followed by a 2-week drug holiday). Treatment will be continued until progressive disease is assessed by the investigator or sub-investigator according to the RECIST Guideline Version 1.1, or due to pre-specified adverse events.[37]

B.2.6.1.3.1. Treatment beyond progression

Nivolumab, docetaxel, and paclitaxel treatment could be continued beyond disease progression if the following criteria is met, providing that continuing of the study treatment is expected to be beneficial for the patient.[37]

- No rapid disease progression and the continuation of study treatment is expected to lead to clinical benefits
- Treatment (nivolumab, docetaxel, or paclitaxel) was tolerated
- A stable ECOG Performance Status Score
- Continuation of study treatment will not cause a delay of any prophylactic intervention for serious complications associated with disease progression (e.g., brain metastasis).

B.2.6.1.4. *Study endpoints and assessments*

The primary, secondary, and exploratory endpoints of ATTRACTION-3 are provided in Table 10. Treatment will be assessed by the investigator and sub-investigator according to the RECIST Guideline Version 1.1.

Table 10. Study endpoints in ATTRACTION-3[37]

ATTRACTION-3 Study Outcomes	
Primary endpoint	• **Overall survival (OS)** o defined as the time from randomisation until death from any cause.
Secondary and exploratory endpoint	• **Objective response rate (ORR)** o defined as the percentage of patients whose best overall response is assessed as either CR or PR according to RECIST Guideline Version 1.1 • **Disease control rate (DCR)** o defined as the percentage of patients whose best overall response is assessed as CR, PR or SD according to RECIST Guideline Version 1.1 • **Progression free survival (PFS)** o calculated from the following equation: ("Time from date of randomisation until either the overall response was assessed as progressive disease or the patient died of any cause, whichever was the earlier"+1)/30.4375 • **Duration of response** o calculated from the following equation: ("Time from date of randomisation until either the overall response was assessed as progressive disease for the first time after confirmed response or the patient died of any cause"+1)/ 30.4375 • **Time to response** o calculated from the following equation: ("Time from randomisation until first assessment of confirmed CR or PR"+1)/30.4375 • **Best overall response (BOR)** o assessed as CR, PR or SD according to RECIST Guideline Version 1.1 • **Maximum percent change from baseline in the sum of diameters of the target lesion**
CR: complete response, disappearance of all (non-lymph node) target lesions. Any pathological lymph nodes (whether target or non-target) must have reduction in the short axis to <10 mm. PR: partial response, at least a 30% decrease in the sum of diameters of target lesions, taking as reference the baseline sum diameters. SD: stable disease, neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD, taking as reference the smallest sum diameter while on study. PD: progressive disease, at least a 20% increase in the sum of diameters of target lesions, taking as reference the smallest sum on study (this includes the baseline sum if that is the smallest on study). In addition to the relative increase of 20%, the sum must also demonstrate an absolute increase of at least 5 mm.	

B.2.6.1.5. *Statistical analyses*

The statistical analysis software East (version 6.3) was used for the calculation of the required events and sample size at the time of planning the study.[37]

The primary analysis set for efficacy endpoints was based on the intention-to-treat (ITT) population. Statistical significance will be assessed using a two-sided 5% significance level, except for testing of interactions, which will be performed using the two-sided 15% significance level. No statistical tests will be performed on safety.

The null hypothesis was that the nivolumab group is superior to the control group in terms of OS, which was calculated in days from date of randomisation till date of death from any cause. Data is compared between the two treatment groups using the stratified log-rank test with specified stratification factors. The hazard ratio and its 95% CI for the nivolumab group relative to control group is calculated using the stratified Cox proportional-hazards model. Median OS and its 95% CI is calculated using the Kaplan-Meier method, together with survival rates at 3, 6, 9, 12, 15, 18, 21, 24,27 and 30 months.[37,38]

B.2.6.1.6. *Sample size and power calculation*

The planned sample size comprises 195 patients in each treatment arm, totalling 390 patients. This study was intended to verify the superiority of the nivolumab group over the control groups (docetaxel or paclitaxel) in terms of OS (the primary endpoint).[37]

An exponential distribution was assumed for OS and the average hazard ratio of the nivolumab group vs. control group was assumed to be 0.70 (and a median OS of the control and nivolumab groups equivalent to 7.2 and 10.3 months, respectively), the number of events required to detect superiority of the nivolumab group over the control groups with two-sided significance level of 5% and 90% power by the log-rank test was calculated to be 331.[37]

Assuming the enrolment period to be 24 months and the follow-up period after the last patient's enrolment to be 15 months, the number of patients required to ensure the required 331 events was estimated to be 384. The target sample size was set at 390 in consideration of drop-outs.[37]

B.2.6.1.7. *Patient disposition*

A total of 590 patients were enrolled in the study and 419 patients were randomised to receive either nivolumab (210 patients) or docetaxel/paclitaxel (209 patients; docetaxel: 65 patients, paclitaxel: 144 patients). A total of 171 patients were not randomised due to patient's request to be withdrawn from study (11 patients), patient not fulfilling inclusion criteria (77 patients), patient meeting exclusion criteria (90 patients), failure to return to study site (1 patient) and for other reasons not specified (7 patients). 209 out of 2010 patients randomised to the nivolumab arm and 208 of 209 patients randomised to the control arm received at least one dose of the investigational product.[38] A summary of the patient disposition is provided in Table 11.

Table 11. ATTRACTION-3: Patient disposition[38]

	Nivolumab	Control group		
		Total	Docetaxel	Paclitaxel
Number of patients (intention-to-treat)	210	209	65	144
Number of treated patients	209	208	65	143
Continuation in the treatment period				
Patients continued in the treatment period	[redacted]	[redacted]	[redacted]	[redacted]
Patients discontinued in the treatment period	[redacted]	[redacted]	[redacted]	[redacted]
Reasons for discontinuation of the treatment period				
PD	[redacted]	[redacted]	[redacted]	[redacted]
Worsening of clinical symptoms due to PD	[redacted]	[redacted]	[redacted]	[redacted]
Onset of grade 2 or more adverse event				
Not received a dose of investigational product within past 6 weeks	[redacted]	[redacted]	[redacted]	[redacted]
Investigator judged to be inappropriate for other reasons	[redacted]	[redacted]	[redacted]	[redacted]
Other reasons	[redacted]	[redacted]	[redacted]	[redacted]
PD: progressive disease				

B.2.6.1.8. *Baseline patient characteristics*

A total of 419 patients were enrolled (Table 13). All patients had squamous-cell carcinoma. At the data cut-off (November 2018) median follow-up was 10.55 months for patients in the nivolumab arm and 8.02 for patients in the total control arm. Patients received a median of 2 cycles (range 1 to 20) of nivolumab.[38]

The demographics and baseline characteristics for patients enrolled in ATTRACTION-3 are summarised in Table 12. The median age in the nivolumab and total control group were 64 (range:37-82) and 67 (range: 33-87), respectively. A substantial proportion of the patients in the nivolumab arm (46.7%) and the total control arm (59.3%) were aged 65 years or older. Patients randomised to the nivolumab arm were overall comparable to patients randomised to the control arm in terms of baseline characteristics. The majority of patients in the nivolumab had and ECOG performance score of 1 (51.9%) while in the total control arm slightly more patients had an ECOG score of 0 (51.2%). The most common sites of metastatic disease were the lymph nodes (75.7% and 78% in the nivolumab and total control groups, respectively), lung (46.7% and 44%), and the liver (27.1% and 25.8%). 95.7% of patients were of Asian ethnicity.[38]

Table 12. ATTRACTION-3: Baseline characteristics[38]

Characteristic		Nivolumab	Control group		
			Total	Docetaxel	Paclitaxel
Cohort size		210	208	65	143
Age	Median (range), years	64 (37-82)	67 (33-)	[redacted]	[redacted]
	<65 years, n(%)	112 (53.3)	85 (40.7)	[redacted]	[redacted]
Gender, n (%)	Female	31 (14.8)	24 (11.5)	[redacted]	[redacted]
	Male	179 (85.2)	185 (88.5)	[redacted]	[redacted]
Race, n (%)	Asian	201 (95.7)	200 (95.787)	[redacted]	[redacted]
Histological type	Squamous cell carcinoma	210 (100.0)	209 (100.0)	[redacted]	[redacted]
ECOG PS	0	101 (48.1)	107 (51.2)	[redacted]	[redacted]
	1	109 (51.9)	102 (48.8)	[redacted]	[redacted]
Disease stage[a)]	IIA	1 (0.9)	3 (2.5)	[redacted]	[redacted]
	IIIA	4 (3.7)	5 (4.2)	[redacted]	[redacted]
	IIIB	1 (0.5)	1 (0.8)	[redacted]	[redacted]
	IIIC	2 (1.4)	4 (3.3)	[redacted]	[redacted]
	IV	94 (87.9)	100 (83.3)	[redacted]	[redacted]
	Not evaluated	5 (4.7)	7 (5.8)	[redacted]	[redacted]
Site of recurrence	Lymph Node	159 (75.7)	163 (78.0)	[redacted]	[redacted]
	Peritoneum	5 (2.4)	11 (5.3)	[redacted]	[redacted]
	Liver	57 (27.1)	54 (25.8)	[redacted]	[redacted]
	Lung	98 (46.7)	92 (44.0)	[redacted]	[redacted]
	Pleural Tissue	22 (10.5)	13 (6.2)	[redacted]	[redacted]
	Adrenal Gland	6 (2.9)	7 (3.3)	[redacted]	[redacted]
	Brain	5 (2.4)	1 (0.5)	[redacted]	[redacted]
	Bone	23 (11.0)	25 (12.0)	[redacted]	[redacted]
	Bone Marrow	0	0	[redacted]	[redacted]
	Skin	1 (0.5)	1 (0.5)	[redacted]	[redacted]
	Stomach	0	3 (1.4)	[redacted]	[redacted]
	Other	26 (12.4)	28 (13.4)	[redacted]	[redacted]

ECOG PS: Eastern Cooperative Oncology Group Performance Score
[a)]Summarised for subjects with non-recurrent oesophageal cancer.

B.2.6.1.9. *Results*

At the data cut-off (November 2018), median follow-up in surviving patients was 10.5 months (range: 0.4-33.8) in the nivolumab group and 8.02 months (range: 0.6-34.1) in the control group.[38] A summary of the key outcomes from ATTRACTION-3 is provided in Table 13.

Table 13. ATTRACTION-3: Nivolumab efficacy[38]

Endpoint	Nivolumab	Control group		
		Total	Docetaxel	Paclitaxel
Responses				
Evaluable patients	171	158	49	109
ORR, n (%) [95% CI]	33 (19.3%) [13.7, 26.0]	34 (21.5%) [15.4, 28.8]	[redacted]	[redacted]
Complete response, n (%)[95% CI]	1 (0.6) [0.0, 3.2]	2 (1.3) [0.2,4.5]	[redacted]	[redacted]
Partial response, n (%) [95% CI]	32 (18.7) [13.2,25.4]	32 (20.3) [14.3,27.4]	[redacted]	[redacted]
Stable disease, n (%) [95% CI]	31 (18.1) [12.7,24.7]	65 (41.1) [33.4,49.2]	[redacted]	[redacted]
Progressive disease, n (%) [95% CI]	94 (55.0)	51 (32.3)	[redacted]	[redacted]
UD, n (%)	13 (7.6)	8 (5.1)	[redacted]	[redacted]
OS				
Evaluable patients	210	209	65	144
Median, months (95% CI)	10.91 (9.23, 13.34)	8.38 (7.20, 9.86)	[redacted]	[redacted]
Hazard Ratio (95% CI)	0.77 (0.62,0.96)		[redacted]	[redacted]
Number of events, n/N	160/210	173/209	[redacted]	[redacted]
6 months, % (95% CI)	[redacted]	[redacted]	[redacted]	[redacted]
9 months, % (95% CI)	[redacted]	[redacted]	[redacted]	[redacted]
12 months, % (95% CI)	46.9% (39.9, 53.5)	34.4% (27.8, 40.9)	[redacted]	[redacted]
18 months, % (95% CI)	30.5% (24.4, 36.9)	20.7% (15.4, 26.6)	[redacted]	[redacted]
21 months, % (95% CI)	[redacted]	[redacted]	[redacted]	[redacted]
24 months, % (95% CI)	[redacted]	[redacted]	[redacted]	[redacted]
PFS				
Evaluable patients	210	209	65	144
Median, months (95% CI)	1.68 (1.51,2.73)		[redacted]	[redacted]
Hazard Ratio (95% CI)		1.08 (0.87,1.34)	0.97 (0.71, 1.33)	[redacted]
Number of events (%)	167 (79.5)	162 (77.5)	[redacted]	[redacted]
3 months, % (95% CI)	[redacted]	[redacted]	[redacted]	[redacted]
6 months, % (95% CI)	24.2% (18.6, 30.3)	17.2% (12.1, 23.1)	[redacted]	[redacted]
9 months, % (95% CI)	[redacted]	[redacted]	[redacted]	[redacted]
12 months, % (95% CI)	11.9% (7.8, 16.8)	7.2% (3.8, 12.0)	[redacted]	[redacted]
18 months, % (95% CI)	[redacted]	[redacted]	[redacted]	[redacted]
21 months, % (95% CI)	[redacted]	[redacted]	[redacted]	[redacted]

CI: confidence interval; NA: not available; ORR: Objective response rate; OS: overall survival; PFS: progression-free survival; NR: not reported; UD: unable to determine

B.2.6.1.9.1. Survival

Patients treated with nivolumab had significantly longer median OS than those treated with chemotherapy (10.91 vs 8.38 months, HR 0.77 [95% CI: 0.62, 0.96] P <0.0001). OS rates were also notably higher in the nivolumab group than with chemotherapy at 12 months (46.9% vs 34.4%) and 30 months (16.3% vs 4.8%).

Patients treated with nivolumab had a shorter median PFS than those treated with chemotherapy (1.68 months vs. 3.35 months, respectively (p <0.0001, hazard ratio: 1.08 [95% CI: 0.87, 1.34]), but there was a significant PFS benefit for nivolumab-treated patients at all time points from three months through to 21 months.[38] The PFS and OS results are shown in Figure 10 and Figure 11, respectively.

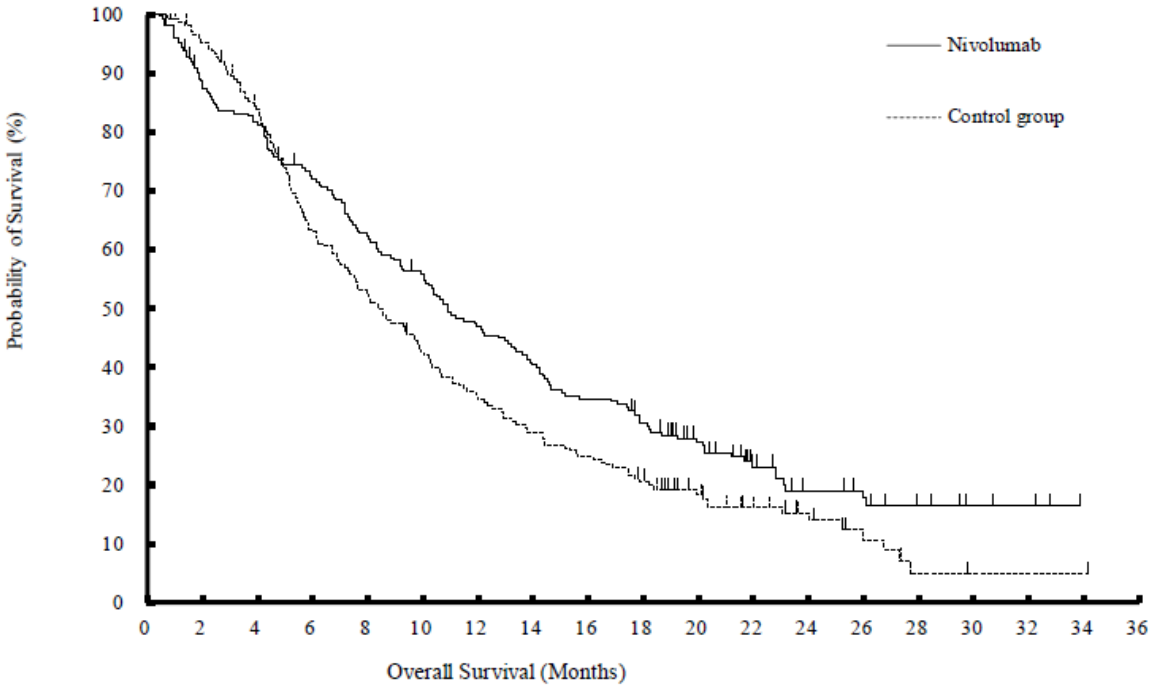


Figure 10. ATTRACTION-3: Kaplan-Meier plot of overall survival in patients receiving nivolumab (ONO-4524) or docetaxel/paclitaxel[38]

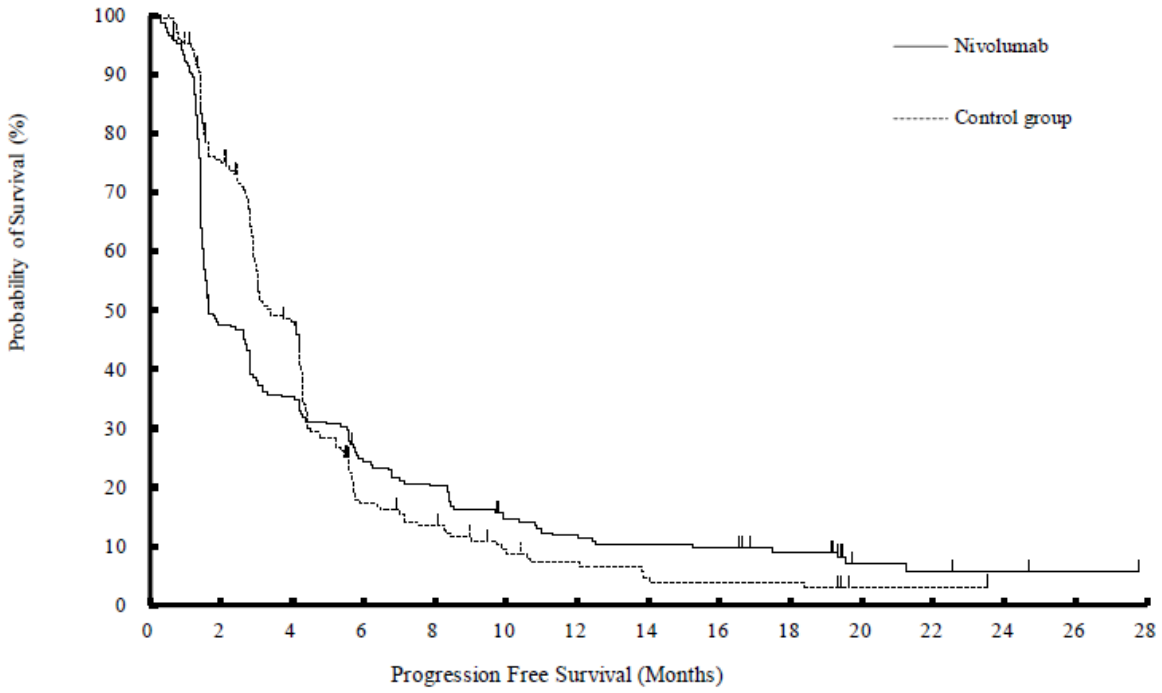


Figure 11.ATTRACTION-3: Kaplan-Meier plot of progression-free survival in patients receiving nivolumab (ONO-4524) or docetaxel/paclitaxel[38]

B.2.6.1.9.2. Response

Overall response rate was similar between the nivolumab and chemotherapy groups (ORR: 19.3% vs. 21.5%, CP: 0.6% vs.1.3%, PR: 18.7%, 20.3%, SD: 18.1 vs. 41.1%, PD: 55% vs. 32.2%).[38]

B.2.6.1.9.3. Patient reported outcomes

ATTRACTION-3 collected patient reported outcomes through the EuroQoL 5 Dimension (EQ-5D) questionnaire. A summary of EQ-5D index scores at each timepoint in the trial (up to 54 weeks) is provided in Table 14 and Figure 12. Additional timepoints are available in the clinical study report but represent smaller patient numbers.

In the nivolumab arm, no meaningful changes in the proportion of patients who reported QoL-related problems were observed during the treatment period in any of the EQ-5D categories. In the control arm, however, the proportion of patients who reported QoL-related problems in the mobility, self-care and usual activities categories after commencing chemotherapy increased by >10% compared with the proportion at the screening stage. [38]

A summary of EQ-VAS scores at each timepoint is presented in Table 15.

EQ-VAS scores remained relatively stable among patients treated with nivolumab. A worsening of >6 scores was observed among patients treated with chemotherapy from early timepoints after commencing treatment.

Table 14. ATTRACTION-3: Summary of EQ-5D index scores at each time point up to 54 weeks (UK based scoring) [38]

	Nivolumab			Control		
Timepoint	**N**	**Mean (SD)**	**Median**	**N**	**Mean (SD)**	**Median**
Screening	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 6	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to Week 6	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 12	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to Week 12	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 18	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to Week 18	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 24	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to Week 24	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 30	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to Week 30	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 36	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to Week 36	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 42	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to Week 42	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 48	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to Week 48	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 54	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to Week 54	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 15. ATTRACTION-3: Summary of EQ-VAS scores at each time point up to 54 weeks[38]

	Nivolumab			Control		
Timepoint	**N**	**Mean (SD)**	**Median**	**N**	**Mean (SD)**	**Median**
Screening	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 6	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to Week 6	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 12	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to Week 12	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 18	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to Week 18	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 24	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to Week 24	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 30	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to Week 30	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 36	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to Week 36	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 42	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to Week 42	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 48	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to Week 48	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Week 54	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Change from baseline to Week 54	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Figure 12. ATTRACTION-3: summary of EQ-5D index data

B.2.6.1.10. *Subgroup analyses*

Results of subgroup analyses for the ATTRACTION-3 nivolumab and chemotherapy arms, as at the database lock on 30 November 2018, are shown in Table 13, Figure 13 and Figure 14. For OS, the superior treatment effect of nivolumab over chemotherapy was consistently observed across the majority of subgroups (Figure 13 and Figure 14). Similar results were observed for PFS, and ORR.[38]

Table 16. ATTRACTION-3: Subgroup analyses for OS and PFS[38]

Stratification factor		OS		PFS		ORR	
		Nivolumab (n)	Chemotherapy (n)	Nivolumab (n)	Chemotherapy (n)	Nivolumab (n)	Chemotherapy (n)
Cohort (ITT)		210	209	210	209	171	158
Age							
<65 years	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
>=65 years	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
65-<75 years	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
>=75	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Sex							
Male	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Female	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Race							
Asian	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
White	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
ECOG Performance Status							
0	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
1	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Recurrent							
No	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Yes	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Lesion sites (TNM classification)							
	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Stratification factor		OS		PFS		ORR	
		Nivolumab (n)	Chemotherapy (n)	Nivolumab (n)	Chemotherapy (n)	Nivolumab (n)	Chemotherapy (n)
Cervical Oesophagus	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Thoracic Oesophagus	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Cervical and Thoracic Oesophagus	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Unknown	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Histological classification							
Squamous Cell Carcinoma	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number of organs with metastases (IWRS)							
<=1	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
>=2	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Lymph Node metastasis							
No	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Yes	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Liver metastasis							
No	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Yes	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Lung metastases							
No	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Yes	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Bone metastasis							
No	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Stratification factor		OS		PFS		ORR	
		Nivolumab (n)	Chemotherapy (n)	Nivolumab (n)	Chemotherapy (n)	Nivolumab (n)	Chemotherapy (n)
Yes	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Target lesion							
No	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Yes	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Past treatments for cancer (surgery)							
No	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Yes	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Past treatments for cancer (radiotherapy)							
No	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Yes	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
History of smoking							
Never	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Former	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Current	No. of patients	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	No. of events	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

CI: Confidence interval; ECOG: Eastern Corporative Oncology Group; DOR: duration of response; ITT: intention to treat; ORR: Objective response rate; OS: overall survival; PFS: progression-free survival.

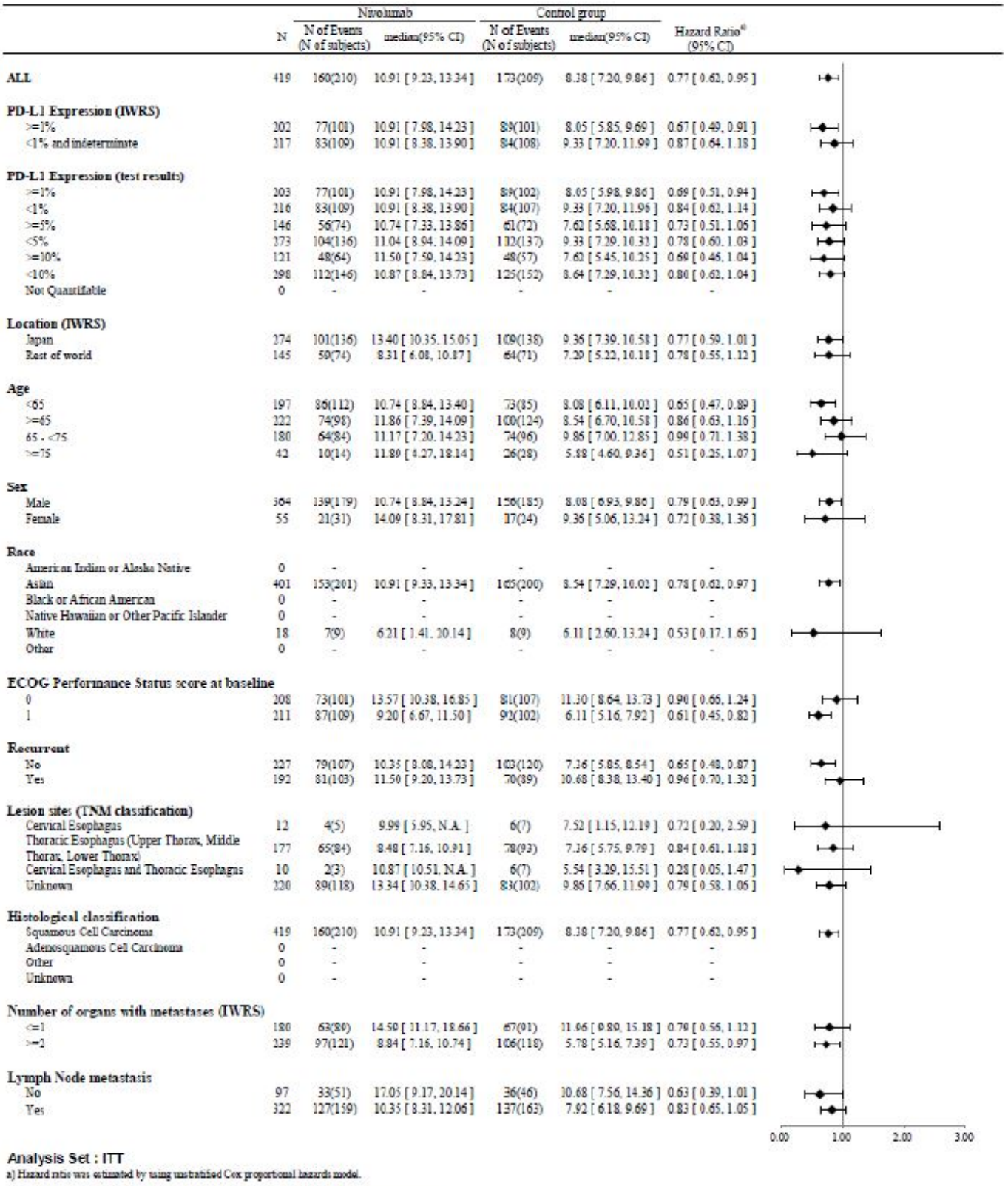


Figure 13. ATTRACTION-3: Forest Plots of Subgroup Analyses for Overall Survival 1[38]

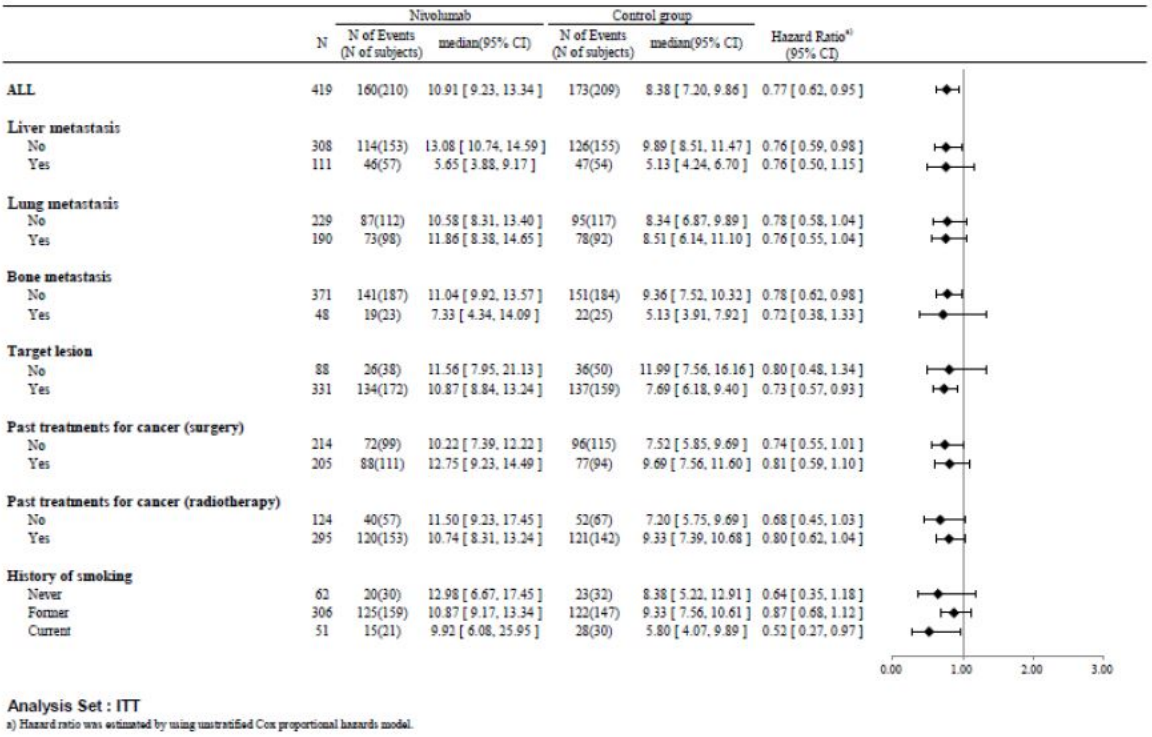


Figure 14. ATTRACTION-3: Forest Plots of Subgroup Analyses for Overall Survival 2[38]

B.2.6.2. ATTRACTION-1

B.2.6.2.1. *Study design*

ATTRACTION-1 (ONO-4538-07, JapicCTI-No.142422) is a Phase II, open-label, single arm multicentre trial in Japanese patients with advanced oesophageal cancer who are refractory or intolerant to fluoropyrimidine, platinum and taxane based chemotherapy.[39] A total of 65 patients with oesophageal carcinoma were enrolled in this study. 64 patients were evaluated for efficacy. It was calculated that a sample of at least 53 patients would be required to provide power of at least 80% to detect a significant overall response (complete response or partial response).[39] The study design of ATTRACTION-1 is shown in Figure 15.

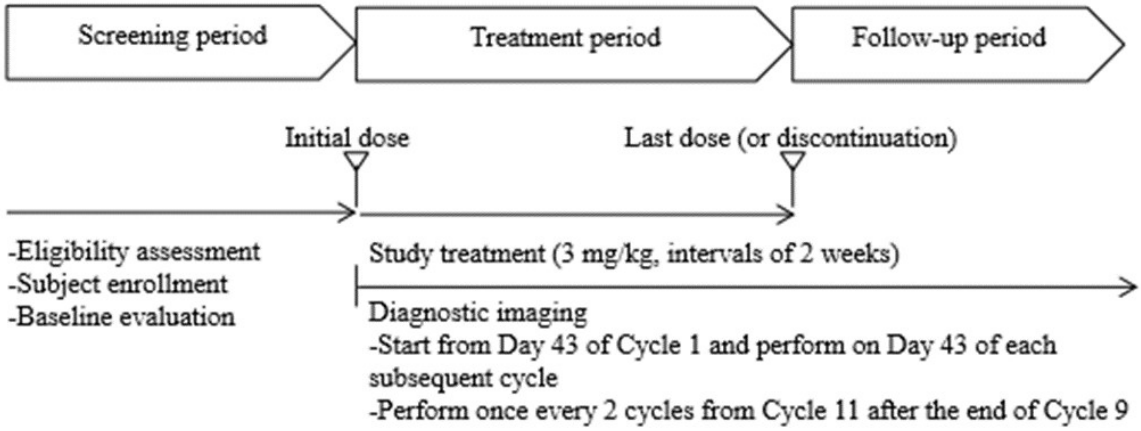


Figure 15**. Study design of ATTRACTION-1**[39]

B.2.6.2.2. *Eligibility criteria*

ATTRACTION-1 enrolled patients with oesophageal cancer refractory or intolerant to standard therapies. The main eligibility criteria are set out below.[39]

Table 17. Inclusion and exclusion criteria for ATTRACTION-1[39]

Key inclusion criteria	Key exclusion criteria
• Men and women of at least 20 years of age • Oesophageal cancer with the major lesion in the oesophagus, histological type of the primary lesion is squamous, adenosquamous, or adenomatous cancer as confirmed by pathological diagnosis • Refractory or intolerant to fluoropyrimidine-, platinum-, and taxane-based chemotherapies and not candidates for radical resection • At least one measurable lesion (as defined by the RECIST guideline v1.1[43]) • ECOG Performance Status score 0 or 1 • A life expectancy of at least three months • Adequate organ function (assessed by white blood cell, neutrophil, and platelet counts, measurement of haemoglobin concentration, and liver and kidney function tests) • Use of contraception	• Significant malnutrition • Patients with apparent tumour invasion on organs located adjacent to the oesophageal disease • Multiple primary cancers • Residual adverse effects of previous therapy • History of, or current severe hypersensitivity to any other antibody products • Presence of chronic or recurrent autoimmune disease • Interstitial lung disease, pulmonary fibrosis, diverticulitis, or symptomatic gastrointestinal ulcerative disease • Pregnant or breastfeeding women • History of treatment with nivolumab (ONO-4538 or MDX-1106 or BMS-936558), anti-CTLA-4 antibody, or other antibody or pharmacological therapies intended to control T cells.

B.2.6.2.3. *Study medications*

Patients received nivolumab 3 mg/kg IV three times at intervals of two weeks, followed by diagnostic imaging at Week 6, which constituted one treatment cycle. Each cycle lasted six weeks. Treatment continued until unacceptable toxicity or disease progression, diagnosed by the investigator or sub investigator according to the RECIST guideline (version 1.1).[39]

B.2.6.2.3.1. Treatment beyond disease progression

Patients with progressed disease were allowed to continue nivolumab treatment upon meeting the following criteria:[39]

- Patients show no acute disease progression and continuing nivolumab treatment is expected to be clinically useful
- Patient tolerates nivolumab treatment
- Patient has a stable ECOG performance score
- Continuing nivolumab treatment will not delay preventive interventions for significant complications associated with disease progression (e.g. brain metastases).

B.2.6.2.4. *Study endpoints*

The primary, secondary and exploratory endpoints of ATTRACTION-1 are shown in Table 18.

Table 18. Study endpoints in ATTRACTION-1[39]

ATTRACTION-1 Endpoints	
Primary endpoint	• **Centrally assessed objective response** defined as the proportion of patients whose best overall response was CR or PR according to RECIST guideline version (1.1)
Secondary and exploratory endpoints	• **Objective response rate (investigator-assessed)** • **Overall survival** defined as the time from the first dose of nivolumab to death from any cause • **Progression-free survival** defined as the time from the first dose of nivolumab to disease progression or death from any cause • **Response of primary oesophageal lesion (**investigator assessed • **Duration of response** calculated from the following equation: ("Time from date of randomisation until either the overall response was assessed as progressive disease for the first time after confirmed response or the patient died of any cause"+1)/ 30.4375 • **Disease control rate (**percentage of patients whose best objective response was CR, PR, or SD) • **Time to response** calculated from the following equation: ("Time from randomisation until first assessment of confirmed CR or PR"+1)/30.4375 • **Immune-related objective response:** Based on the outcome of diagnostic imaging assessed by the central image analysis centre using RECIST Guideline Version 1.1[43] • **Immune-related progression-free survival d**efined as the ("Time from start of study treatment until immune-related overall response of immune-response, PD or day of all-cause death, whichever earlier"+ 1)/30.4375 • **Immune-related best objective response** based on the outcome of diagnostic imaging assessed by study site investigator according to the evaluation criteria for immune-related responses[44] • **Tumour burden** defined as percentage change from baseline in the sum of tumour diameters • **Best overall response** using RECIST Guideline Version 1.1[43], up to study completion Safety will be analysed through the incidence of adverse events, serious adverse events Safety events assessed through adverse events, laboratory tests, vital signs, 12-lead ECG, chest X-ray and ECOG performance scores.
CR: complete response, disappearance of all (non-lymph node) target lesions. Any pathological lymph nodes (whether target or non-target) must have reduction in the short axis to <10 mm. PR: partial response, at least a 30% decrease in the sum of diameters of target lesions, taking as reference the baseline sum diameters. SD: stable disease, neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD, taking as reference the smallest sum diameter while on study. PD: progressive disease, at least a 20% increase in the sum of diameters of target lesions, taking as reference the smallest sum on study (this includes the baseline sum if that is the smallest on study). In addition to the relative increase of 20%, the sum must also demonstrate an absolute increase of at least 5 mm.	

B.2.6.2.5. *Statistical analyses*

The proportion of patients with each RECIST-defined response was noted together with 95% CIs (calculated with normal approximation). OS and PFS were analysed with Kaplan-Meier curves and estimated median values were determined, with 95% CIs. Adverse events were summarised by grade, system organ class and preferred term, with incidence calculated for each event. All analyses were performed with SAS version 9.3.[40]

B.2.6.2.6. *Sample size and power calculation*

The planned number of patients for ATTRACTION-1 is 60, which is expected to detect a significant difference in the response rate by a binomial test (normal distribution) at a one-sided significance level of 2.5% and a minimum power of 80% was calculated. The estimation of sample size was based on an assumed threshold response rate of 5% and an expected response rate of 15%. A minimum power of approximately 64.0% may be ensured when the expected response rate is at least 12.5%.[39]

B.2.6.2.7. *Baseline patient characteristics*

A total of 65 patients were enrolled between 25 February 2014 and 14 November 2014 (Table 19). All patients were histological type squamous-cell carcinoma and exclusively of Asian race. One patient was excluded from the analysis of primary and secondary endpoints due to having multiple primary cancers; however, this patient was included in the safety analysis. At the data cut-off (17 November 2016) median follow-up was 10.8 months and patients received a median of three cycles (range one to ten) of nivolumab.[41]

Table 19. ATTRACTION-1: Baseline characteristics[41]

Characteristic		Nivolumab
Cohort size		65
Age	Median (range), years	62 (49 – 80)
Gender, n (%)	Female	11 (16.9)
	Male	54 (83.1)
Race, n (%)	Asian	65 (100)
Histological type (%)	Squamous cell carcinoma	65 (100)
ECOG PS (%)	0	29 (44.6)
	1	36 (55.4)
Disease stage (%)	I	9 (14.0)
	II	11 (17.0)
	III	24 (37.0)
	IV	20 (31.0)
	Not evaluable	1 (1.0)
Site of metastasis (%)	Cervical lymph node	[redacted]
	Abdominal lymph node	[redacted]
	Other lymph nodes	[redacted]
	Liver	[redacted]
	Lung	[redacted]

	Bone	[redacted]
	Others	[redacted]
No. of prior chemotherapy treatments (%)	≤ 2 regimens	21 (32.3)
	3 regimens	24 (36.9)
	≥ 4 regimens	20 (3)
ECOG PS: Eastern Cooperative Oncology Group Performance Score		

B.2.6.2.8. *Patient disposition*

All of the 65 patients enrolled in the study received nivolumab. Out of these patients, [redacted] continued the treatment phase as of data cut-off November 17 2016, [redacted] terminated the treatment phase and shifted to the follow-up phase, and [redacted] discontinued without sifting to the follow-up phase. Of the [redacted] of patients who completed the treatment phase and shifted to the follow-up phase, the most common reason for shifting to the follow-up as reported by [redacted] of patients was an overall response of PD. A summary of the patient disposition is provided in Table 20.[41]

Table 20. ATTRACTION-1: Patient disposition[41]

	Nivolumab
Number of patients (intention-to-treat)	65
Continuation of treatment phase	[redacted]
Completion of the treatment phase (shifting to the follow-up phase)	[redacted]
Overall response of PD	[redacted]
Worsening of clinical symptoms due to disease progression	[redacted]
Onset of ≥ 2 of interstitial lung disease	[redacted]
Not received a dose of nivolumab within 6 weeks after the last dose	[redacted]
Investigation or subinvestigator judged that continuation of study treatment was in appropriate for other reasons	[redacted]
Withdrawal (not shifting to the follow-up phase)	[redacted]
Patients requests to withdraw from the study	[redacted]
Patient is found to meet any of the exclusion criteria	[redacted]
Investigation or subinvestigator judged it inappropriate to continue with further study procedures because of progressive disease	[redacted]
Patient has failed to return to the study site	[redacted]
Investigation or subinvestigator judged that continuation of study treatment was in appropriate for other reasons	[redacted]
Completion of the follow-up phase	[redacted]
Discontinuation	[redacted]
Patients requests to withdraw from the study	[redacted]
Patient is found to meet any of the exclusion criteria	[redacted]
Investigation or subinvestigator judged it inappropriate to continue with further study procedures because of progressive disease	[redacted]
Patient has failed to return to the study site	[redacted]
Investigation or subinvestigator judged that continuation of study treatment was in appropriate for other reasons	[redacted]
.1 PD: progressive disease	

B.2.6.2.9. *Results*

Clinical efficacy results from ATTRACTION-1 are summarised in Table 21 and represent data from the database lock in 17 November 2016.[41]

B.2.6.2.9.1. Response

Of 64 evaluable patients, 11 (17.2%) had an objective response by central assessment (defined as the proportion of patients whose best overall response was complete response [CR] or partial response [PR]). Three (4.7%) had CR and 8 (12.5%) had PR. Fourteen patients (21.9%) had an objective response by investigator assessment. [41]

Disease control (CR, PR or stable disease, by central assessment) was achieved in 27 (42.2%, 95% CI 30.9–54.4) of patients. Immune-related objective response and disease control were reported in 16 patients (25%, 95% CI 16.0–36.8) and 43 patients (67.2%, 55.0–77.4).[41]

B.2.6.2.9.2. Survival

At the time of data cut-off (17 November 2016), 55 (85.9%) patients had experienced an event for OS analysis.

Median duration of OS was 10.78 months (95% CI 7.39–13.93, Figure 16). Median centrally assessed PFS was 1.5 (95% CI 1.4–2.8) months. Median investigator assessed PFS was 2.3 (1.5–3.0) months (Figure 17).

PFS and OS at one year were 10.3% and 45.3% respectively. The median time to progression was 2.8 months (95% CI 1.4–2.8) and tumour burden and target lesion size decreased in 29 patients (45%) by investigator assessment.[41]

Table 21. ATTRACTION-1: Nivolumab efficacy[41]

Endpoint	Centrally assessed (n=64)		Investigator assessed (n=64)	
	N	% (95% CI)	N	% (95% CI)
Response				
Complete response	3	4.7 (1.6, 12.9)	2	3.1 (0.9, 10.7)
Partial response	8	12.5 (6.5, 22.8)	12	18.8 (11.1, 30.0)
Stable disease	16	25.0 (16.0, 36.8)	20	31.3 (21.2, 43.4)
Progressive disease	29	45.3	29	45.3
Not assessable	3*	4.7	1	1.6
Objective response†	11	17.2 (9.9, 28.2)	14	21.9 (13.5, 33.4)
Disease controlled §	27	42.2 (30.9, 54.4)	34	53.1 (41.1, 64.8)
PFS				
Median PFS (months), 95% CI	1.51 (1.41, 2.79)		2.33 (1.5, 3.0)	
PFS at Month 3 (%), 95% CI	[redacted]		[redacted]	
PFS at Month 6 (%), 95% CI	[redacted]		[redacted]	
PFS at Month 9 (%), 95% CI	15.4 (7.7, 25.5)		20.9 (11.9,31.7)	
PFS rate at one year (%), 95% CI	10.3 (4.2,19.4)		12.9 (6.0, 22.4)	
PFS rate at two years (%), 95% CI	8.6 (3.2, 17.3)		9.7 (3.9, 18.5)	
PFS rate at Month 30 (%), 95% CI	[redacted]		[redacted]	
OS				
Median OS (months), 95% CI	10.78 (7.39, 13.3)			
OS rate at Month 3 (%), 95% CI	[redacted]			
OS rate at Month 6 (%), 95% CI	[redacted]			
OS rate at Month 9 (%), 95% CI	56.3 (43.3, 67.4)			
OS rate at one year (%), 95% CI	45.3 (32.9, 56.9)			
OS rate at two years (%), 95% CI	17.2 (9.2, 27.3)			
OS rate at 30 months (%), 95% CI	[redacted]			

*Including patients who have no target lesion. †Complete or partial response. §Complete response, partial response or stable disease. CI: confidence interval' PFS: progression-free survival; OS: overall survival

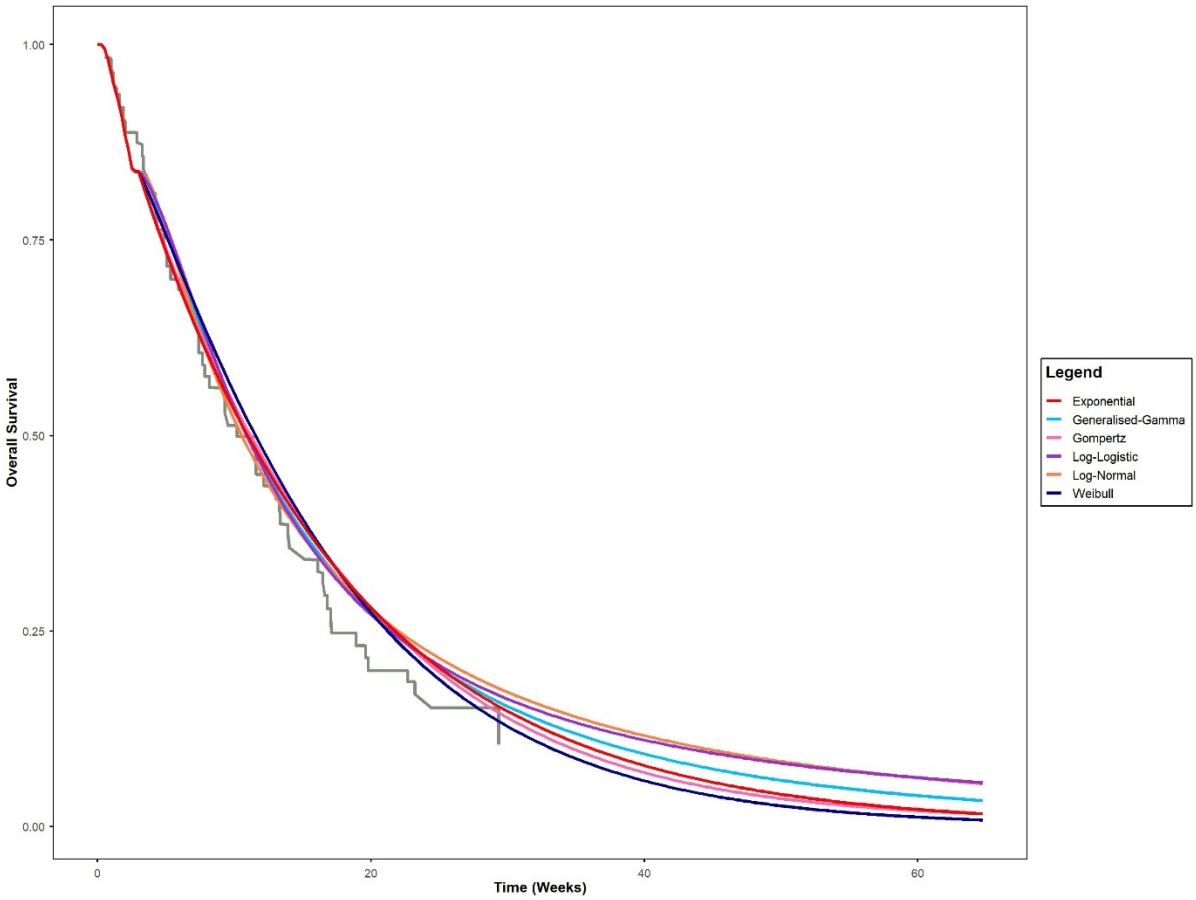


Figure 16.ATTRACTION-1: Kaplan-Meier analyses of overall survival

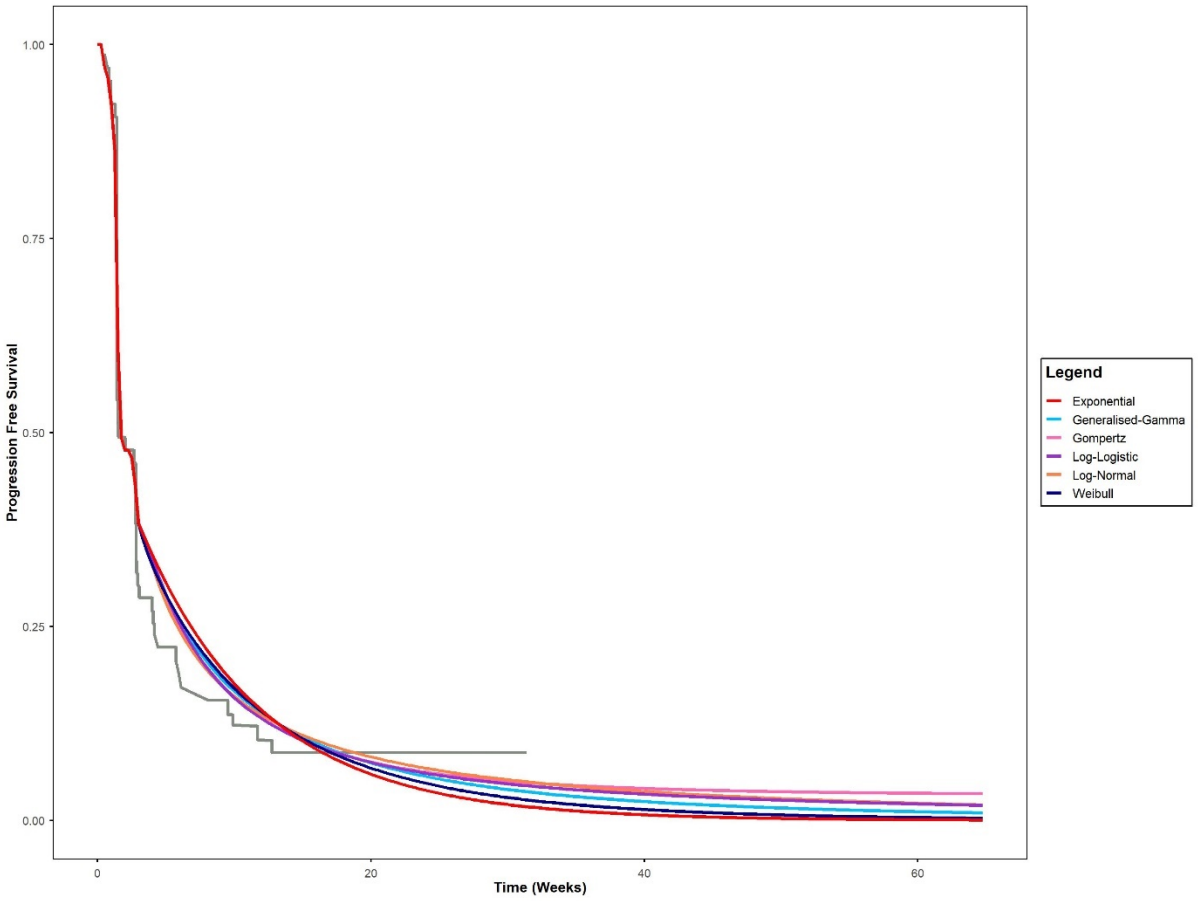


Figure 17. ATTRACTION-1: Kaplan-Meier analyses of progression-free survival

B.2.6.2.10. *Subgroup analysis*

The results of subgroup analyses on the response rate (central assessment), OS, an PFS (central assessment) are provided in Table 22.

[redacted]

[redacted]

Table 22. ATTRACTION-1: Subgroup analyses on Response Rate, Overall Survival, Progression-free survival[41]

Stratification factor	Response Rate		Overall Survival		Progression-free survival	
	Central assessment				Central assessment	
	n/N (%)	95% confidence interval[e)]	n/N (%)	Median [95% confidence interval][e)]	n/N (%)	Median [95% confidence interval][e)]
Past treatments for oesophageal cancer (radiotherapy)						
Absent	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Present	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
History of alcohol consumption[a)]						
Absent	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Present	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Years of alcohol consumption (years)[b)]						
<35	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
35-<40	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
>=40	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
History of smoking[c)]						
Absent	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Present	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Years of smoking (years)[d)]						
<25	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
25 - <40	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
>=40	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Performance Status (ECOG)						
0	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Histological type						
Squamous cell carcinoma (well differentiated type)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Squamous cell carcinoma (moderately differentiated type)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Squamous cell carcinoma (moderately-poorly differentiated type)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Squamous cell carcinoma (poorly differentiated type)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Stratification factor	Response Rate		Overall Survival		Progression-free survival	
	Central assessment				Central assessment	
Squamous cell carcinoma (detail unknown)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Past treatments for oesophageal cancer (the number of regimens in pharmacotherapy)						
<=2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
>=4	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Past treatments for oesophageal cancer (history of surgery)						
Absent	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Present	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

By the central image analysis laboratory, best overall response was assessed in accordance with RECIST Guideline Ver. 1.1.
The subjects with CR or PR in best overall response were included.
a) The subjects responded as “Nondrinker (never been a habitual drinker)” in history of alcohol consumption was classified as “Absent”. The rest of the subjects were classified as “Present”.
b) The subjects whose histories of alcohol consumption were “Present” were included.
c) The subjects responded as “Nonsmoker (never been a smoker)” in history of smoking were classified as “Absent”. The rest of the subjects were classified as “Present”.
d) The subjects whose histories of smoking were “Present” were included.
e) 95% confidence interval based on Wilson method

B.2.7. Subgroup analysis

Available subgroup analyses for ATTRACTION-3 are described in 0, in line with results reported in the interim clinical study report (CSR), as detailed in Appendix E.

B.2.8. Meta-analysis

Direct evidence for comparative efficacy of nivolumab versus chemotherapy may be drawn from the ATTRACTION-3 study, so that no meta-analysis is required. Indirect treatment comparisons deriving comparative efficacy using ATTRACTION-3 are presented in Section B.2.9.

B.2.9. Indirect and mixed treatment comparisons

Key points

- A network could not be formed that included both BSC and irinotecan.
- As can be anticipated, outcomes are worse for patients receiving BSC than those receiving docetaxel, with a HR of 1.6.
- Docetaxel was associated with slightly worse outcomes than paclitaxel (HR: 0.89). This can be anticipated based on the published comparisons of paclitaxel and docetaxel.

As outlined in Section B.1.3.4, UK guidelines recommend chemotherapy for patients who have progressed on first-line therapy; however, specific chemotherapy regimens are not defined in the NICE clinical guidelines in the second-line setting.[8-10] ESMO guidelines recommend taxane monotherapy for the second-line treatment (after failure of first-line treatment with taxane combination therapy) of OC but highlight a lack of current evidence in relation to specific chemotherapies in this patient population.[11] Clinical expert opinion obtained during a clinical advisory board meeting supported evidence on a lack of standard of care for treatment-experienced OC patients. In the second line setting, decisions on treatment options for unresectable, advanced or metastatic OC patients were described as highly individualised. Chemotherapy agents such as paclitaxel or docetaxel are usually the treatment of choice in this setting. Thus, the comparators applied in ATTRACTION-3, docetaxel and paclitaxel, represents the most appropriate comparator in the UK setting for previously treated OC patients.

Best supportive care (BSC) represents a further treatment option for patients with OC failing first-line therapy, particularly those unable to receive second-line therapies due to age or comorbidities. For this reason, it is included in the scope of this submission. As no direct

evidence is available to describe outcomes for patients receiving BSC, an indirect comparison is considered relevant.

Another treatment that was discussed as possibly relevant was irinotecan. However, advice obtained from clinical experts during an advisory board confirmed that irinotecan is currently not routinely used in UK clinical practice, so that clinicians did not consider it a relevant comparator. BMS market research conducted in 2019 estimated that clinicians treat approximately [redacted] of OC patients in the second line setting with irinotecan. Networks including irinotecan were examined but were considered insufficient and so did not form the basis of the analysis presented here. As irinotecan is not considered relevant to the decision problem in terms of clinical practice in the UK, it was deemed that any attempt to include it this way would add uncertainty and therefore not further the objective of estimating the cost-effectiveness of nivolumab in the current UK setting. The decision was made to estimate the efficacy of BSC with studies that report docetaxel and paclitaxel separately as they are more numerous and detailed. Therefore, fewer assumptions are required and the efficacy of BSC can be estimated with more certainty.

B.2.9.1. Identification of evidence

As described in Section B.2.1, an SLR was undertaken to identify the clinical effectiveness evidence (efficacy and safety) of interventions for the treatment of unresectable advanced oesophageal cancer where standard chemotherapy has failed. This SLR was used to inform the indirect comparison outlined below. Full details of the process and methods to identify and select the relevant clinical evidence are summarised in Appendix D.

B.2.9.2. Study Selection for the NMA

Of the 54 unique studies that were found in the SLR, 12 studies[21-24, 45-53] reported at least one treatment of interest for this NMA. Of those, three[51, 52 45]were single arm studies and could not provide information about comparable efficacy in the network. Of the remaining 9 studies, six had KM data (which could verify or contribute results) and three were able to provide comparative values between two nodes in the network; one for BSC versus docetaxel[21] and two for docetaxel versus paclitaxel[22, 47]. Additionally, one study[54] was a publication of the ATTRACTION-3 trial and so was not considered useful as patient level data was available for his study.

The 9 studies were examined for their suitability for inclusion in terms of population, treatment, inclusion and exclusion criteria, and availability of outcomes. There was considerable inconsistency in treatments included. For example, while a number of studies included BSC, all the comparators were different. This would introduce considerable heterogeneity and reduce transitivity if all were to be included in the network.

For six[23, 24, 45, 46, 48, 49] of the nine studies, only one arm from each of these studies was of interest. Including all of these can decrease the power of the NMA to estimate the links of interest. It was therefore considered sensible to only include those studies that provided direct links. Details of these studies with direct links and the populations can be seen in Appendix A.

All studies reported in the literature applied inclusion criteria that allowed patients with an ECOG PS score of 2 to be included (Table 23). This is contradictory to the inclusion criteria of ATTRACTION-3 (B.2.6.1.2). However, due to the absence of other studies to inform these links, they were all included in this NMA. The impact of including these different populations is discussed in Section B.2.9.3.1

Table 23: Prognostic factors of patients in studies included in the network meta-analysis from ATTRACTION-3[38]

	ATTRACTION-3[38]			Moriwaki et al., 2014[21]		Nakatsumi et al. 2016[22]		Shirakawa et al., 2014[47]	
Treatment	**Nivolumab**	**Paclitaxel**	**Docetaxel**	**Docetaxel**	**BSC**	**Paclitaxel**	**Docetaxel**	**Paclitaxel**	**Docetaxel**
Dose	240 mg	100g/m2 weekly for 6 weeks, 2-week break	75mg/m2 once every 3 weeks	70mg/m2 every 3 weeks	NR	Weekly, 100mg/m2	70mg/m2 week 1, 4 and 7	100 mg/ m2 weekly for 6 weeks, followed by 1 week's rest	70mg/m2 every 3 weeks
Study Design	Randomised open label			Retrospective		Retrospective		Retrospective	
ECOG 0 %	48.1	[redacted]	[redacted]	27	18	28.6	32	12.9	18.9
ECOG 1 %	51.9	[redacted]	[redacted]	61	33	64.3	52	80.6	73.5
ECOG 2 %	0	[redacted]	[redacted]	12	49	0	16	6.5	7.6
ECOG 3 %	0	[redacted]	[redacted]	0	0	7.1	0	0	0
Med Age	64	[redacted]	[redacted]	64	67	65	63	61	64
Items of note	49% recurrent	[redacted]	[redacted]	93.9% metastatic	95.6% metastatic	48.4% recurrent 51.6% metastatic	56.8% recurrent 43.2% metastatic		

B.2.9.2.1. *Networks Including Irinotecan*

Although irinotecan was initially considered to be a potential comparator, the decision was taken not to include it in the NMA due to the lack of informing studies, assumptions required and the instability of the resulting network geometry.

Only four studies were identified that could support the inclusion of irinotecan; however, two of these studies compared to irinotecan combination therapy and thus could not be mixed with the remaining study[45, 50, 53, 55]. An additional study linked irinotecan to a mixed docetaxel/paclitaxel arm, although this did not report the ratio of docetaxel and paclitaxel received or the dosing regimens.[46] This would require the assumption that ratio and dosing of docetaxel/paclitaxel are equivalent to the control arm of ATTRACTION-3. While this would allow a link between the combined control arm and irinotecan, there would be no link to BSC: docetaxel could not be included separately due to the lack of studies comparing docetaxel with combined taxanes and there is only one link from docetaxel to BSC available in the network. The resulting network would be minimal and offer no information about BSC, which clinicians have confirmed is more relevant to the UK clinical setting than irinotecan.

As stated above, taxanes (docetaxel and paclitaxel) represent the main comparator to nivolumab for OC patients in the second-line setting. Nevertheless, it is relevant to consider BSC as a potential treatment option despite the lack of head to head studies to estimate efficacy between nivolumab and BSC, and between BSC and taxanes. However, the available literature allows BSC to be included into a network in which a reasonable number of studies are available to support the estimation of the taxanes as separate treatment arms (docetaxel and paclitaxel) and therefore stabilise the network more than considering them as a combined “taxane” arm. Thus, this approach offered a much more robust analysis with more informing studies for the relevant treatments.

B.2.9.3. Evidence Network

Combining the three studies from the clinical SLR with the ATTRACTION-3 data enabled a network to be constructed for OS (Figure 18).

The study that links BSC to the network does not report on PFS. However, the study does report on post-progression survival (PPS). The study authors noted that while the time to death from initiation with docetaxel was measurable, it was difficult to quantify the time of initiation with BSC. Therefore, the start date was defined as the date of disease progression on platinum-based chemotherapy from any cause or to the last follow up (censored). This is comparable to the measurements used in other studies and so it was included as if it were a measure of OS.

There are two studies that inform an estimate of OS between paclitaxel and docetaxel from the SLR. There is also an additional estimate from the ATTRACTION-3 study.[22, 47] ATTRACTION-3 is the only study that links nivolumab into the network.

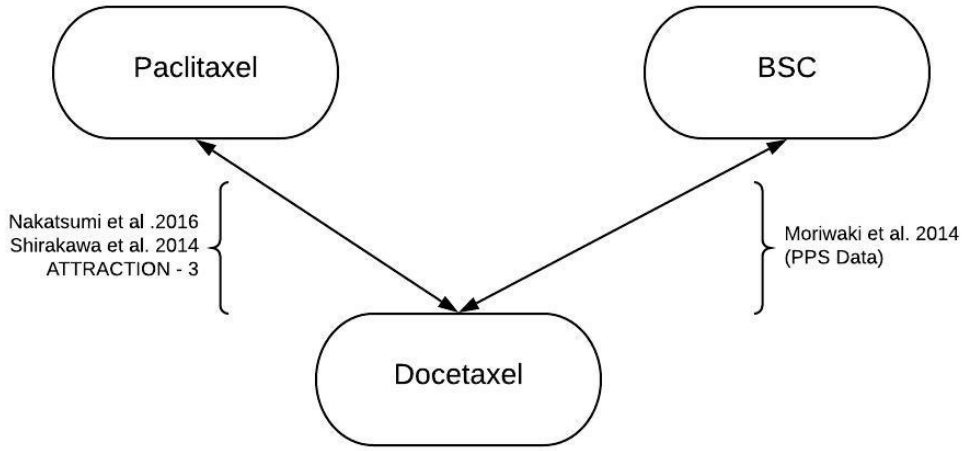


Figure 18: Network Geometry for indirect treatment comparison

The estimate of PFS between taxanes is informed by the same two studies as in the OS network. The study that reported PPS between BSC and docetaxel did not report PFS. Therefore, there is no link to make an estimate. For use in an economic model, a suitable assumption about the relationship between OS and PFS in the BSC arm has to be made; the assumption used in the economic model for this decision problem assumes the same OS to PFS ratio as seen in the docetaxel arm, where PPS is assumed to be equivalent to OS for BSC as described.

The resulting HRs estimated by the model from these networks will be applied to the docetaxel arm of the ATTRACTION-3 study. This is appropriate because the ATTRACTION-3 PLD is available, therefore reconstruction does not require assumptions. The docetaxel arm should be scaled as this is the arm for which the most information is available; it has the greatest number of links in the network. Additionally, application of an HR to a taxane is more appropriate given the proportional hazards assumption holding from the information available; seen in Figure 19, which plots the Moriwaki study information.[21] Additionally, global Schoenfeld test for this time-event data produces a p-value of 0.43 for the treatment arm covariate; a significant result would indicate non-proportional hazards.

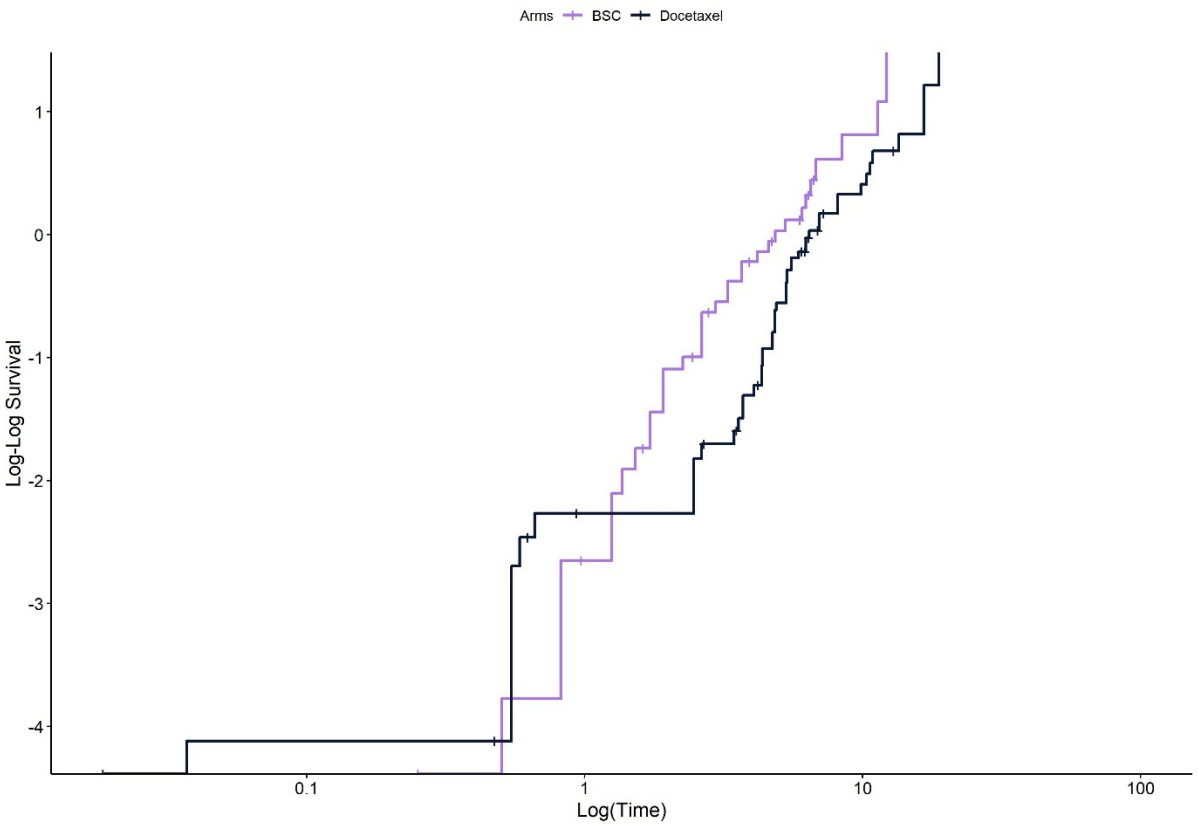


Figure 19: Complimentary hazard plot for evidence of BSC and docetaxel

B.2.9.3.1. *Assessment of consistency and transitivity in included trials*

The authors of the study comparing BSC with docetaxel note that the groups in each arm were significantly different.[21] Notably, 49% of patients treated with BSC had an ECOG PS score of 2 compared to 12% treated with docetaxel. This is particularly prognostic of outcomes and as expected, the testing showed that the ECOG PS and GPS scores are significant factors in outcomes. Therefore, the NMA uses the adjusted values as reported in the literature. The value used in the NMA (0.62) adjusts for six significant factors. While this cannot be replicated, the unadjusted value could be reconstructed. Using the adjusted value reduces the apparent differences between trials and balances the population between arms within trial.

As discussed, all studies other than ATTRACTION-3 contained patients who had an ECOG PS score of 2. For all arms this was over 5% and for the docetaxel arm in Nakatsumi et al., 2016 over 15%.[22] Within the study reported in Nakatsumi et al., no patients in the paclitaxel arm are reported to have an ECOG PS of 2, but 7.1% have an ECOG PS score of 3 and no adjustment between these populations is reported.[22] Additionally, while Moriwaki et al. adjusted for the differences between arms, the resulting HR is for a group of patients that has been adjusted such that approximately 12% might have an ECOG PS of 2.[21] This contrasts with the ATTRACTION-3 study where no patients had an ECOG PS score of 2.

Nivolumab has a different mechanism of action to both docetaxel and paclitaxel. Therefore, to include them in the same network can be problematic and applying any generated HR to

nivolumab data from another arm or vice versa would assume a similar profile and distribution of events, which may not hold. It would be considered appropriate to apply a generated HR for the BSC arm to taxane arm however given evidence of proportional hazards between BSC and docetaxel (Figure 19).

Of the examined studies, none of the median time to event or HRs were notably different to the reported values. While there was a difference in the absolute time to event and HRs between treatments, the order of efficacy was not different between studies. Paclitaxel was always considered to be more effective than docetaxel. Only one study informs the link between docetaxel and BSC with docetaxel considered more effective.[21]The magnitude of the difference was not constant across studies and neither was the time to event.

The results for ATTRACTION-3 for PFS are in line with the literature in that paclitaxel is shown to be more effective than docetaxel. However, for OS this is not upheld, and docetaxel is estimated to be slightly more effective than paclitaxel.

B.2.9.3.2. *Studies Excluded from the Network*

A number of studies identified in the clinical SLR were excluded from analysis. These are described in B.2.9.2 and also tabulated in Appendix D.

B.2.9.4. Methods of Analysis

The Technical Support Document (TSD) 2 outlines methods that can be used to conduct an NMA, which informed the methods used.[56] Additionally, TSD3 was used to support assessments of heterogeneity in line with recommendations by NICE for good practice.[57]The Technical Support Document (TSD) 2 outlines methods that can be used to conduct an NMA, which informed the methods used.[56] Additionally, TSD3 was used to support assessments of heterogeneity in line with recommendations by NICE for good practice.[57]

While an NMA of survival analysis endpoints may often use other method, e.g. fractured polynomials, this was not deemed necessary for this analysis because this is more useful where the proportional hazards assumption is violated. As this is not the case (B.1.1.1, Figure 19) adopting a more complex approach where unnecessary can add to uncertainty and detract from the usefulness. Therefore, adopting the method proposed in TSD2 for estimating differences with HRs was deemed appropriate. This if further outlines in Section B.2.9.4.2.

B.2.9.4.1. *Software Used*

To facilitate and validate the inputs to the NMA any available KM data from literature that was to be used in the network were digitised using DigitizeIt Version 2.3.3. Median times for OS and PFS were calculated in R Version 3.5.1 with the Survival package (version 2.43-3) and compared to reported values. Additionally, cox proportional hazard models were used to estimate the hazard ratio (HR) between treatments. For ATTRACTION-3, as PLD was available, it was used to calculated outcomes and HRs. This practice allowed for validation of the published findings and for the generation of HRs. The HRs were used as the treatment effect input to the NMA.

Where an HR was reported, this value was used. Only if there was no HR reported, the reconstructed value was used. This is because the reported values in the literature were calculated with PLD and are therefore considerably more accurate than HRs calculated with digitised data.

Analysis was run in WinBUGS Version 1.4.3.

B.2.9.4.2. *Model used*

A Bayesian approach was taken as this is promoted in TSD 2.

Analysis was run in WinBUGS using the model outline in TSD2. As the input data was given as HRs, these were log transformed and assessed as continuous outcomes with a normal distribution as recommended.

This model can assume that even if underlying data is skewed, the sample means are approximately normally distributed. The likelihood function can therefore be assumed as:

$$Y_{ik} \sim N(\theta_{ik}, se^2_{ik})$$

This can be directly interpreted so the identify link can be used where the parameter of interest (Θ_{ik}) can used for the linear model directly.

As nivolumab has a different mechanism of action, survival profile and distribution of events to other arms in the network, a point estimate HR may not be fully capable to describe the time to event in this arm. Applying a point estimate HR to docetaxel to estimate nivolumab would assume the same distribution and would see the “new” nivolumab arm lose the tail that it is known for. Similarly, using a HR to describe the difference between nivolumab and docetaxel or paclitaxel may unduly influence the assessment of the efficacy of docetaxel and paclitaxel in the network as the underlying distributions may be quite different. As such, it was considered best to omit it from the base case. The relative efficacy of docetaxel and paclitaxel in ATTRACTION-3 is retained in the base case network because it is pertinent to the analysis and the resulting HR will be applied to data from the ATTRACTION-3 study and thus it should be influential.

B.2.9.4.3. *Choice of model*

Both random and fixed effects models were run. This is because of the differing assumptions; namely fixed effect model assume that the treatment effects can be estimated directly from the included population and that it represents the whole population. A random effects model assumes the treatment effects are from a section of the population and that there will be an additional parameter equal to the between-study variance.

B.2.9.4.4. *Assessment of fit*

Model fit was assessed as directed by TSD2, with the use of the DIC and examination of residuals.[56]

B.2.9.5. Results

B.2.9.5.1. *Overall Survival*

The base case analysis shows that, in line with all included studies, BSC is estimated to be less efficacious than docetaxel. Paclitaxel is estimated to be more efficacious than docetaxel. The results, displayed as log hazard ratios, are presented in Table 24. The results indicate that the fixed effects model provide a better fit to the data as the DIC statistics are broadly similar. The HR for docetaxel vs BSC is 1.6, indicating 1.6 death events with BSC for every 1 event for patients receiving docetaxel. The HR for docetaxel vs paclitaxel is 0.89 indicating 0.89 events with paclitaxel for every 1 with docetaxel.

Table 24. Base case results - ITC

	Fixed Effect				Random Effect			
	Mean	**SD**	**Median**	**95% CI**	**Mean**	**SD**	**Median**	**95% CI**
Docetaxel vs BSC log HR	0.4772	0.202	0.4771	0.0806, 0.8729	0.4798	1.226	0.4784	-2.08, 3.029
Docetaxel vs Paclitaxel log HR	-0.1165	0.129	-0.1162	-0.3695, 0.1366	-0.2189	0.7189	-0.193	-1.755, 1.239
σ	-	-	-	-	0.814	0.9005	0.4888	0.030, 3.6
Residual Deviance	5.477	1.992	4.868	3.539, 10.77	3.932	2.636	3.43	0.499, 10.5
pD	1.990	-	-	-	3.627	-	-	-
DIC	3.320	-	-	-	3.422	-	-	-

The outputs from this analysis are in line with expectations. Broadly, the estimate for paclitaxel is between the input estimates of 0.62, 0.67, and [redacted]. However, the input of [redacted] (from ATTRACTION-3) is greater than the other two and it [redacted]. The estimate for BSC is less than the input HR may suggest. However, the difference in estimates between docetaxel and paclitaxel can affect the estimated efficacy of docetaxel and in turn its relative efficacy to BSC.

B.2.9.5.2. *Assessment of heterogeneity*

TSD3 describes that the use of vague priors, despite this being the recommendation in TSD2, can result in counter-intuitive or unrealistic heterogeneity parameters. This is a documented issue and TSD3 recommends the use of deviance statistics and knowledge of the inputs studies to determine the most appropriate model.

While the statistical indication of heterogeneity is used to determine the model type used for these analyses, it is recognised that there may be some uncertainty in the values. Qualitative assessment of the included studies, examination of the log cumulative hazard profiles,

proportional hazards and the between study variance calculated in the analysis all were used to assess the most appropriate model and the interpretation of results.

The fit statistics indicate that the fixed effects model and its assumptions are suitable. The difference between the model results are minimal, although the random effects model reports much wider credible intervals indicating greater uncertainty. As only one study is able to inform the estimate of BSC it is right that credible intervals might be very wide.

B.2.9.6. Validation

When the base case HR is used to estimate the BSC arm, estimated survival is slightly higher than the reported values (Table 7). This is not unexpected as the population in the literature have higher ECOG scores than those in ATTRACTION-3 (to which the HR is applied) and so outcomes would be expected to be worse (Table 25). Nomura et al. [23] reports that 23.5% of BSC patients were surviving at 6 months. The base case estimate is 48.23%, which is considerably higher, however, the base case population is considered to initiate with no patients having an ECOG score above 1. In contrast, the population reported by Nomura et al. [23] have 38% patients in the BSC arm with an ECOG PS of 2, which is strongly associated with reduced survival.8 Moriwaki et al.[21] reports survival estimates at 6 and 12 months that are slightly closer to those estimated by the NMA, though the median is much lower. Another study, Tsushima et al. [24], also reports a lower median than the base case estimate, but it does not report survival at any time points.

The results produced by the NMA are considered to be a reasonable estimate, when taking into consideration the nature of heterogeneity and intransitivity in the studies included in the NMA. Comparing the estimated survival at different time points suggests that the survival of patients receiving BSC may not be identical to that of patients receiving active treatment (HR applied to the docetaxel arm of ATTRACTION-3). Differences in the medians and survival at different time points can sometimes suggest that the shape of the survival curve and, therefore, the distribution of events may be different between active and BSC treatment.

Table 25. Validation of NMA outcomes

Time	**Base case**	**Moriwaki et al., 2014[21]**	**Nomura et al.,2016[23]**	**Tsushima et al., 2015[24]**
6 months	48.23%	40.00%	23.50%	-
12 months	18.08%	13.00%	5.90%	-
median	5.3	3.3 (3.6)	4.3 (4.26)	4.2

B.2.9.7. Conclusions

The results of the NMA indicate that BSC is less effective than docetaxel at preventing death events. However, there are uncertainties due to the limited number of reports that were able to be included into the NMA. Validation exercises show that the outcomes from the NMA are credible and that the uncertainty intervals around the point estimate are in line with the variation in reporting.

B.2.9.8. Uncertainties in the indirect treatment comparisons

It is important to note also that, while median values are available for all the studies, the follow up times are different. This is important because an incomplete or heavily censored KM curve may give a different HR value than if the data were complete.

There are several marked limitations of this analysis. Only one study informs the link to BSC, the heterogeneity in the included studies, and the application of an HR to nivolumab.

Having only one study to inform the relative efficacy of docetaxel is difficult because it increases uncertainty and relies on the study populations between Moriwaki et al. and ATTRACTION-3 to be the same.[21, 38] This is not upheld entirely. The docetaxel doses are slightly different as is the distribution of ECOG scores (particularly the proportion with an ECOG PS of 2) and the difference between recurrent and metastatic disease. A random effects model goes some way to adjust for these differences, but it is important to note that comparing studies that are not truly comparable may not result in robust estimates. Comparing estimates of survival from the base case to the reports in the literature suggests that, while the estimates may be reasonable and are fit for use in a cost-effectiveness analysis, the assumption that the distribution of events is the same as active treatment may not hold.

Another important limitation is the quality of the input studies. The included studies were all retrospective, aside from ATTRACTION-3. Therefore, patients included from these trials were not randomised and so this would not be considered high quality input data for analysis. While this does not mean they are uninformative, it should be considered while examining the outputs of analysis. This is often a limitation of any evidence synthesis in indications that are sparsely reported on.

It is also important to note that the evidence network is constructed to provide an estimate of BSC efficacy where direct evidence is not available. However, the patients in the retrospective trial taking BSC may be too frail to receive the chemotherapy option. Therefore, it may not be appropriate to suggest that these patients would receive nivolumab had it have been available. Given this limitation, the results should be considered with caution.

B.2.10. Adverse reactions

Key points

- Based on available evidence, nivolumab has an acceptable safety profile in patients with oesophageal cancer refractory intolerant to combination therapy with fluoropyrimidine- and platinum-based drugs.
- This safety profile is well-established based on that observed in other indications for nivolumab.

- The safety profile of nivolumab was improved over that for the taxanes: 65.6% of patients in the nivolumab arm reported a drug-related AE (grade 3-5: 18.2%) versus 95.2% for patients receiving paclitaxel or docetaxel (grade 3-5: 64.0%)

Safety data for nivolumab in advanced or recurrent unresectable oesophageal cancer are available from the ATTRACTION-3 and ATTRACTION-1 studies.

In general, nivolumab presents with a good safety profile, which is well characterised and consistent with other indications. In the pooled dataset of nivolumab 3 mg/kg as monotherapy across tumour types (n = 2578) with minimum follow-up ranging from 2.3 to 28 months, the most frequent adverse reactions (reported in ≥ 10% of patients) were fatigue (30%), rash (17%), pruritus (13%), diarrhoea (13%), and nausea (12%). The majority of adverse reactions were mild to moderate (Grade 1 or 2).[58]

B.2.10.1. ATTRACTION-3

Safety data from ATTRACTION-3 is available as of the database lock date of 12 November 2018, representing 18 months of follow up data (mean), based on the safety set population (SAF) comprising 210 and 209 patients in the nivolumab and chemotherapy arms, respectively.[38]

B.2.10.1.1. *Extent of exposure*

[redacted]

[redacted]**Table 26.ATTRACTION-3: Extent of nivolumab exposure**[38]

Variable	Nivolumab arm	Control arm		
		Total	Docetaxel	Paclitaxel
N	209	208	65	143
Number of doses received				
Mean (SD)	[redacted]	[redacted]	[redacted]	[redacted]
Median (Range)	[redacted]	[redacted]	[redacted]	[redacted]
Cumulative dose (mg/kg)				
Mean (SD)	[redacted]	[redacted]	[redacted]	[redacted]
Median (Range)	[redacted]	[redacted]	[redacted]	[redacted]
Relative dose intensity (n, %)				
<50%	[redacted]	[redacted]	[redacted]	[redacted]
50-<70%	[redacted]	[redacted]	[redacted]	[redacted]
70-<90%	[redacted]	[redacted]	[redacted]	[redacted]
90 - <110%	[redacted]	[redacted]	[redacted]	[redacted]
>=110	[redacted]	[redacted]	[redacted]	[redacted]
Number of Cycles				
Mean (SD)	[redacted]	[redacted]	[redacted]	[redacted]
Median (Range)	[redacted]	[redacted]	[redacted]	[redacted]
Duration of treatment (months)				
Mean (SD)	[redacted]	[redacted]	[redacted]	[redacted]
Median (Range)	[redacted]	[redacted]	[redacted]	[redacted]

B.2.10.1.2. *Overall adverse events*

[redacted][38]

Drug-related AEs (incidence ≥5) reported in patients treated with nivolumab or chemotherapy are summarised in Table 27. The only AE with a higher incidence (difference ≥5%) in patients treated with nivolumab than the patients treated with chemotherapy was hypothyroidism (8.1% in the nivolumab arm, 0.5% in the control arm). For all the remaining drug-related AEs a higher incidence was reported in the control arm.[38]

Grade 3-4 AEs were reported by 18.2% (38 patients) and 63% (198 patients). A summary of drug-related AEs impacting ≥5% of the patient population of ATTRACTION-3 for any grade

and grade 3-4 is provided in Table 30. In addition, drug-related select AEs impacting ≥5% of the patient population for any grade and grade 3-4 is provided in Table 31.[38]

Table 27. ATTRACTION-3: Frequency of patients with drug-related AEs with Incidence Rate >5% Classified by PT[38]

	Nivolumab	Control arm
	137 (65.6)	198 (95.2)
Rash	23 (11.0)	31 (14.9)
Diarrhoea	22 (10.5)	20 (9.6)
Hypothyroidism	17 (8.1)	1 (0.5)
Pruritus	17 (8.1)	11 (5.3
Decreased appetite	16 (7.7)	56 (26.9)
Fatigue	15 (7.2)	43 (20.7)
Pyrexia	15 (7.2)	17 (8.2)
Malaise	9 (4.3)	45 (21.6)
Anaemia	5 (2.4)	49 (23.6)
Stomatitis	5 (2.4)	25 (12.0)
Constipation	4 (1.9)	16 (7.7)
Lymphocyte count decreased	4 (1.9)	18 (8.7)
Nausea	4 (1.9)	34 (16.3)
Alopecia	3 (1.4)	98 (47.1)
Arthralgia	3 (1.4)	21 (10.1)
Dysgeusia	3 (1.4)	14 (6.7)
Myalgia	3 (1.4)	18 (8.7)
Neutrophil count decreased	3 (1.4)	76 (36.5)
Lung infection	3 (1.4)	11 (5.3)
White blood cell count decreased	2 (1.0)	72 (34.6)
Neutropenia	1 (0.5)	40 (19.2)
Peripheral sensory neuropathy	1 (0.5)	47 (22.6)
Vomiting	1 (0.5)	14 (6.7)
Febrile neutropenia	0	22 (10.6)
Leukopenia	0	17 (8.2)
Neuropathy peripheral	0	22 (10.6)
AEs and drug-related AEs occurring between the start date of the first administration of the product and 28 days after the last dose or the start date of subsequence anti-cancer therapy after the last dose whichever comes first were tabulated. Drug-related AEs were defined as any AEs with causal relationship with the product is "Related" or missing.		

Table 28. ATTRACTION-3: Summary of Adverse Events[38]

	Nivolumab arm (N =209)		Control arm (N =208)	
	Any grade	Grade 3-4	Any grade	Grade 3-4
Number of patients with AEs	[redacted]	[redacted]	[redacted]	[redacted]
Number of patients with SAEs	[redacted]	[redacted]	[redacted]	[redacted]
Number of patients with AEs leading to discontinuation of study treatment	[redacted]	[redacted]	[redacted]	[redacted]
Number of patients with AEs leading to dose delay	[redacted]	[redacted]	[redacted]	[redacted]
Number of patients with AEs leading to dose reduction	[redacted]	[redacted]	[redacted]	[redacted]
Number of patients with drug-related-AEs[i]	[redacted]	[redacted]	[redacted]	[redacted]
Number of patients with drug-related SAEs	[redacted]	[redacted]	[redacted]	[redacted]
Number of patients with drug-related AEs leading to discontinuation of study treatment	[redacted]	[redacted]	[redacted]	[redacted]
Number of patients with drug-related AEs leading to dose-delay	[redacted]	[redacted]	[redacted]	[redacted]
Number of patients with drug-related AEs leading to dose reduction	[redacted]	[redacted]	[redacted]	[redacted]
AEs, drug-related AEs occurring between the start date of the first administration of the investigational product and 28 days after the last dose or the start date of subsequence anti-cancer therapy after the last dose whichever comes first were tabulated [i]Drug-related AEs were defined as any AEs with causal relationship with the investigational product is “related” or missing				

Table 29. ATTRACTION-3: Summary of Deaths[38]

	Nivolumab	Control arm		
		Total	Docetaxel	Paclitaxel
N				
Number of patients who died				
Initial Disease				
Drug Toxicity				
Other Cancer				
Other				
Number of patients who died within 28 days of last dose				
Initial Disease				
Drug Toxicity				
Other Cancer				
Other				
Number of patients who died within 28 days of last dose or the start date of subsequence anti-cancer therapy after the last dose whichever comes first				
Initial Disease				
Drug Toxicity				
Other Cancer				
Other				
Number of patients who died within 100 days of last dose				
Initial Disease				
Drug Toxicity				
Other Cancer				
Other				

Table 30. ATTRACTION-3: Summary of drug-related adverse event profile impacting ≥5% of population[38]

AE (n, %)	Nivolumab arm (N =209)		Control arm (N =208)	
	Any grade	Grade 3-4	Any grade	Grade 3-4
Total	137 (65.6)	38 (18.2)	198 (95.2)	131 (63.0)
Anaemia	5 (2.4)	4 (1.9)	49 (23.6)	19 (9.1)
Decreased appetite	16 (7.7)	2 (1.0)	56 (26.9)	10 (4.8)
Febrile neutropenia	0	0	18 (27.7)	18 (27.7)
Leukopenia	0	0	3 (4.6)	2 (3.1)
Lymphocyte count decreased	4 (1.9)	2(1.0)	9 (13.8)	7 910.8)
Neutropenia	1 (0.5)	0	5 (3.5)	5 (3.5)
Neutrophil count decreased	3 (1.40	1 (0.5)	48 (33.6)	35 (24.5)
White blood cell count decreased	1 (0.5)	1 (0.5)	24 (11.5)	20 (9.6)
Drug-related AEs occurring between the start date of the first administration of the product and 28 days after the last dose or the start date of subsequence anti-cancer therapy after the last dose whichever comes first were tabulated. Drug-related AEs were defined as any AEs with causal relationship with the product is "Related" or missing.				

Table 31. ATTRACTION-3: Summary of drug-related select adverse event profile impacting ≥5% of population[38]

AE (n, %)	Nivolumab arm (N =209)		Control arm (N =208)	
	Any grade	Grade 3-4	Any grade	Grade 3-4
Alanine aminotransferase increased	[redacted]	[redacted]	[redacted]	[redacted]
Aspartate aminotransferase increased	[redacted]	[redacted]	[redacted]	[redacted]
Diarrhoea	[redacted]	[redacted]	[redacted]	[redacted]
Rash	[redacted]	[redacted]	[redacted]	[redacted]
Drug-related AEs occurring between the start date of the first administration of the product and 28 days after the last dose or the start date of subsequence anti-cancer therapy after the last dose whichever comes first were tabulated. Drug-related AEs were defined as any AEs with causal relationship with the product is "Related" or missing.				

B.2.10.1.3. *Discontinuation due to adverse events*

AEs leading to discontinuation of study treatment were reported in [redacted] in the nivolumab arm and [redacted] in the chemotherapy. Drug-related AEs leading to discontinuation of study treatment were reported in [redacted] and [redacted] respectively.

B.2.10.1.4. *Deaths*

As of the data cut-off date in November 2018, deaths of any cause during the study were reported in [redacted] in the nivolumab group and [redacted]) in the control group ([redacted] in the docetaxel arm and [redacted]] in the paclitaxel arm). The most common reason for death was “initial disease” in both groups

B.2.10.1.5. *Serious adverse events*

[redacted]

B.2.10.2. ATTRACTION-1

Safety data from ATTRACTION-1 is available as of the database lock date of 17 November 2016, based on the safety set population (SAF) comprising 65 patients who received nivolumab. [41]

B.2.10.2.1. *Extent of exposure*

All patients received at least one dose of nivolumab or chemotherapy. The majority of nivolumab-treated patients received at least 90% of the planned dose intensity. The median number of times treatment was received was 6. The median duration of treatment 78 days. The median number of cycles administered to patients was 3.0 (range 1-23).[41] Dose intensity and duration of treatment for both treatment arms are summarised in Table 32.

Table 32. ATTRACTION-1: Extent of exposure[41]

Variable	Nivolumab arm
Number of treatments (times)	
≤10	[redacted]
>10-20	[redacted]
>20	[redacted]
Mean (SD)	[redacted]
Median (Range)	[redacted]
Number of cycles	
Mean (SD)	[redacted]
Median (Range)	[redacted]
Duration of treatment (days)	
≤100	[redacted]
>100-200	[redacted]
>200	[redacted]
Mean (SD)	[redacted]
Median (Range)	[redacted]
Total dose (mg/kg)	
Number of patients	[redacted]
Mean (SD)	[redacted]
Median (Range)	[redacted]
Relative dose intensity (%)	
<50	[redacted]
50-<70	[redacted]
70-<90	[redacted]
90-<110	[redacted]
≥110	[redacted]
Mean (SD)	[redacted]
Median (Range)	[redacted]

B.2.10.2.2. *Overall adverse events*

AEs were reported in 56 (86.2%) of 65 patients, with grade 3−4 events reported in 19 (29.2%) and grade 3−4 SAEs in 12 (18.5%).[41] Treatment-related AEs were reported in 40 (61.5%) of 65 patients, with grade 3 or worse events being reported in 12 (18.5%). AEs and treatment-related AEs that led to discontinuation were reported in 10.8% (7 patients) each. Deaths within 28 days after the last dose or before the start of the post-study treatment after the last dose, were reported in 9.2% (6 patients). No patient died due to a treatment related AE (Table 34).[41] Incidence of AEs and serious AEs are summarised in Table 33 and Table 34. All 65 patients were assessable for safety. The most common AEs were diarrhoea, decreased appetite, dysgeusia and cough. The most common serious AEs were diarrhoea, dysgeusia, and decreased appetite.[41]

Table 33. ATTRACTION-1: Adverse events and treatment-related adverse events, Reported in ≥5% of patients by Grade[41]

	Adverse events (n, %) Safety population, N = 65			Treatment-related adverse events (n, %) Safety population, N = 65		
	Grade 1	Grade 2	Grade 3	Grade 1	Grade 2	Grade 3
Diarrhoea						
Decreased appetite						
Lung infection						
Cough						
Constipation						
Dysgeusia						
Fatigue						
Nasopharyngitis						
Rash						
Pneumonia						
Pruritus						
Vomiting						
Malaise						
Nausea						
Pyrexia						
Back pain						
Blood creatinine phopsphokinase increased						
Hepatic function abnormal						
Hypothyrodism						
Oedema						
Pain						
Infusion relation reaction						
Adverse events were classified with the Japan Clinical Oncology Group translation of the Common Terminology Criteria for Adverse Events (version 4.0). Treatment-related adverse events were defined as adverse events for which a causal relation to nivolumab could not be ruled out. Some patients had more than one event. No patients died due to adverse events.						

Table 34. ATTRACTION-1: Serious adverse events and serious treatment-related adverse events, Reported in ≥5% of patients[41]

	Serious adverse events (n, %) Safety population, N = 65		Serious treatment-related adverse events (n, %) Safety population, N = 65	
	All Grade	Grade 3-4	All Grade	Grade 3-4
Overall	12 (18.5)	11 (16.9)	9 (13.8)	8 (12.3)
Lung infection	4 (6.2)	4 (6.2)	2 (3.1)	2 (3.1)

Adverse event terms reported by physicians were coded according to MedDRA version 19.1J. Totalling of adverse events that occurred within 28 days after the last dose or before the start of the post-study treatment after the last dose, whichever was earlier, after the start of the treatment phase was conducted. An adverse drug reaction was defined as an adverse event for which a causal relationship to the investigational product could not be ruled out.

Table 35. ATTRACTION-1: Summary of Adverse Events and Death[41]

	Nivolumab arm (N =65)
	Any grade (n,%)
Number of patients with AEs	[redacted]
Number of patients with a Grade 3-4 AE	[redacted]
Number of patients with a SEA	[redacted]
Number of patients with an AE that led to discontinuation of study treatment	[redacted]
Number of patients with an AE that caused death	[redacted]
Number of patients with a treatment-related AE	[redacted]
Number of patients with a Grade 3-4 treatment related AE	[redacted]
Number of patients with an treatment-related SEA	[redacted]
Number of patients with an SEA that led to discontinuation of study treatment	[redacted]
Number of patients with a treatment-related AE that caused death	[redacted]
Number of death	[redacted]
AE: adverse events, SEA: serious adverse event AEs, treatment-related AEs occurring between the start date of the first administration of the investigational product and 28 days after the last dose or the start date of subsequence anti-cancer therapy after the last dose whichever comes first were tabulated [i]Drug-related AEs were defined as any AEs with causal relationship with the investigational product is "related" or missing	

B.2.11. Innovation

Nivolumab is a checkpoint inhibitor immunotherapy agent that provides an innovative mechanism of action that utilises the body's own immune system to destroy cancer cells (see Section B.1.3.5.1). Based on the innovative nature of nivolumab treatment, an application for PIM designation was submitted on 10th May 2017, which has since been granted by the MHRA on 10th July 2017 as being a promising candidate for the Early Access to Medicines Scheme in the treatment, diagnosis or prevention of life-threatening or seriously debilitating conditions with unmet need. Further, nivolumab is viewed by physicians and patients as a 'step-change' in the management of this stage of the disease.

The introduction of nivolumab would change the treatment paradigm for these patients and thus represents a 'step-change' in the management of OC following failure of prior line of therapy. The benefits of nivolumab include:

- **Improved survival outcomes:** Treatment options for OC patients who have failed first-line therapy are limited. Patients can be considered for different options of palliative treatment. However, due to the lack of evidence, second-line treatment discussions are highly individual for each patient and no specific chemotherapy regimens are currently recommended.[8-11] Nivolumab demonstrated a significant extension in overall survival (OS) versus chemotherapy in patients with unresectable

advanced or recurrent OC that is refractory to or intolerant of fluoropyrimidine plus platinum-based therapy.[38]

- **Maintenance of quality of life:** As described in Section B.2.6.1.9, nivolumab was associated with maintained quality of life from baseline in a generic health status measure (EQ-5D), demonstrating the quality of life benefit for nivolumab following treatment discontinuation and post- progression.[38]

- **Improved tolerability:** In comparison with the chemotherapy regimens received at second-line, the safety profile for nivolumab can be considered acceptable to patients, as described in Section 0, and is well-established based on that observed in other indications.[59] Further, nivolumab was found to have an acceptable safety profile when directly compared with chemotherapy in ATTRACTION-3; drug-related adverse events (AEs) of any grade led to the discontinuation of nivolumab treatment in 18 patients (8.6%) and with docetaxel and paclitaxel in 19 patients (9.1%) while drug-related adverse events (AEs) of grade 3-4 led to the discontinuation of nivolumab treatment in 8 patients (8.6%) and with docetaxel and paclitaxel in 12 patients (5.8%).

- **Facilitation of normal life:** Due to the improved quality of life and acceptable safety profile, nivolumab monotherapy has the potential to facilitate continuation of normal life, enabling patients to spend less time at hospital and more at home, which is of significant comfort to patients with advanced oesophageal cancer. Nivolumab requires administration once every two weeks, enabling patients to schedule outpatient attendances into their lives in a predictable manner. This is in comparison to BSC, where patients are likely to require additional ongoing management, which may be unpredictable. Furthermore, the improved tolerability compared to chemotherapy would translate to patients having to seek medical attention for adverse events less often.

- **Additional treatment option:** Current treatment options for OC cancer patients who failed a previous line of therapy are limited, with best BSC and palliative monotherapy likely to be the remaining option for managing squamous cell OC. However, outcomes from BSC and palliative monotherapy studies for pre-treated OC patients are poor. Nivolumab provides a treatment option with proven efficacy and tolerability in patients who may otherwise have been receiving chemotherapy based on limited evidence or only BSC, which would manage the symptoms of a patient's illness, but with limited impact on survival.

In summary, the availability of nivolumab provides an opportunity to make a significant and substantial impact on health-related benefits and addresses a current unmet need, and the adoption of nivolumab in this therapeutic indication in NHS England would represent a further, significant advance in the management of this life-threating condition.

B.2.12. Interpretation of clinical effectiveness and safety evidence

B.2.12.1. Principal findings from the clinical evidence

The clinical evidence supporting the use of nivolumab for advanced or metastatic squamous cell OC when standard therapy has failed was derived from ATTRACTION-3 and ATTRACTION-1. Primary clinical evidence was obtained from the randomised controlled trial, ATTRACTION-3. Supportive evidence is available from ATTRACTION-1.

The ATTRACTION-3 study is a Phase III, open-label, multi-centre, docetaxel/ paclitaxel-controlled study which demonstrates the benefits of nivolumab over chemotherapy in terms of response rate, survival and tolerability as described in Sections B.2.6. Based on the available data, benefits in OS for nivolumab over chemotherapy were observed from Month 6 through Month 30, with six-month OS of 71.9% for nivolumab and 63% for chemotherapy, corresponding to a tripling of OS at Month 30, at 16.3% and 4.8%, respectively. Median OS for patients treated with nivolumab was 10.91 months and 8.38 months for patients treated with chemotherapy.

Clinical trial data presented within this submission demonstrates significant survival improvements for nivolumab-treated patients and demonstrates the novel survival profile associated with immunotherapy agents. Although the patients in the nivolumab arm showed numerical lower values in the secondary endpoints (e.g. DCR and PFS) compared with patients treated with chemotherapy, the results demonstrate that the effect of nivolumab on patients who have responded to the treatment is likely to be sustained for a continued duration. This is in line with the long treatment effect of nivolumab already demonstrated in other indications. Therefore, the clinical meaning of nivolumab in prolonging survival and the inhibitory effect on disease progression shown in this study is significant.

In addition, a favourable tolerability profile was observed in nivolumab and none of the AE were detected as a newly identified risk of treatment with nivolumab. An incidence rate of drug-related AEs requiring a dose delay or dose reduction and affecting treatment was lower in nivolumab treatment compared with those in the conventional therapies (taxane agents).

Overall, nivolumab offers a favourable benefit-risk profile for patients with unrespectable, advanced OC when standard chemotherapy has failed.

B.2.12.2. Strengths and limitations of the clinical evidence base

The main limitations of the clinical evidence base are set out in Section B.2.12.2.1 while strengths of the evidence are outlined in Section B.2.12.2.2. However, these limitations should be viewed within the context of the study strengths and the high unmet need in this patient population.

B.2.12.2.1. *Limitations of study evidence*

Nivolumab clinical efficacy is informed using the two pivotal trials, ATTRACTION-3. There are inherent limitations with both studies. However, these limitations should be viewed within the context of the study strengths and the high unmet need in this patient population.

- Study location – Despite enrolment of patients globally, the majority of patients were from Asian countries. Although the limited number of patients from countries outside of Asia might limit the interpretation and external validity of results, analysis in Asian and non-Asian patients showed favourable survival outcomes for nivolumab compared with chemotherapy in both subgroups. An SLR evaluating differences in patient characteristics and survival outcomes between Asian and Western population with treatment experienced advanced OSCC was undertaken. A detailed description of outcomes for Asian and non-Asian patients is provided in Section B.2.12.4.1.1

- Open-label study design – The open-label study design of ATTRACTION-3 means that there is a possibility the knowledge of the treatment might have influenced patient responses with regards to health-related quality of life. However, an open-label design was considered appropriate because of the differences in the dosing regimens and associated toxicities for each treatment group. The primary endpoint of overall survival is an objective measure, which would not be affected by the open-label nature of the study. Furthermore, involvement of an independent data monitoring committee for safety assessments ensured anonymity of the treatment groups during data review

B.2.12.2.2. *Strengths of study evidence*

ATTRACTION-3 is a well-designed, Phase III randomised controlled trial which provide direct comparative evidence on the clinically efficacy of nivolumab versus chemotherapy. The sizes of the patient cohorts were large (210 and 209 in the nivolumab and chemotherapy arms, respectively) and all patients had received prior therapies, consistent with the current indication. Patient-reported outcomes are available from ATTRACTION-3, where QoL was assessed through collection of EQ-5D data, providing utility estimates which are directly attributable to nivolumab treatment. In addition, ATTRACTION-3 provides survival data which may be considered relatively mature, placing less reliance on the need for survival extrapolation though parametric curve fitting.

The most important treatment outcomes for most oesophageal cancer patients include OS, reduced side effects, improved symptom control and quality of life. Nivolumab provides significant benefits for each of these outcomes:

- Improved survival outcomes

- Maintained quality of life

- Tolerability

The safety and efficacy of nivolumab are of particular importance in the setting of previously treated advanced or recurrent unresectable oesophageal cancer that is refractory or intolerant

to chemotherapy where there has been a lack of new treatments, specifically those with a favourable safety profile, as well as improved efficacy. Following the failure of one or more prior chemotherapy regimens, therapeutic options are severely limited, and additional salvage chemotherapeutic options may not be available to all patients due to tolerability issues, especially in elderly patients with existing comorbidities. In this setting, nivolumab may be a well-tolerated therapeutic option with the potential to offer significant survival benefit in this patient population. The availability of nivolumab would provide an opportunity to make a significant and substantial impact on health-related benefits and address a current unmet need.

The safety profile of nivolumab in esophageal cancer was considered to be almost similar to that of previously approved solid tumor indications, and nivolumab can be acceptable and manageable when properly used by doctors with sufficient knowledge and experience in cancer chemotherapy, in medical facilities that could sufficiently respond to emergencies The results demonstrated that nivolumab could become one of treatment options that would be beneficial for the disease as the benefits outweigh the risks of nivolumab treatment in patients with unresectable, advanced or recurrent esophageal cancer progressing after cancer chemotherapy, and that the clinical benefit of nivolumab is significant.

B.2.12.3. Relevance of the evidence base to the decision problem

The submission presents two studies, one of which is a docetaxel and paclitaxel-randomised trial, evaluating the efficacy of nivolumab in metastatic or advanced oesophageal cancer who received at least one prior line of therapy, in line with the decision problem. Indirect comparison analyses applying different methodologies are presented to provide supportive evidence of comparative effectiveness. These comparisons underscore the clinical efficacy of nivolumab and provide additional certainty around the beneficial impact of nivolumab in a Western patient population. Further, outcomes considered in the submission closely mirror the decision problem set out by NICE.

The evidence base presented within this submission represents the best available evidence and is directly relevant to the decision problem.

B.2.12.4. External validity of study results to patients in routine clinical practice

Patients enrolled in the available studies can be considered broadly representative of UK practice, in terms of baseline characteristics, with subgroups provided for analysis where possible.

B.2.12.4.1.1. Relevance to UK patient population

As outlined in Section B.2.12.2.1, despite enrolment of patients globally, the majority of patients in ATTRACTION-3 were from Asian countries. Although the limited number of patients from countries outside of Asia might limit the interpretation and external validity of results, analysis in Asian and non-Asian patients showed favourable survival outcomes for nivolumab compared with chemotherapy in both subgroups.

An SLR evaluating differences in patient characteristics and survival outcomes between Asian and Western population with treatment experienced advanced OSCC was undertaken. Results indicated that OS was comparable between Asian and Western populations with OSCC (median: 7.5 versus 7.4 months); mean one-year OS was 21.1% in Asian and 27.9% in Western patients. Longer OS was observed in Asian patients in the overall population (OSCC and oesophageal adenocarcinoma [OADC]; median: 8.1 versus 5.7 months for Western patients). These results observed in Western populations were driven by poor outcomes in Western OADC patients than in Western OSCC patients (5.6 versus 7.4 months); no data was identified for Asian patients with OADC. These results suggest that survival in patients with OSCC was comparable between Asian and Western populations.

Based on this evidence, it can be concluded that the ATTRACTION-3 patient population differed from the UK population in terms of ethnicity. However, this did not have an impact in terms of a difference between the patient subgroup, which can be expected in light of the published evidence in this patient population.

B.2.12.4.1.2. UK standard of care

As outlined in Section B.1.3.4, UK guidelines recommend chemotherapy for patients who have progressed on first-line therapy; however, specific chemotherapy regimens are not defined in NICE clinical guidelines in the second-line setting.[8-10] ESMO guidelines recommend taxane monotherapy for the second-line treatment (after failure of first-line treatment with taxane combination therapy) of OC.[11] Clinical expert opinion obtained during a clinical advisory board meeting supported evidence on a lack of standard of care for treatment-experienced OC patients. In the second line setting, decisions on treatment options for unresectable, advanced or metastatic OC patients were described as highly individualised, and chemotherapy agents such as paclitaxel or docetaxel are usually the treatment of choice in this setting.

The ATTRACTION-3 study included a taxane comparator arm, comprising docetaxel and paclitaxel. Outcomes are relatively comparable between docetaxel and paclitaxel in this setting, with median OS of 7.62 months for docetaxel versus 8.51 months for paclitaxel, while OS at 12 months was 34.6% and 34.2% respectively. It is acknowledged that low patient numbers receiving individual treatments may impact on outcomes, particularly during later periods of follow-up. Hence, the combined ATTRACTION-3 control is a relevant comparator to the UK setting for treatment of previously treated OSCC patients.

BSC can also be considered a relevant comparator in patients unable to receive alternative therapies. Although there is no direct comparative evidence for nivolumab versus BSC, ITC evidence has been provided to inform comparative efficacy.

The NICE scope includes irinotecan as a potential comparator. However, advice obtained from clinical experts during an advisory board confirmed that irinotecan is currently not routinely used in UK clinical practice, so that clinicians did not consider it a relevant comparator. BMS market research conducted in 2019 estimated that only irinotecan comprised only 6% of usage for OSCC patients who had received previous treatment. Further, there is a lack of clinical evidence identified to support use, based on the clinical SLR described in Appendix D.

B.2.12.5. Application of NICE end-of-life criteria to nivolumab use in oesophageal cancer

Outcomes are known to be poor in oesophageal cancer patients with unresectable and advanced disease when standard first-line chemotherapy has failed, although there is a paucity of evidence describing this patient population. These patients have highly limited treatment options remaining and estimates of OS at 1 year are around 34.4% (as reported in chemotherapy patients from ATTRACTION-3.[38] Thus, there is a high degree of unmet medical need in this patient population, which would be addressed by availability of nivolumab.

The case for application of NICE end-of-life criteria to nivolumab use in oesophageal cancer is set out in Table 36., and based on this evidence, nivolumab is considered to meet both criteria for end-of-life.

Table 36. End-of-life criteria

Criterion	Data available	Submission reference
The treatment is indicated for patients with a short life expectancy, normally less than 24 months	Available therapies in patients with unresectable, advanced oesophageal cancer when standard chemotherapy has failed are associated with poor outcomes, although data describing this patient population are limited. Based on available data, median OS for combined taxanes, as estimated by the model was 12.0 months.	Section B.2.6.1.9, B.3.3.2.1.3 and B.3.7.1
There is sufficient evidence to indicate that the treatment offers an extension to life, normally of at least an additional 3 months, compared with current NHS treatment	The mean OS is more representative of the survival benefit associated with nivolumab. However, it is acknowledged that extrapolated output is subject to uncertainty, due to the potential variation in extrapolations. However, when data are restricted to the observed period, restricted mean OS is 14.06 months in the nivolumab arm and 11.48 months in the taxane arm. Although this does not account for the long-term divergence in survival benefit outside of the observed trial period, the three month survival benefit criteria is almost met (improvement of 2.58 months). Hence, there is relative certainty that the criteria will be met during longer follow up. Based on model output, mean OS extrapolated over a life-time horizon was 19.8 months in the nivolumab arm and 12.0 months in the control arm (an improvement of 7.8 months). Based on this evidence, it can be concluded that end-of-life criteria are met.	Section B.2.6.1.9, B.3.3.2.1.3 and B.3.7.1

B.3 Cost effectiveness

Base case analysis

- In line with estimates of short life expectancy in patients receiving taxanes, the base case analysis predicts median OS of 0.75 years (mean 1.00 years), informed by a randomised-controlled trial.
- Use of nivolumab will result in an increased mean OS of 1.65 years, as well as additional discounted QALYs and life years of [redacted] and [redacted], respectively.
- Based on mean OS outcomes for patients treated with taxanes (12.0 months) and mean OS benefit associated with nivolumab (incremental 7.8 months), end of life criteria can be considered to be met.
- Discounted incremental costs were estimated to be £20,842 under base case assumptions and the resultant ICER was £45,491 per QALY, which is considered to be cost-effective at a willingness-to-pay threshold of £50,000 per QALY.

Sensitivity analysis

- In the deterministic and probabilistic sensitivity analyses, nivolumab was cost-effective in the majority of scenarios at a willingness-to-pay threshold of £50,000 per QALY.
- Extensive scenario analyses were undertaken, reflecting the assumptions required to undertake plausible, robust and transparent base case analysis.
- Within these scenario analyses, the majority of ICERs remain below the £50,000 per QALY threshold.

B.3.1. Published cost-effectiveness studies

In line with the NICE Guide to the methods of technology appraisal 2013,[60] an SLR was conducted to identify cost-effectiveness studies for the treatment of previously treated advanced or recurrent unresectable oesophageal cancer that is refractory or intolerant to chemotherapy. In brief, electronic database searches (MEDLINE, Embase, the Cochrane library and EconLit) were conducted in December 2017, and subsequently updated in October 2018 and February 2020. Publications describing full economic evaluations of interventions aimed at managing previously treated advanced or recurrent unresectable oesophageal cancer that is refractory or intolerant to chemotherapy were included.

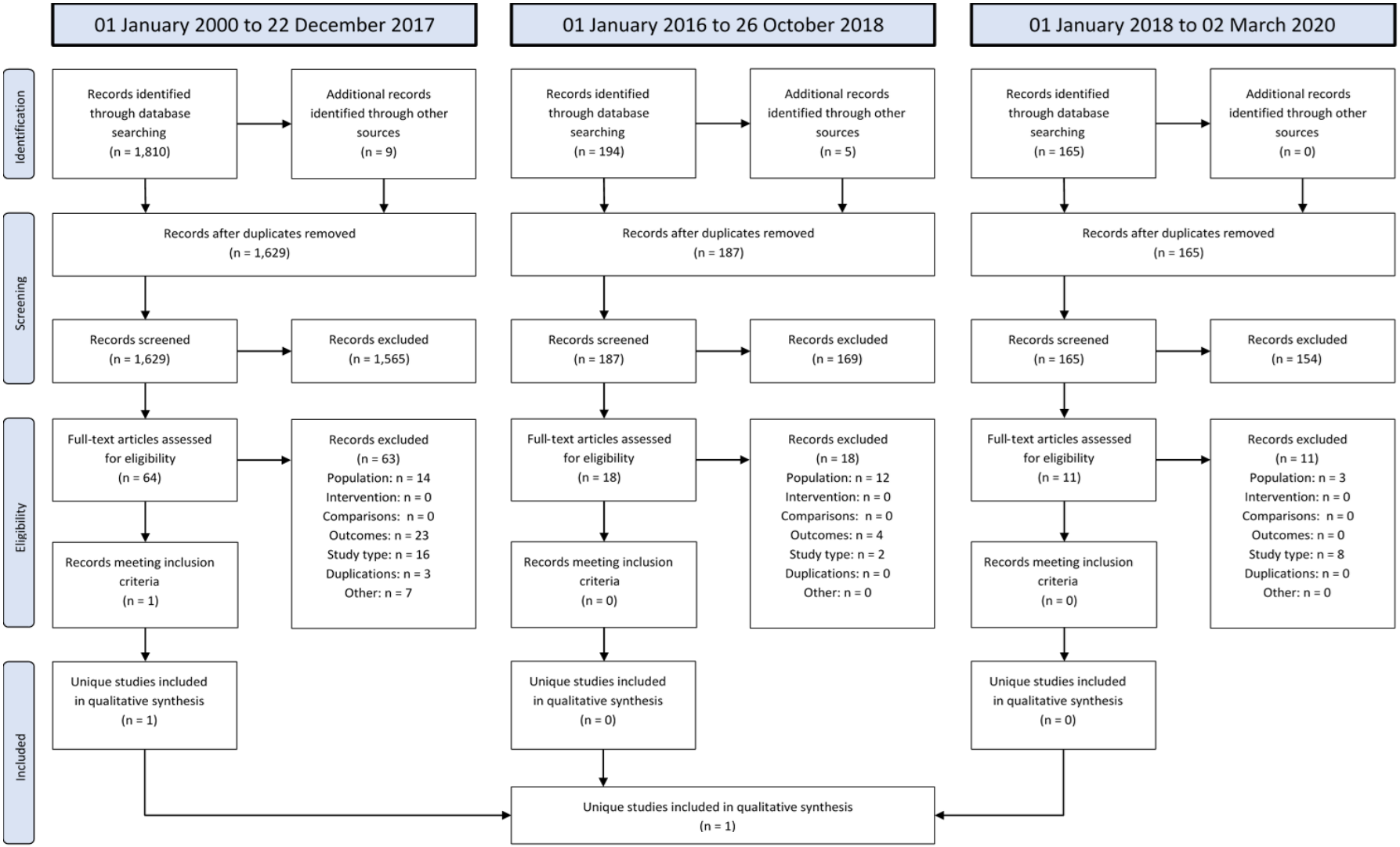


Figure 20. PRISMA diagram illustrating the study selection process for identifying cost-effectiveness studies for the period from 01 January 2000 to 02 March 2020

Table 37. Study characteristics of economic modelling studies of patients with advanced OC (n=8)

Model structure	Intervention	Patient population	Model inputs		Outcomes reported
			Utilities	Costs	
Janmaat (2016)[61] - The Netherlands					
Linear cost-effectiveness model, Monte Carlo sensitivity analysis: Two clinical outcome measures (PFS and OS)	Cisplatin and fluorouracil, either alone or in combination with cetuximab	≥18 years old pts with histologically confirmed and EGFR-expressing advanced, nonresectable ESCC. No prior neoadjuvant therapy (within 6 months) and no prior chemotherapy for recurrent or metastatic disease	0.675 utility at diagnosis It was assumed that utility score decreases linearly from 0.675 during progressive disease to 0 at time of death	Costs were in Euro (€). Unless otherwise specified, unit costs were obtained from the Dutch manual for cost-effectiveness research 2010. All costs and effects were converted to the price level of 2009 according to the general Dutch consumer price index. Cost of Cetuximab 2009: €237.20 Cost of outpatient visit: €251 Cost of evaluation of EGFR expression: €750	Adding cetuximab to standard chemotherapy: 0.187 LYs and 0.105 QALYs. Mean incremental cost: €26,459 per treated patient Adding cetuximab to cisplatin-5-flurorouracil 1st line regimen for advanced ESCC resulted in a mean ICER of €252,203 per QALY

AE: adverse event; AJCC: the American Joint Committee on Cancer; AUD: Australian dollar; c: cost; CRT: Chemoradiation therapy; CT: computed tomography; DFS: disease-free state; EGFR: epidermal growth factor receptor; EQ-5D: EuroQol 5-dimensions; ESCC: esophageal squamous cell carcinoma; EUS: endoscopic ultrasound; FNA: fine needle aspiration; GBP: British pound sterling; I$: international dollars; ICER: incremental cost-effectiveness ratio; INMB: incremental net monetary benefit; LYs: life years; NCCRT: Neoadjuvant concurrent chemoradiotherapy; NCRT: Neoadjuvant chemoradiotherapy; NR: not reported; OC: oesophageal cancer; OS: overall survival; PCA: prescription cost analysis; PET-CT: positron emission tomography – computed tomography; PFS: progression-free survival; QALY: quality-adjusted life years; RT: radiotherapy; S: surgery; SF-12: short-form 12 questionnaire; USD: USA dollar.

B.3.2. Economic analysis

The economic case presented in this submission is based on conventional cost-utility analysis, assessing the use of nivolumab versus taxanes for the treatment of unresectable advanced, recurrent or metastatic oesophageal squamous cell carcinoma that is refractory or intolerant to fluoropyrimidine and platinum-based combination therapy, taking into account a simple discount patient access scheme (PAS) for nivolumab.

A partitioned survival model structure has been utilised. The economic modelling of nivolumab and the comparator in this particular indication does not require extensive complexity with regard to subsequent lines of treatment or time-dependency of model inputs, which may necessitate use of a Markov model. Further, a partitioned survival model may replicate survival outcomes with a higher degree of accuracy compared with a Markov model, although differences in outcomes should be minimal, particularly where appropriate transition rates have been derived.[62]

The model utilises three health states (pre-progression, post-progression and death) to reflect disease progression, and the subsequent cost and utility consequences of different health states; in line with clinical practice, patients may receive treatment beyond progression. The model structure has been chosen to reflect the most important treatment outcomes for most oesophageal cancer patients: survival (progression free and overall), side effects, symptom control and quality of life. Survival curves have been applied to estimate PFS and OS in each treatment arm, while health state utilities and costs have been applied to reflect the symptom control and quality of life experienced by patients receiving nivolumab or taxanes. Treatment-specific AE probabilities, alongside AE event-specific costs, are used to estimate the incidence and economic consequences associated with treatment-related AEs (Section B.3.3.2.4)

Of note, the structure of the partitioned survival model accommodates treatment discontinuation and subsequent lines of therapy. This is of importance in the appraisal of nivolumab, where therapies may be continued beyond progression, subject to a stopping rule or discontinued upon disease progression.

B.3.2.1. Description of analyses

Within this submission, ATTRACTION-3 has been used to inform decision making and provide certainty around the beneficial clinical impact of nivolumab in oesophageal cancer in the UK. ATTRACTION-3 has been used to inform comparative efficacy in the base case analysis, as this is a Phase III randomised controlled trial providing direct evidence for nivolumab versus taxanes, and so can be considered the best available evidence. All analyses within this submission have been conducted from the payer perspective, in this case the NHS.

B.3.2.2. Patient population

This economic evaluation considers the use of nivolumab for the treatment of unresectable advanced, recurrent or metastatic oesophageal squamous cell carcinoma that is refractory or intolerant to fluoropyrimidine and platinum-based combination therapy, in line with the anticipated licensed indication.

A outlined in Section B.1.3.4, UK guidelines recommend chemotherapy for patients who have progressed on first-line therapy; however, specific chemotherapy regimens are not defined in the NICE clinical guidelines in the second-line setting.[8-10] ESMO guidelines recommend taxane monotherapy for the second-line treatment (after failure of first-line treatment with taxane combination therapy) of OC.[11] Clinical expert opinion obtained during a clinical advisory board meeting supported evidence on a lack of standard of care for treatment-experienced OC patients. In the second line setting, decisions on treatment options for unresectable, advanced or metastatic OC patients in terms of efficacy and toxicity were described as highly individual and chemotherapy agents such as paclitaxel or docetaxel are usual the treatment of choice in this setting. Thus, the comparators applied in ATTRACTION-3, docetaxel and paclitaxel, represent current UK standard of care as second-line treatment for OC patients. ATTRACTION-3 was powered to show differences in efficacy for nivolumab against the combined taxane arm, as opposed to docetaxel and paclitaxel, thus providing justification for using combined taxanes as the main comparator. As a scenario analysis, a comparison for nivolumab against docetaxel and paclitaxel separately is also conducted. In addition, a scenario for comparing nivolumab and BSC is conducted using ITC to inform the comparator efficacy. Based on a lack of evidence for use in clinical practice, irinotecan is not considered as an appropriate comparator. This is also supported by a lack of evidence found within the conducted SLR.

In the base case analysis, baseline patient parameters are derived from the baseline characteristics of patients enrolled in ATTRACTION-3, as detailed in Table 38.

Table 38. Baseline parameters

Parameter	Mean	SE	Source
Base case analysis			
Age (years)	63.82	0.45	ATTRACTION-3[38] patient-level data
Proportion of cohort male	0.869	0.016	
Cohort size	1,000	-	Assumption

B.3.2.3. Model structure

A de novo partitioned survival model was developed, applying health states representing pre-progression, post-progression and death (Figure 21). Unlike a Markov model, the number of people in any state at successive points in time is not dictated by transition probabilities. Instead, the model estimates the proportion of a cohort in each state based upon parametric or semi-parametric survival equations. These health states reflect disease severity and determine use of healthcare resources, health-related quality of life and mortality rates. To reflect the nature of oesophageal cancer and available evidence, the model assumes that oesophageal cancer phases are consecutive, which means patients are not able to revert to pre-progression from more advanced phases of the disease; this assumption has been validated by clinicians.[63]

Using a weekly cycle length, the model predicts the proportion of the population who experience a progression or death event. Weekly cycles were considered appropriate for this evaluation because it enables the model to reflect the timings of drug administrations associated with both nivolumab and comparator therapies. Weekly cycles further capture a realistic minimum time during which the symptoms or responses can change in UK clinical practice.

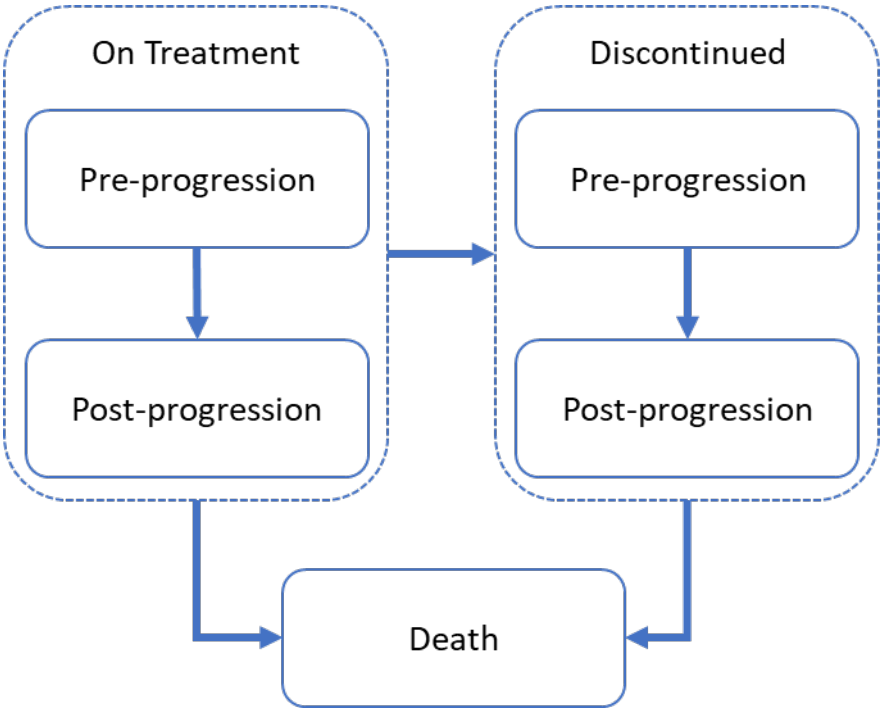


Figure 21. Conceptual model schematic

B.3.2.3.1. *Derivation of health state occupancy estimates*

Health state occupancy is defined by treatment-specific PFS and OS extrapolations, derived from available data (as described in Section B.3.3.2). An overview of model implementation of survival curves is presented in Figure 22.

As these PFS and OS data implicitly include the effects of any subsequent treatment that may have been administered, the need to explicitly incorporate the survival effects of these subsequent treatments is negated. Due to the short life expectancy in patients, a treatment waning effect is deemed in appropriate.

For nivolumab and taxanes, parametric curves for PFS and OS were fitted using patient-level data from the patient cohort in ATTRACTION-3; methods for deriving these curves are provided in Section B.3.3.2. For the scenario analyses, the same methodology was applied for deriving data for docetaxel and paclitaxel, separately. Data for the BSC comparator is derived from the SLR and ITCs described in Section B.2.9.

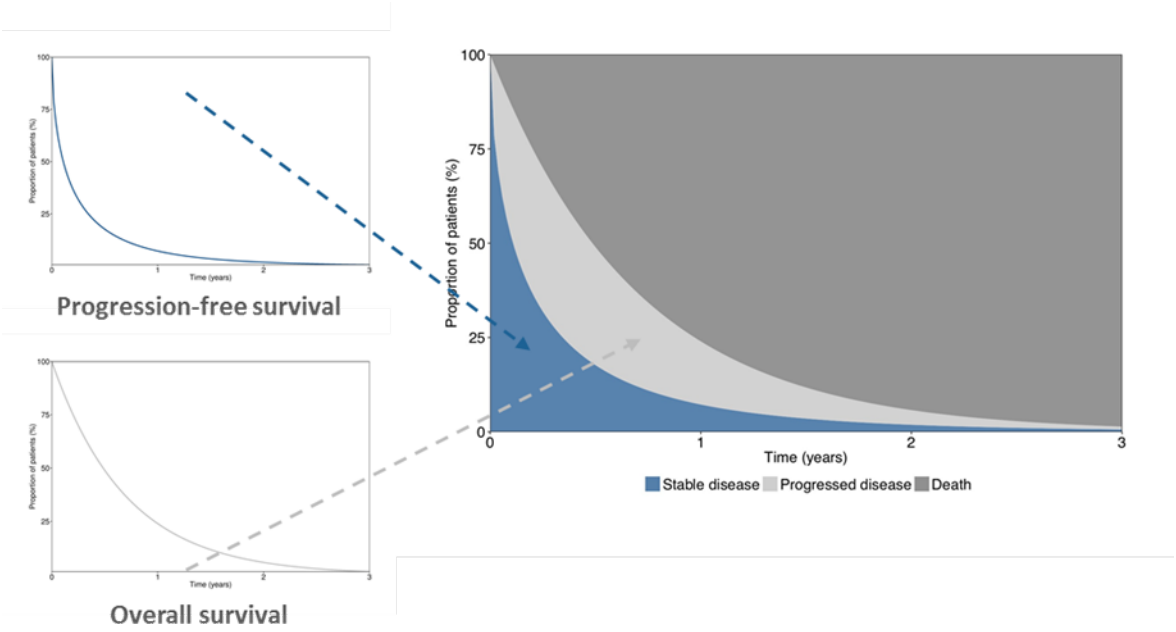


Figure 22. Overview of survival curve implementation in the model

B.3.2.3.2. *Derivation of Treatment Line Occupancy*

Patients enter the model following failure of prior therapy and can receive nivolumab or a comparator treatment. Following treatment cessation or progression, patients can receive a subsequent therapy, as detailed in Section B.3.5.1.5. However, as a simplifying assumption, it is assumed that patients may not discontinue this final line of therapy, as it is assumed to be comprised of all possible therapies that patients may subsequently receive, either sequentially or concurrently.

In clinical practice, treatment cessation may be caused by loss of clinical benefit or may be related to other factors, such as AEs. Clinicians may choose to cease treatment on progression, treat beyond progression or may choose to undertake a stopping rule, in line with previous nivolumab indications. Hence, the proportion of patients on initial or subsequent treatment lines is based on one of the following criteria:

- Base case analysis: All-cause discontinuation (excluding discontinuation due to progression) based on ATTRACTION-3 discontinuation rate data,
- Scenario analysis: Treatment cessation (where treatment duration is specified, for example in stopping rules),
- Scenario analysis: Disease progression in addition to discontinuation due to AEs.

B.3.2.3.3. *Treatment sequences*

Patients enter the model following failure of prior therapy and can receive nivolumab or a comparator treatment. Following treatment discontinuation, patients in both arms can receive subsequent therapy, described in Section B.3.5.1.5. It is assumed that this subsequent therapy is BSC, with composition and cost derived from the clinician survey described in Section B.3.5.1.5.

B.3.2.3.4. *Outcome measures*

The primary model output is the incremental cost-effectiveness ratio (ICER) expressed as incremental costs per QALY gained. Additionally, the model provides an overview of other outcomes, such as life years gained, and clinically relevant outcomes, such as predicted median OS and PFS.

No previous NICE Technology Appraisals have been identified for oesophageal cancer therapies. Table 39 provides a comparison versus a previous appraisal for gastric cancer in previously treated patients.

Table 39. Features of the economic analysis

Factor	Current appraisal		Previous appraisal
	Chosen values	Justification	TA378[64] (ramucirumab)
Time horizon	Lifetime (up to 40 years or 2,080 weeks)	This ensures that all events have occurred, and all patients are accounted for. However, a shorter time horizon is assessed in sensitivity analysis.	Lifetime (~7 years)
Treatment waning effect	None	This is in line with previous NICE appraisals.[65] Additionally, due to the short life expectancy in these patients, a treatment waning effects is deemed in appropriate.	None
Source of utilities	ATTRACTION-3 provides EQ-5D-3L data that can be used to derive utility inputs for use in nivolumab and comparator arms.	ATTRACTION-3 collected utility data using the EQ-5D-3L. In line with the NICE reference case, trial utilities collected as part of ATTRACTION-3 (baseline and every 6 weeks until the end of the treatment phase and subsequently ever 12 weeks during the follow-up phase) have been applied in the base case analysis for both treatments.	Pre- and post-progression health state utility values obtained from EQ-5D data from RAINBOW trial
Source of costs	As per TA378[64]	This TA is relevant to the licensed indication for nivolumab and applying these values will facilitate cross comparison between the TAs.	Costs of intervention and comparators included drug acquisition, administration and monitoring costs and costs of tests. Costs of available generic chemotherapies were sourced from the electronic market information tool which uses the actual price paid by hospitals over the last 12 months. Costs of BSC were identified from a review of hospital medical records. Further costs consisted of follow-up, adverse event, hospitalisation, third-line therapy (drug costs, administration and follow-up care), terminal care costs and adverse events.

B.3.2.4. Intervention technology and comparators

As outlined in Section B.1.3.4, UK guidelines recommend chemotherapy for patients who have progressed on first-line therapy; however, specific chemotherapy regimens are not defined in NICE clinical guidelines in the second-line setting.[8-10] ESMO guidelines recommend taxane monotherapy for the second-line treatment (after failure of first-line treatment with taxane combination therapy) of OC.[11] Clinical expert opinion obtained during a clinical advisory board meeting supported evidence on a lack of standard of care for treatment-experienced OC patients. In the second line setting, decisions on treatment options for unresectable, advanced or metastatic OC patients were described as highly individualised, and chemotherapy agents such as paclitaxel or docetaxel are usually the treatment of choice in this setting. Further, published clinical outcomes are comparable between docetaxel and paclitaxel in this setting.

The ATTRACTION-3 study included a taxane comparator arm, comprising docetaxel and paclitaxel. The trial was powered to show differences in efficacy for nivolumab against the combined taxane arm, as opposed to docetaxel and paclitaxel separately. Low patient numbers receiving individual treatments may impact on outcomes, particularly during later periods of follow-up. Hence, it is more appropriate to use the combined taxane arm as a comparator. However, a comparison of nivolumab against docetaxel and paclitaxel separately is provided as a scenario analysis.

In line with the NICE scope, is provided comparing nivolumab and BSC, using ITC evidence to inform the comparator efficacy. However, BSC is only a valid comparator in patients unable to receive alternative therapies. Further, there is no direct comparative evidence for nivolumab versus BSC. For these reasons, it is appropriate to provide this as a scenario analysis only.

The NICE scope includes irinotecan as a potential comparator. However, advice obtained from clinical experts during an advisory board confirmed that irinotecan is currently not routinely used in UK clinical practice, so that clinicians did not consider it a relevant comparator. BMS market research conducted in 2019 estimated that only irinotecan comprised only 6% of usage for OSCC patients who had received previous treatment. Further, there is a lack of clinical evidence identified to support use, based on the clinical SLR described in Appendix D.

B.3.3. Clinical parameters and variables

B.3.3.1. Evidence synthesis

Evidence to describe the effectiveness of nivolumab for the treatment of unresectable advanced, recurrent or metastatic oesophageal squamous cell carcinoma that is refractory or intolerant to fluoropyrimidine and platinum-based combination therapy is primarily derived from ATTRACTION-3, a randomised docetaxel/paclitaxel-controlled, phase III study evaluating nivolumab as monotherapy for the treatment of unresectable advanced or recurrent OC. In the base case analysis, nivolumab efficacy has been derived from the nivolumab arm of ATTRACTION-3, while taxane efficacy has been derived from the combined taxane arm.

B.3.3.2. Parameterisation of overall survival and progression-free survival

B.3.3.2.1. *Base case analysis; ATTRACTION-3*

B.3.3.2.1.1. Survival analysis approach

Clinical data to inform the base case analysis can be derived from ATTRACTION-3. However, follow-up was substantially less than the 40-year time horizon of the model. Therefore, parametric extrapolation of survival data from the study was required to inform long-term outcomes, undertaken with reference to the guidance from the NICE Decision Support Unit (DSU)[66] and Bagust and Beale (2014).[67]

A full description of methods used to undertake parametric extrapolation is provided in Appendix D.1.3.4. In brief, parametric functions that inform survival curves were developed using patient-level data from ATTRACTION-3 12 November 2018 database lock.

Progression events were based on investigator-assessed outcomes from ATTRACTION-3 and were defined as in this study. Death events from ATTRACTION-3 were used to inform OS modelling. Parametric survival functions were fitted to the extracted pooled data using the R statistics environment, including exponential, Weibull, log-logistic, lognormal, Gompertz and generalised-gamma survival distributions. Additionally, spline models were considered, as well as semi-parametric models assessing the impact of different split points and subsequent parametric functions, in line with the approach taken in recent appraisals of immuno-oncology agents.[68, 69]

Goodness-of-fit was evaluated using the Akaike and Bayesian Information Criteria (AIC and BIC, respectively); minimisation of these measures is used to indicate goodness-of-fit whilst penalising overfitting, so that a smaller value demonstrates a more appropriate fit. In addition to assessment of goodness-of-fit statistics, the appropriateness of the parametric extrapolation was evaluated by visual inspection of the fit over the observed period and consideration of the log cumulative hazard plots.

It is worth noting that while the above methods for validating the extrapolation of progression and death events are appropriate, they are also necessarily constrained by derivation from observed data, which is, as previously indicated, limited by the availability of follow-up data. Therefore, the plausibility of the extrapolation was assessed through consideration of the long-term hazard profile and the extrapolated mean survival estimates.

A more detailed description of survival extrapolation and outcomes is provided in Appendix D.1.3.4. In summary, Kaplan-Meier plots describing PFS and OS in the nivolumab and taxane arms demonstrated a high initial hazard during the initial study period, with a significant number of events occurring immediately after study entry, perhaps reflecting the high mortality impacting patients with oesophageal cancer. This was followed by a lower hazard in the longer-term in both study arms. Parametric models didn't adequately reflect this change in hazard. By contrast, a semi-parametric approach was considered appropriate as it reflected

the high initial hazard but applied the maximum amount of data to inform the long-term extrapolation.

Applying Kaplan-Meier data until 2.99 months followed by parametric extrapolation enabled the initial hazard to be modelled appropriately and captured the high rate of events between study entry and second assessment, which was scheduled for 12 weeks. Switching to parametric extrapolation from 2.99 months used the maximum number of events to inform long-term extrapolation and describe the lower long-term hazard. This semi-parametric approach was applied for both PFS and OS in both nivolumab and taxane arms.

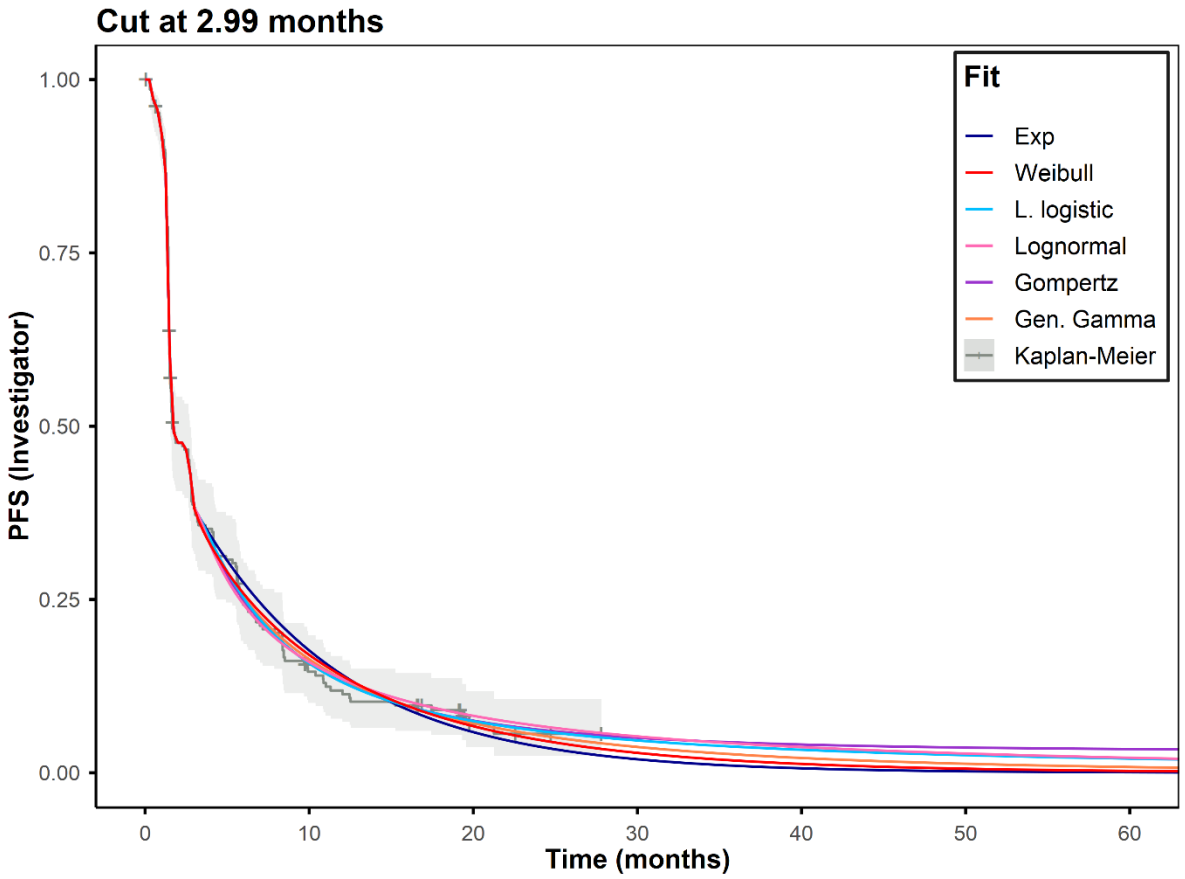
In order to model PFS in the nivolumab arm, Kaplan-Meier data was applied until 2.99 months followed by parametric extrapolation using the Weibull distribution to provide an appropriate fit. This approach predicted a median PFS of 1.7 months (observed 1.7 months) and a mean PFS of 5.8 months.[38] Similarly, a semi-parametric approach was considered to be most appropriate for modelling OS, where Kaplan-Meier data was applied until 2.99 months followed by parametric extrapolation using the log-logistic distribution. This approach was deemed appropriate as it provided an adequate fit to the data, providing a median OS of 10.9 months (observed 10.9 months) and a mean OS of 24.3 months.[38]

In order to model PFS in the taxane arm, Kaplan-Meier data was applied until 2.99 months followed by parametric extrapolation using the Weibull distribution to provide an appropriate fit. This approach predicted a median PFS of 3.3 months (observed 3.4 months) and a mean PFS of 4.8 months.[38] Similarly, a semi-parametric approach was considered to be most appropriate for modelling OS, where Kaplan-Meier data was applied until 2.99 months followed by parametric extrapolation using the exponential distribution. This approach was deemed appropriate as it provided an adequate fit to the data, providing a median OS of 8.9 months (observed 8.4 months) and a mean OS of 12.0 months.[38]

A summary of survival outcomes following extrapolation is provided in Table 40. Parametric extrapolation following the split points at 2.99 months for nivolumab and taxanes PFS and OS from ATTRACTION-3 are shown in Figure 23 through Figure 26.

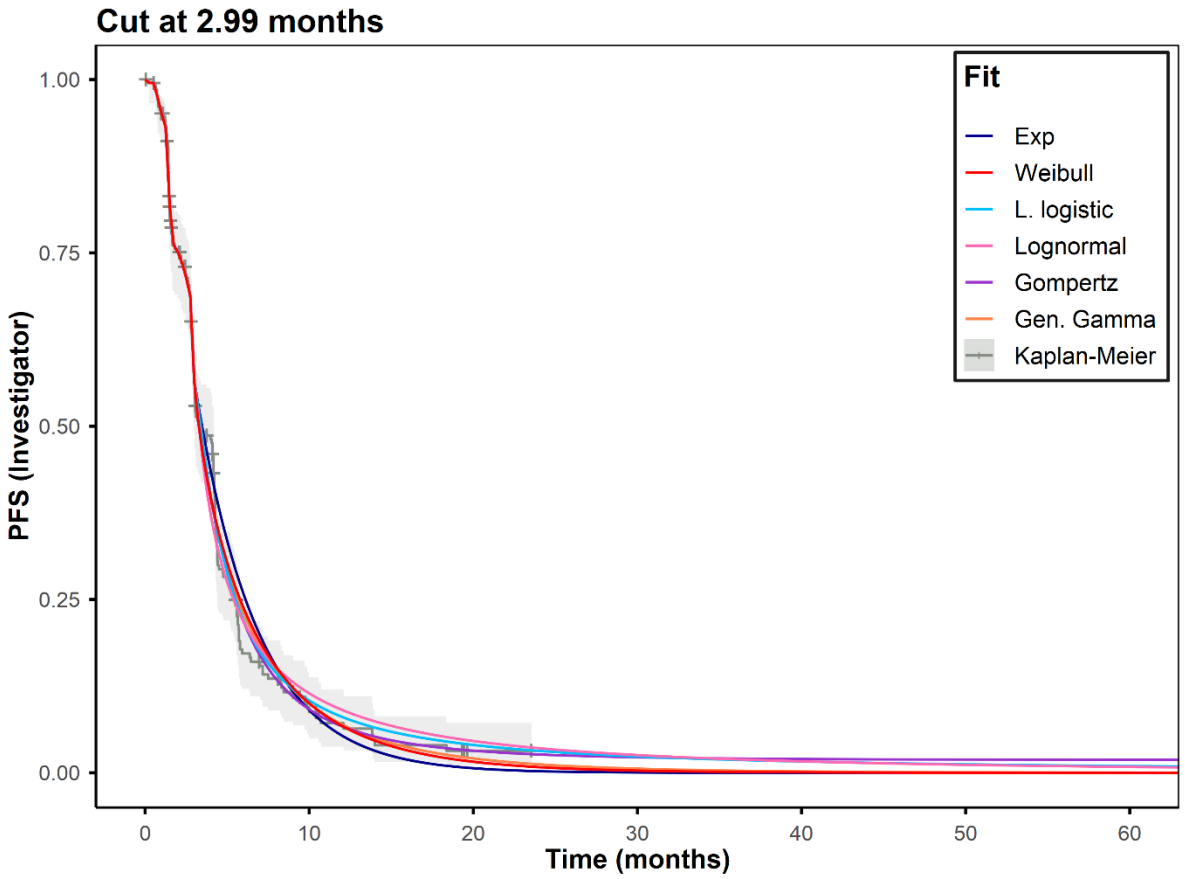
Table 40. Extrapolation of survival outcomes from ATTRACTION-3

	Progression-free survival	Overall survival
Nivolumab		
Median	1.68 months	10.91 months
Extrapolation method	Semi-parametric Kaplan-Meier to 2.99 months with parametric extrapolation using Weibull distribution	Semi-parametric Kaplan-Meier to 2.99 months with parametric extrapolation using log-logistic distribution
Median (from extrapolation)	1.68 months	10.87 months
Mean (from extrapolation)	5.78 months	24.33 months
Taxane		
Median	3.35 months	8.37 months
Extrapolation method	Semi-parametric Kaplan-Meier to 2.99 months with parametric extrapolation using Weibull distribution	Semi-parametric Kaplan-Meier to 2.99 months with parametric extrapolation using exponential distribution
Median (from extrapolation)	3.27 months	8.90 months
Mean (from extrapolation)	4.79 months	11.96 months



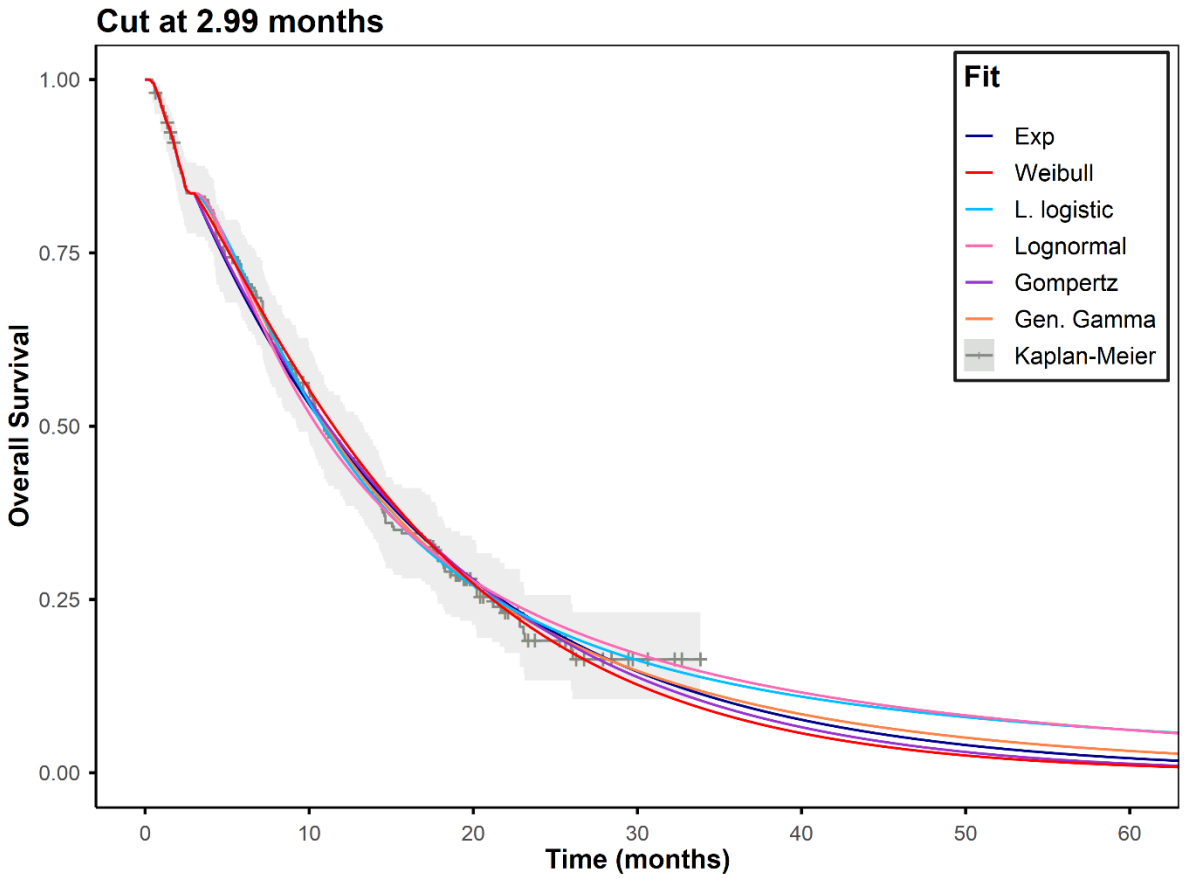
Fit Name	AIC	BIC	Parameters	Median (Months)	Mean (Months)
Exp.	386.75	389.09	lambda: 0.1101	1.68 (1.54, 2.79)	5.55 (4.45, 6.92)
Weibull	386.72	391.41	shape: 0.8621; scale: 8.9647	1.68 (1.54, 2.79)	5.78 (4.55, 7.54)
L. logistic	383.62	388.30	shape: 1.2032; scale: 5.2410	1.68 (1.54, 2.79)	12.41 (6.09, 67.74)
Lognormal	387.60	392.28	mu: 1.6315; sigma: 1.5250	1.68 (1.54, 2.79)	8.33 (5.53, 15.33)
Gompertz	383.61	388.30	shape: -0.0625; rate: 0.1552	1.68 (1.54, 2.79)	N/A (4.31, N/A)
Gen. Gamma	387.13	394.16	mu: 1.9519; sigma: 1.3202; Q: 0.5449	1.68 (1.54, 2.79)	6.24 (4.76, 11.37)

Figure 23. Progression-free survival: ATTRACTION-3 nivolumab arm – Kaplan-Meier data to 2.99 months followed by parametric extrapolation



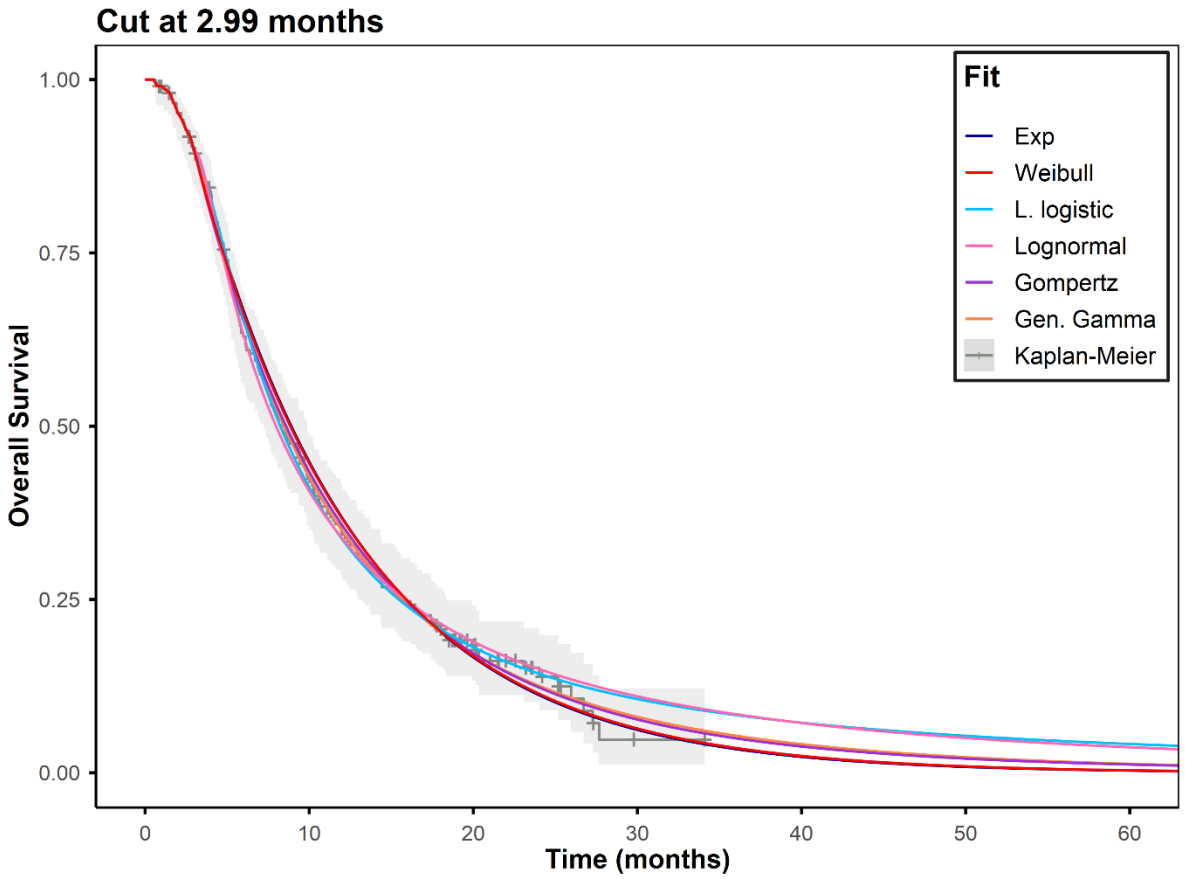
Fit Name	AIC	BIC	Parameters	Median (Months)	Mean (Months)
Exp.	427.95	430.62	lambda: 0.2617	3.47 (2.99, 3.94)	4.69 (4.10, 5.42)
Weibull	423.12	428.46	shape: 0.8166; scale: 3.5782	3.27 (2.99, 3.65)	4.79 (4.14, 5.72)
L. logistic	421.34	426.69	shape: 1.2208; scale: 2.0693	3.39 (2.99, 3.78)	8.13 (5.03, 48.69)
Lognormal	430.28	435.62	mu: 0.6395; sigma: 1.5689	3.28 (2.99, 3.63)	6.20 (4.81, 8.73)
Gompertz	417.92	423.26	shape: -0.1091; rate: 0.3720	3.33 (2.99, 3.67)	N/A (3.86, N/A)
Gen. Gamma	423.71	431.73	mu: 1.1037; sigma: 1.3038; Q: 0.7127	3.30 (2.99, 3.69)	4.87 (4.30, 6.08)

Figure 24. Progression-free survival: ATTRACTION-3 taxane arm – Kaplan-Meier data to 2.99 months followed by parametric extrapolation



Fit Name	AIC	BIC	Parameters	Median (Months)	Mean (Months)
Exp.	944.27	947.42	lambda: 0.0646	10.94 (9.49, 12.78)	15.70 (13.55, 18.48)
Weibull	944.04	950.33	shape: 1.1237; scale: 15.3711	11.49 (9.96, 13.33)	15.07 (13.14, 17.82)
L. logistic	941.44	947.73	shape: 1.4773; scale: 10.3094	10.87 (9.38, 12.66)	24.33 (18.18, 38.88)
Lognormal	944.13	950.42	mu: 2.3124; sigma: 1.1952	10.49 (9.05, 12.31)	20.00 (15.95, 26.97)
Gompertz	946.11	952.41	shape: 0.0058; rate: 0.0616	11.14 (9.37, 13.09)	15.24 (12.99, 17.93)
Gen. Gamma	943.17	952.61	mu: 2.5291; sigma: 1.0531; Q: 0.4844	10.99 (9.32, 12.71)	16.44 (14.15, 25.94)

Figure 25. Overall survival: ATTRACTION-3 nivolumab arm – Kaplan-Meier data to 2.99 months followed by parametric extrapolation



Fit Name	AIC	BIC	Parameters	Median (Months)	Mean (Months)
Exp.	1009.02	1012.23	lambda: 0.0990	8.90 (7.87, 9.98)	11.96 (10.52, 13.52)
Weibull	1010.98	1017.39	shape: 0.9866; scale: 10.0788	8.85 (7.73, 10.02)	12.00 (10.52, 13.72)
L. logistic	1009.69	1016.11	shape: 1.3641; scale: 6.2024	8.24 (7.18, 9.35)	20.14 (14.81, 32.41)
Lognormal	1013.76	1020.18	mu: 1.7919; sigma: 1.2954	7.98 (6.97, 9.06)	15.36 (12.24, 19.87)
Gompertz	1010.14	1016.56	shape: -0.0135; rate: 0.1086	8.59 (7.41, 9.92)	N/A (10.60, N/A)
Gen. Gamma	1009.75	1019.38	mu: 2.0868; sigma: 1.1447; Q: 0.5460	8.46 (7.35, 9.59)	12.64 (10.93, 16.03)

Figure 26. Overall survival: ATTRACTION-3 taxane arm – Kaplan-Meier data to 2.99 months followed by parametric extrapolation

B.3.3.2.1.2. Clinical rationale for survival curves applied in the economic evaluation

The PFS hazard profile for nivolumab shows two distinct portions and the optimal cut point for a semi-parametric model was calculated to be [redacted] months, which is the timepoint that shows a maximum rate of change in the hazard. Visual examination of the hazard plot for this outcome shows a noticeable change in the slope of hazard after one month and thus this seems reasonable. The parametric forms fit from after the first month show little variability and all improve on the parametric forms fit from time zero.

The hazard profile for the taxane arm shows a similar profile to the nivolumab arm although the distinct parts of the hazard profile are less exaggerated. As numerous events occur immediately after initiation in all arms, it is more appropriate to consider semi-parametric forms

for PFS where the extrapolated period is not informed by those who contribute to the initially high hazard. The extrapolated period should be informed by those who are still pre-progressed where observed data ends ideally.

Parametric forms fit to the OS outcome follow the same assumptions as when fit to PFS. As can be seen in (Appendix D.1.3.4) a number of these assumptions do not hold. Examination of the hazard plots for nivolumab showed distinct portions of the hazard profile. This is similar to those seen in the pre-progression outcome but are slightly later in time. It is possible that the similarity in the survival curves and hazard profiles in the OS outcome and the PFS outcome are driven by patients who are moving from PFS quickly.

The parametric forms fit from time zero to the OS outcome for the nivolumab arm are reasonable but could certainly be improved upon for an economic model. The optimal cut time is estimated to be [redacted] months and most semi-parametric fits show an improvement on the fully parametric estimates. The semi-parametric curve using the optimal cut point also shows some improvement on the representation of a short period of stabilisation at the end of the observed period.

The optimal cut point for OS in nivolumab is later in time than the optimal cut point for PFS, which is expected. This makes clinical sense, as with nivolumab there is a risk of false progression being recorded in trial due to a delayed onset of action; a documented issue with immunosuppressants.[70] Therefore, it would be expected that the point of maximum rate of change of the hazard for PFS in this population would be soon after initiation of treatment. It is more appropriate to consider semi-parametric models for both OS and for PFS to ensure that those who were marked as progressed but may be responding long term to treatment represent the extrapolated portion of the curves.

Clinically, it is reasonable to assume that PFS for both arms is more likely than other outcomes to be best represented by a semi-parametric form; for nivolumab there is the possible existence of a subpopulation who experience false progression and for the taxane arm, normal variance is seen in the population where some patients will progress quickly or treatment is not well tolerated or effective. It is important to clarify that the speculation of the existence of subpopulations is driven by the hazard profiles and not by any description of the trial group or design. Additionally, it is possible that these profiles have appeared due to the mechanism of action of the treatments rather than being present at initiation. Distinct groups may exist in the cohort treated with nivolumab as well as the control although these may take very different profiles as these two technologies have very distinct pharmacological mechanisms of action. This is evident particularly for the hazard profiles in the control arms for the OS outcome where any change in hazard is quite slight but still present. It is also important to clarify that where the maximum rate of hazard of change is found is not necessarily particularly large when compared to other points; it is just the largest. This is less likely to be the case for profiles with obvious parts (such as the nivolumab arm) but more so for the taxane arm.

The hazard profile for the taxane arm for the OS outcome shows a smoother profile than for nivolumab. However, the fully parametric models provide poor fits to the observed data and overestimate outcomes.

The optimal cut for semi-parametric models in the taxane arm is [redacted] month, which is longer after the optimal cut for the PFS outcome than is seen in the nivolumab arm. This may be expected if the reason for the change of hazard rates is due to the differences in clinical progression of the disease when treated with different technologies. False progression would not be expected to be an issue for either docetaxel or paclitaxel as neither are immunotherapies. Therefore, when a patient's disease progresses and is not responsive to therapy, it is reasonable to assume these patients continue to progress at a similar rate and certainly sooner than in the immunotherapy arm. Therefore, the maximum change of the rate of the hazard would be expected to be later after that of PFS in the control than immunotherapy arm but not as pronounced which is what is seen. Only the underlying patient group who are responding or have a slower disease progression remain to inform the curves and extrapolated periods.

It is important to consider that, where cut points are positioned later in time, less patients inform the parametric portion of the curve and thus the extrapolated portion. As this patient number decreases, there is potential for uncertainty to increase. Consistently positioning the cut point at 2.99 months ensures that any patients who may quickly progress and die are appropriately represented by the non-parametric part. This method also ensures that as many patients as possible, who most accurately represent the long-term outcomes, are retained to inform the parametric and extrapolated components for a cost-effectiveness model.

B.3.3.2.1.3. Validation of survival curves applied in the economic evaluation

There are no other studies with which to validate the results for extrapolation of the nivolumab arm other than the informing trial, ATTRACTION-3. The extrapolated curves and approaches were compared to the observed values as much as possible. This method informed selection of the most appropriate modelling approach and fit as a form of validation. The results for nivolumab can be seen in Table 41 and Table 42 for OS and Table 43 and Table 44 for the taxane arm. Table 45 to Table 48 report for each arm respectively for PFS estimates. Restricted means reported used the ATTRACTION-3 data set minus the last 10 events in each arm; the time point for this event was used for the semi-parametric estimates.

Overall, the semi-parametric models show less overall variation in the estimates and are closer to the observed values than the parametric models. This is particularly important with reference to the median values as there are more events initially and these incur cost which need to be well represented in cost-effectiveness analysis.

The only other source available to validate the estimates for the taxane arm are from Auzolle et al[46]. This study reported an OS median of 7.5 months and 3.9 for PFS although ITT numbers were low (n=29 in the taxane arm). These estimates are lower for OS than the observed and predicted values for the taxane arm but are broadly in line with the estimates for the taxane arm that were modelled.

Table 41: Observed and predicted estimates of overall survival for nivolumab (mean and median values)

	Observed	Parametric	Semi-parametric	Observed	Parametric	Semi-parametric
Distribution	**Median**			**Restricted mean**		
Exponential	10.91	10.93	10.94	[redacted]	12.88	12.94
Generalised Gamma	10.91	10.73	10.99	[redacted]	12.91	12.98
Gompertz	10.91	11.21	11.14	[redacted]	12.96	12.99
Log-Logistic	10.91	10.52	10.87	[redacted]	12.99	13.00
Log-Normal	10.91	10.14	10.49	[redacted]	12.83	12.95
Weibull	10.91	11.35	11.49	[redacted]	13.04	13.09

Table 42: Observed and predicted estimates of overall survival for nivolumab (proportion surviving at specific time points)

	Observed	Parametric	Semi-parametric	Observed	Parametric	Semi-parametric	Observed	Parametric	Semi-parametric
Distribution	**Survival at 6-Months**			**Survival at 1-Year**			**Survival at 2-Years**		
Exponential	[redacted]	68.3%	68.8%	46.9%	46.7%	46.7%	[redacted]	21.8%	21.5%
Generalised Gamma	[redacted]	69.5%	71.4%	46.9%	45.9%	46.5%	[redacted]	21.5%	21.0%
Gompertz	[redacted]	69.5%	69.3%	46.9%	47.5%	47.3%	[redacted]	21.1%	21.1%
Log-Logistic	[redacted]	69.5%	71.9%	46.9%	45.2%	45.9%	[redacted]	23.0%	21.6%
Log-Normal	[redacted]	67.0%	70.6%	46.9%	44.4%	45.0%	[redacted]	23.4%	22.6%
Weibull	[redacted]	71.1%	71.2%	46.9%	47.8%	48.3%	[redacted]	20.3%	20.2%

Table 43: Observed and predicted estimates of overall survival for taxane arm (mean and median values)

	Observed	Parametric	Semi-parametric	Observed	Parametric	Semi-parametric
Distribution	**Median**			**Restricted mean**		
Exponential	8.38	8.55	8.90	[redacted]	10.60	10.82
Generalised Gamma	8.38	8.73	8.46	[redacted]	11.11	10.69
Gompertz	8.38	8.32	8.59	[redacted]	10.83	10.73
Log-Logistic	8.38	8.61	8.24	[redacted]	10.61	10.69
Log-Normal	8.38	8.65	7.98	[redacted]	10.74	10.67
Weibull	8.38	9.63	8.85	[redacted]	10.99	10.81

Table 44: Observed and predicted estimates of overall survival for taxane arm (proportion surviving at specific time points)

	Observed	Parametric	Semi-parametric	Observed	Parametric	Semi-parametric	Observed	Parametric	Semi-parametric
Distribution	**Survival at 6-Months**			**Survival at 1-Year**			**Survival at 2-Years**		
Exponential	[redacted]	61.5%	66.7%	34.4%	37.8%	36.8%	[redacted]	14.3%	11.2%
Generalised Gamma	[redacted]	64.7%	65.4%	34.4%	37.7%	35.0%	[redacted]	16.2%	12.5%
Gompertz	[redacted]	65.2%	65.2%	34.4%	39.7%	65.2%	[redacted]	11.4%	12.3%
Log-Logistic	[redacted]	67.0%	65.4%	34.4%	34.3%	33.7%	[redacted]	11.8%	14.3%
Log-Normal	[redacted]	66.2%	63.1%	34.4%	35.5%	33.8%	[redacted]	12.2%	15.0%
Weibull	[redacted]	69.1%	66.3%	34.4%	39.5%	36.7%	[redacted]	9.7%	11.4%

Table 45: Observed and predicted estimates of progression free survival for nivolumab (mean and median values)

	Observed	Parametric	Semi-parametric	Observed	Parametric	Semi-parametric
Distribution	**Median**			**Restricted Mean**		
Exponential	1.68	3.61	1.68	[redacted]	5.08	4.99
Generalised Gamma	1.68	2.62	1.68	[redacted]	5.01	4.93
Gompertz	1.68	2.66	1.68	[redacted]	4.83	4.85
Log-Logistic	1.68	2.60	1.68	[redacted]	4.24	4.86
Log-Normal	1.68	2.89	1.68	[redacted]	4.65	4.88
Weibull	1.68	3.30	1.68	[redacted]	5.03	4.93

Table 46: Observed and predicted estimates of progression free survival for nivolumab (proportion surviving at specific time points)

	Observed	Parametric	Semi-parametric	Observed	Parametric	Semi-parametric	Observed	Parametric	Semi-parametric
Distribution	**Survival at 6-Months**			**Survival at 1-Year**			**Survival at 2-Years**		
Exponential	24.2%	31.6%	27.4%	11.9%	10.0%	14.2%	[redacted]	1.0%	3.8%
Generalised Gamma	24.2%	25.4%	25.3%	11.9%	13.0%	13.6%	[redacted]	6.4%	5.4%
Gompertz	24.2%	26.0%	24.9%	11.9%	12.2%	13.1%	[redacted]	6.3%	6.2%
Log-Logistic	24.2%	20.8%	25.2%	11.9%	7.9%	13.1%	[redacted]	2.8%	6.0%
Log-Normal	24.2%	25.0%	24.3%	11.9%	9.4%	13.6%	[redacted]	2.5%	6.8%
Weibull	24.2%	30.7%	25.8%	11.9%	11.2%	14.0%	[redacted]	1.7%	4.8%

Table 47: Observed and predicted estimates of progression free survival for taxanes (mean and median values)

	Observed	Parametric	Semi-parametric	Observed	Parametric	Semi-parametric
Distribution	**Median**			**Restricted Mean**		
Exponential	3.35	3.35	3.47	[redacted]	4.29	4.38
Generalised Gamma	3.35	3.41	3.30	[redacted]	4.29	4.29
Gompertz	3.35	3.42	3.33	[redacted]	4.32	4.21
Log-Logistic	3.35	3.43	3.39	[redacted]	4.21	4.25
Log-Normal	3.35	3.46	3.28	[redacted]	4.30	4.23
Weibull	3.35	3.83	3.27	[redacted]	4.47	4.29

Table 48: Observed and predicted estimates of progression free survival for taxanes (proportion surviving at specific time points)

	Observed	Parametric	Semi-parametric	Observed	Parametric	Semi-parametric	Observed	Parametric	Semi-parametric
Distribution	**Survival at 6-Months**			**Survival at 1-Year**			**Survival at 2-Years**		
Exponential	17.2%	28.9%	25.8%	7.2%	8.4%	5.4%	-	0.7%	0.2%
Generalised Gamma	17.2%	24.1%	23.0%	7.2%	6.2%	7.0%	-	1.0%	1.2%
Gompertz	17.2%	29.2%	21.8%	7.2%	7.8%	6.7%	-	0.4%	2.6%
Log-Logistic	17.2%	22.1%	22.0%	7.2%	5.6%	8.1%	-	1.2%	3.2%
Log-Normal	17.2%	24.2%	21.8%	7.2%	5.7%	9.1%	-	0.7%	3.5%
Weibull	17.2%	28.9%	23.8%	7.2%	4.7%	6.8%	-	0.1%	0.8%

B.3.3.2.2. *All-cause mortality*

Individuals randomised into clinical trials are likely to be slightly younger and healthier than the overall oesophageal cancer patient population in the UK. The mean age of patients in ATTRACTION-3 is 63.8 years, increasing the likelihood that most deaths observed over the trial period were cancer-related.

Therefore, the model includes age and gender-adjusted mortality based on information from UK life tables, described in Table 49.[71] These values are included in every cycle in addition to the disease-related mortality values and are applied multiplicatively. While some form of double counting occurs, this effect applies equally to all comparators and is likely to have a minimal impact on predicted survival (and hence cost-effectiveness).

Table 49. Excerpt from England and Wales life tables[72]

Age	Annual probability of mortality	
	Males	Females
50	0.0034	0.0021
51	0.0035	0.0024
52	0.0039	0.0026
53	0.0041	0.0028
54	0.0044	0.0030
55	0.0049	0.0033
-	-	-
95	0.2627	0.2304
96	0.2851	0.2491
97	0.3067	0.2708
98	0.3220	0.2903
99	0.3650	0.3164
100	0.3882	0.3397

B.3.3.2.3. *Treatment discontinuation*

B.3.3.2.3.1. Treatment switching and subsequent therapies

The model incorporates treatment switching due to progression and AEs via a time on treatment curve. This allows the application of treatment beyond progression for patients who resume receiving second line therapy post-progression. Patients on nivolumab or taxane are then switched to receive BSC following cessation of treatment.

Table 50. Subsequent therapy applied in model

Treatment arm	Subsequent therapy (pre-progression and post-progression)
Nivolumab	BSC
Taxane	BSC
Scenario comparators	BSC

B.3.3.2.3.2. Time on treatment

A full description of extrapolation of discontinuation events is provided in Appendix D.1.3.4. In brief, patient-level data were obtained describing discontinuation due to progression, study drug toxicity, AEs unrelated to study therapy and withdrawal of patient consent. Data informing this extrapolation was derived from ATTRACTION-3 (base case analysis). In line with the survival analysis outlined in Section B.3.3.2.1, appropriateness of the extrapolation was evaluated by visual inspection of the fit, consideration of the log-cumulative hazard profile and minimisation of goodness-of-fit statistics (AIC and BIC). Based on this approach, the most appropriate extrapolation was considered to be generalised-gamma for nivolumab and exponential for taxane. Inputs are summarised in Table 51 and presented in Figure 27 and Figure 28.

Table 51. ATTRACTION-3: Time on treatment (applied to nivolumab and taxane)

	Nivolumab	Taxane
Distribution	Generalised-Gamma	Exponential
Parameters	Mu: [redacted] Sigma: [redacted] Q: [redacted]	Lambda: [redacted]

[redacted]

B.3.3.2.4. *Adverse events*

Treatment-related AEs are an inevitable consequence of any intervention, and these events are applied in the model, affecting the costs accrued by patients on each intervention. In order to reflect the adverse events that occurred in ATTRACTION-3, grade 3-4 adverse events in any arm (regardless of causality) are modelled, as well as 'select' adverse events deemed to be appropriate within clinical practice. Thus, the model includes anaemia, febrile neutropenia, leukopenia, lymphocyte count decreased, neutropenia, alanine aminotransferase increased, aspartate aminotransferase increased, diarrhoea and rash.

Data from ATTRACTION-3 (base case analysis) assumed to comprise all available evidence describing the safety profile of nivolumab and taxane for the treatment of unresectable, advanced oesophageal cancer when standard chemotherapy has failed. Grade 3-4 treatment-related adverse event rates were sourced from the database lock on November 2018.

Incidence probabilities were converted into monthly equivalents based on number of patients experiencing an event and follow-up time using standard formulae; inputs are summarised in Table 52. For entry into the model, these were converted to weekly probabilities and applied to all patients in the model in all cycles while receiving nivolumab. Thus, the model assumes that there is a constant rate of adverse events during treatment. As the majority of events are likely to occur in the initial trial period, this may overestimate the rate of adverse events over long-term treatment and impacts the nivolumab arm disproportionately.

B.3.3.2.4.1. Derivation of adverse event model inputs

Data describing number of adverse events, number of patients exposed and exposure time in months were obtained for nivolumab and its comparators. These were used to calculate an initial event rate, as outlined in the formula below.

$$Initial\ event\ rate = \frac{-\ln\left(1 - \left(\frac{number\ of\ events}{number\ of\ patients\ exposed}\right)\right)}{\left(\frac{exposure\ time\ in\ months}{number\ of\ patients\ exposed}\right)}$$

The initial event rate was then used to calculate a one-month probability of each event, applying the formula below.

$$One\ month\ probability\ of\ event = 1 - \ exp^{(-initial\ event\ rate)}$$

The one-month probability of each event is used as the model input. The model then converts the one-month probability of each adverse events to a weekly cycle length by converting the probability to a rate, which can be converted to the correct time frame, then transformed to probabilities using the formula outlined below.

$$rate = \frac{-\ln(1 - probability)}{t_1}$$

$$probability = 1 - \ exp^{-rt_2}$$

Where t_1 is the transformation time (i.e. when converting, month to weekly cycle t_1 is 4.348) and t_2 is weekly cycle (i.e. 1 in this example).

Table 52. Base case analysis: weekly adverse event probabilities for nivolumab and taxane (ATTRACTION-3)

Adverse event	Nivolumab		Taxane	
	Mean	SE	Mean	SE
Anaemia	0.00036	0.00019	0.00226	0.00050
Febrile neutropenia	0.00000	0.00000	0.00323	0.00059
Leukopenia	0.00009	0.00010	0.01488	0.00127
Lymphocyte count decreased	0.00018	0.00014	0.00377	0.00064
Neutropenia	0.00009	0.00010	0.01901	0.00144
Alanine aminotransferase increased	0.00000	0.00000	0.00011	0.00012
Aspartate aminotransferase increased	0.00009	0.00010	0.00000	0.00000
Diarrhoea	0.00018	0.00014	0.00022	0.00016
Rash	0.00009	0.00010	0.00022	0.00016

B.3.4. Measurement and valuation of health effects

B.3.4.1. Health-related quality-of-life studies

As described in Appendix G, an SLR was conducted to identify health-related quality-of-life studies.

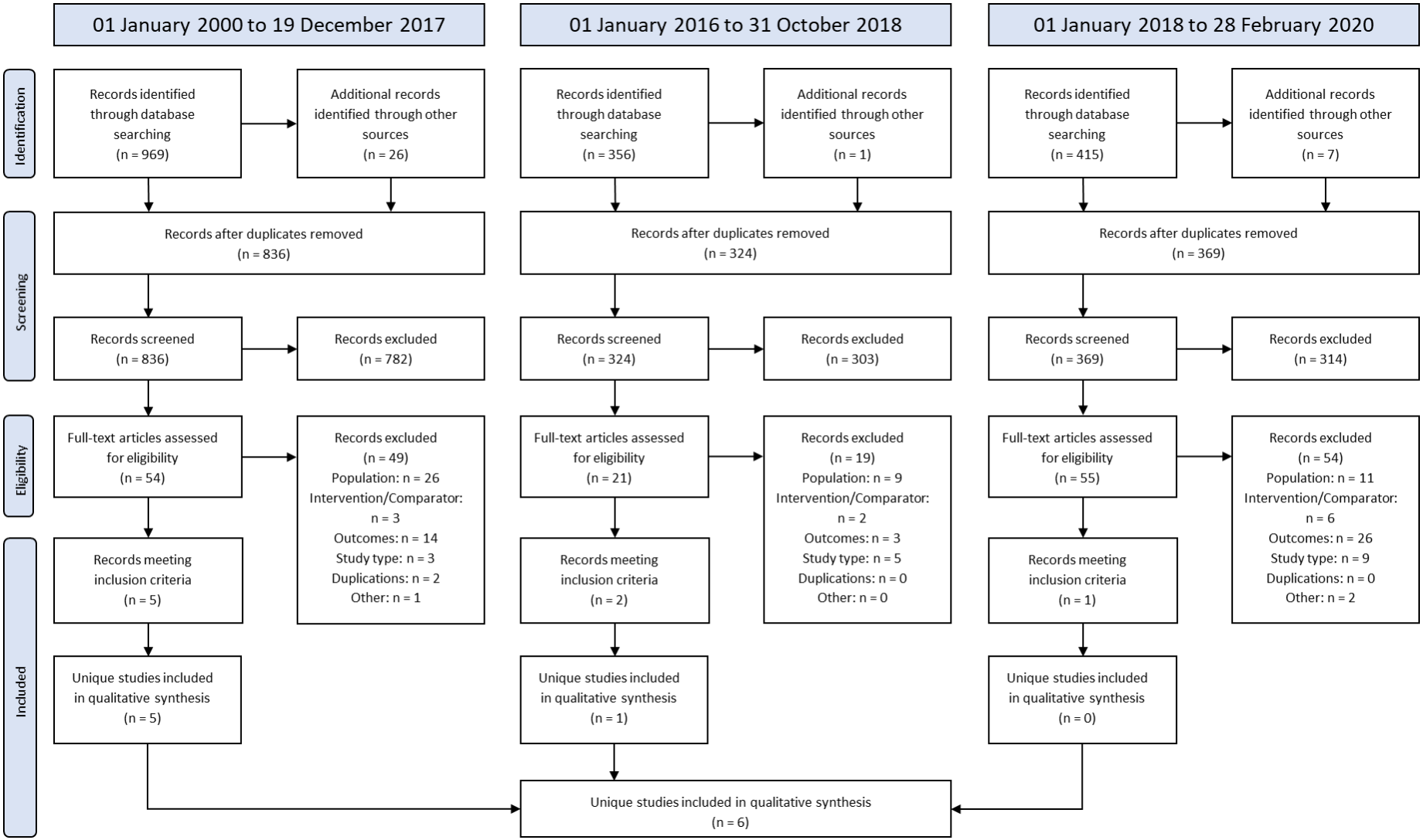


Figure 27. PRISMA diagram illustrating the study selection process for identifying cost and healthcare resource studies for the period from 01 January 2000 to 03 March 2020

B.3.4.2. Health-related quality-of-life data from clinical trials

Patient-reported outcomes were reported during ATTRACTION-3.,[38] specifically EQ-5D-3L measures from patients throughout the trial. Patient baseline demographic and clinical characteristics were collected, and patient time of clinical progression and death were recorded if these events occurred within the follow-up period. This was used to calculate the utility values most appropriate to each health state and arm.

B.3.4.2.1. *Analysis Methods*

B.3.4.2.1.1. Pre-processing of HRQoL data

Patient-assessed HRQoL data was collected with varying frequency through the trial, dependent upon treatment status, which was closely associated with progression status. To allow fitting of a model assuming an AR(1) autocorrelation structure between observations upon a single patient, the period between observations needed to be regularised. HRQoL observations were regularised to a 12-week period, corresponding to the lowest frequency of collection on trial. Observations taken within 6 weeks post or prior to the target day were deemed eligible, with the nearest eligible observation used as the observation for that analytical timepoint. In the event of two or more equidistant observations, the earliest observation had priority

Observations (missing or complete) were recorded as being of patients who had clinically progressed only if the date of questionnaire completion was greater than or equal to the date of observed clinical progression of that patient. If the observation target date exceeded the date of censorship for observation of clinical progression for that patient and the patient was not known to have progressed, the progression status of that observation was marked for imputation. If a patient had any observations within the analysis window, they were assumed eligible. If no observations were available for a patient within a window but the patient was known to survive to a date greater than the target observation day, then the patient was considered eligible and the observation missing.

For patients with an unknown date of death (due to administrative censoring), if the date of last survival observation was less than or equal to the target observation day, the patient was considered ineligible for observation for the purpose of assessing within-trial missingness; within datasets of imputed utility, the eligibility of these patients was determined by their imputed time to death.

B.3.4.2.1.2. Covariate identification

Regression models were developed to characterise the utility data using a fixed set of covariates. Covariates deemed clinically plausible to influence utility in the context of patients with unresectable advanced, recurrent or metastatic oesophageal squamous cell carcinoma were identified by a pragmatic literature review and related to the available data collected from ATTRACTION-3. The following were considered for input:

- Age

- Sex
- Eastern Cooperative Oncology Group Performance Score (ECOG-PS)
- Progression status
- Number of metastatic sites
- Location of metastases (liver, lung, lymph nodes, intraperitoneal, other)

B.3.4.2.1.3. Description of Missingness

Missingness diagrams were produced to describe the patterns of missing data present in the study. By plotting each patient's history of observation and missingness in a block fashion, missingness was observed to be monotonic (missing constantly from one assessment until end of follow-up) or non-monotonic (sporadic). It was also visually assessed whether missingness in patients was temporally correlated with death.

The indexed utility data were described using complete-case analysis. In complete-case analysis, any records with missing observations (utility) or covariates (progression status) are removed to leave a dataset with no missing values. Complete-case analysis is valid under the assumption that data is missing completely at random (MCAR), as the remaining observations describe a reduced but unbiased sample of the overall distribution of observations.

The distribution of utilities observed at each observation time, conditional upon progression status, was described using box plots and simple estimation of mean and standard error of mean.

The progression-state specific mean utility (per treatment arm) was estimated using simple means, and the standard error of this mean was evaluated using the Prais-Winsten correction for autocorrelation within patient observations (assuming that non-monotonic missing data could be ignored).[73, 74]

B.3.4.2.1.4. Multiple imputation of missing progression and utility observations

The assumption that utility observations were MCAR was thought to be unlikely to be met, as the rate of missingness was higher after patients had progressed and was associated with short time to death. A plausible mechanism for missingness would be that worsening physical status associated with end of life would cause patients to be unable or unwilling to attend clinic to complete EQ-5D-3L questionnaires. On the assumption that the missing utility values could be predicted conditional upon the observed data (baseline covariates, progression status, previous utility, time to death), this data could be assumed to be missing at random (MAR) and could be imputed by the method of Multiple Imputation by Chained Equations (MICE).

If the missing data could not be predicted conditional upon the observed data (e.g. if the unobserved patients had systematically lower utility than observed patients with the same covariates) then the data would be missing not at random (MNAR), and MICE alone would be insufficient to impute the data; instead, a joint system of equations predicting both missingness

(a “selection” model) and the utility value (an “outcome” model) would be required, as per the methods of Heckman (1976).[75] An example of such approach is given in Galimard et al (2018)[76], but such models would require a continuous outcome to be conditionally normally distributed, which is not the case for utility data on its natural scale; also, due to the bounded nature of the measure and the high number of observations with no disutility causing inflation at one boundary, it is difficult to transform utility data onto a scale where it approximates a normal distribution. As such, the assumption of MNAR was left unexplored for this data, awaiting further development of numerical techniques in this field.

The assumption of MAR was expected to hold if patients with certain utility values were not selectively removed from the dataset (i.e. if missingness and outcome conditional upon the observed variables was random; this would be the case if missingness was dependent upon some instrumental variable such as distance from home to clinic). A number of predictive measures were expected to condition this model, but among the most important was the measure of time until death. Time to death has been shown to be highly predictive of patient-reported utility in advanced metastatic cancer[77], and observation of time to death in clinical trials among such patients is generally good in comparison with other prognostic variables, particularly after clinical progression when imaging and laboratory measurement schedules are frequently relaxed from the on-treatment period. Administrative censoring is present, preventing full observation of time to death, and whilst censoring time is assumed to be independent of the time of death in any patient, observation that a patient’s time of death has been censored in a clinical trial is not independent of time to death. The patient time of death is not MCAR, and a conditional imputation model is necessary to impute time of death under the assumption of MAR.

In order to use time to death as a predictor in the imputation model for utility, death times, where missing, were imputed. Harel et al (2007) describe a two-stage multiple imputation (MI) model where time to death is first multiply imputed conditional upon the observed variables to provide several datasets with complete observation of time to death.[78] These datasets are then used in turn to multiply impute the missing outcomes of interest, allowing computation of statistics on these complete datasets that are conditional upon survival.

A parametric model of time to death was created conditional upon baseline characteristics, including utility at baseline. A model using measurements at times other than baseline was not considered appropriate, as selection of any single time after baseline would selectively reduce the available data, selection of multiple observation times on single patients would cause under-estimation of the uncertainty in the parameter estimates, and creation of a fully time-varying model would necessitate simulation of the time variation of the independent covariates determining the hazard of death. Due to this limitation, progression status could not be used as a predictor of mortality.

This model was then used to impute the missing time of death after their censoring time for all patients with censored overall survival observations, conditional upon their baseline covariates and their most recent utility value. The structure of this model and method of imputation implies that all increase in hazard experienced by the patient is explained solely by the patient-assessed HRQoL; clinical measurements are assumed unchanged from baseline, and there

is no additional increase in hazard due to any time-dependent effect. Regularly collected measures used for demonstrating clinical effect, such as sum of target tumour diameters, were considered inappropriate to use as predictive of time to death as their influence on mortality hazard is conditional upon other factors, such as location and depth.[79]

The imputed datasets with complete time of death observations were then used to multiply impute missing observations of progression status and utility.

The two-stage imputation method resulted in nested datasets that are analysed using the method attributed to Shen (2000) in Harel et al (2007).[78, 80] Mean utility at analysis times and pooled mean utility conditional upon progression status were calculated as for complete case analysis, with standard error of the latter mean estimate corrected for autocorrelation. The mean and standard error estimates from these datasets were then pooled to form unbiased estimates of the parameters under the assumption that utility data was MAR.

The resulting values assuming data is MAR(MI) used in the cost-effectiveness model can be seen in Table 53. In summary, the values used in the model are a product of the two-stage multiple imputation process where the time to death is imputed conditional on observed variables to make a complete data set and then this in turn is used to impute the missing outcomes of interest.

Table 53: Summary of utility values for cost-effectiveness analysis

	Nivolumab	Control
State	**Utility value: mean (standard error)**	**Utility value: mean (standard error)**
Pre-progression	[redacted]	[redacted]
Post-Progression	[redacted]	[redacted]
Abbreviations: HS, health state; AR, adverse reaction		

B.3.4.3. Health-related quality-of-life data used in the cost-effectiveness analysis

In the base case analysis, no age-related utility decrement was applied. ATTRACTION-3 trial data has been used to inform utility inputs, wherein OS at 24 months is 19.1% for nivolumab and 15.1% for taxanes. Hence, quality of life for the majority of patients is captured by the available trial data and any impact of aging is implicitly captured in the available data.

B.3.4.3.1. *Rationale for application of treatment-specific ATTRACTION-3 values in economic evaluation*

Utilities from ATTRACTION-3 have been applied in the economic evaluation, with values derived from specific treatment arms. This approach should be considered consistent with the NICE reference case, as it reflects the trial evidence.

With ATTRACTION-3, the utility associated with the pre-progression state for the taxane arm was [redacted], while post-progression utility was [redacted]. Although there is limited evidence in the

oesophageal cancer setting, both these utility values can be considered comparable with the published literature for gastric cancer ([redacted] versus 0.737[81] in pre-progression and [redacted] versus 0.587[81] for post-progression). However, the utility value in the nivolumab arm is higher for both the pre-progression and post-progression state, which may be expected due to the novel mechanism of action that may account for this improvement. In contrast to common oncology therapies, nivolumab enables the patient's own immune system to directly destroy cancer cells (in the same way that it would any other "foreign" cell), resulting in destruction of the tumour through pre-existing, intrinsic processes.

The clinical benefits of nivolumab are described in Section B.2, but in brief nivolumab is associated with several benefits that impact directly on patient quality of life, including improved rate of response and a tolerable AE profile, as well as improved PFS and OS. Significantly, observed ATTRACTION-3 data demonstrates that there is a large post-progression survival benefit compared with taxanes, supporting the impact of nivolumab on quality of life. Further, this is supported by the safety profile of nivolumab compared with chemotherapy; 65.6% of patients in the nivolumab arm reported a drug-related AE (grade 3-5: 18.2%) versus 95.2% for patients receiving paclitaxel or docetaxel (grade 3-5: 64.0%). It should be noted that quality of life outcomes during ATTRACTION-3 remained relatively stable in the nivolumab arm, as determined by EQ-5D and EQ-VAS; however, patients receiving taxanes frequently reported worsened quality of life outcomes during the trial period. Additional analysis of ATTRACTION-3 data is provided in Section B.3.4.3.1.1, which supports the beneficial impact of nivolumab.

Further, the utility values observed during ATTRACTION-3 are broadly equivalent to utility values observed from other nivolumab indications,[82-87] indicating that this utility gain may be due to the novel mechanism of action for nivolumab. In addition, it is of note that pre- and post-progression utility estimates for comparator treatments were different from those estimated for nivolumab, consistent with the application of nivolumab-specific utilities in this submission. Thus, the quality of life data derived from patients during ATTRACTION-3 reflects the expected benefits of nivolumab over taxanes, including the potential for immune system stimulation following progression.

Table 54: Summary of utility values for cost-effectiveness analysis

Indication	Health state	Instrument	Utility estimate (mean)		Source study
			Nivolumab	Comparator^	
Renal cell carcinoma	Progression-free	EQ-5D	0.80	0.76	CheckMate 025[87]
	Progressed state		0.73	0.70	
SCCHN	Progression-free	EQ-5D-3L	0.74	0.69	CheckMate 141[88]
	Progressed state		0.66	0.56	
Melanoma	Pre-progression	EQ-5D-3L	0.80	0.89	CA209-066 (range based on response status at landmark event) [85]
	Post-progression		0.84	0.74	
	Pre-progression	EQ-5D-3L	0.66-0.74	0.66	CA209-037 (range based on response status at landmark event)[85]
	Post-progression		0.73-0.82	0.76-0.85	

^Comparator treatments: RCC: everolimus; SCCHN: investigator's choice; melanoma: DTIC in CA209-066 and investigator's choice in CA209-037
SCCHN: squamous cell carcinoma of the head and neck

B.3.4.3.1.1. Additional analysis of ATTRACTION-3 data

The decision to use arm specific utilities was made based on examination of the domain scores while patients were on treatment. The data suggests that there are distinct differences in the scores for nivolumab and the control arm of ATTRACTION-3.

A higher proportion of patients remaining on treatment in the control arm experienced problematic symptoms (domain score of 2 or 3) compared with patients receiving nivolumab (Table 55). This was apparent in the early weeks and became more pronounced as time continued. In the final weeks on treatment, most or all patients in the control arm scored 2 or 3 in each domain. This is particularly pronounced in the mobility and self-care domains (Figure 30. Treatment arm specific utilities by domain). In contrast, very few patients in the nivolumab arm reported a score of 2 or 3 in later weeks indicating that they experienced no difference to their utility as a result of being on treatment with nivolumab.

This pattern appears to be primarily driven by patients in the control arm scoring 2 rather than 3 (the worst outcome). While the proportion of patients on treatment reporting a domain score of 3 is for the most part higher in the control arm than in the nivolumab arm, this seems to be confined to the first 36 weeks. After 36 weeks, aside from one patient who reported a domain score of 3 for mobility and usual activities in the control arm, no patients in either arm scored 3 after 36 weeks.

However, a large proportion of patients in the nivolumab arm do not report any problems in any domain (score of 1) while on treatment. Given this difference in arms, it is considered appropriate to reflect the differences by using arm specific utilities in the cost-effectiveness model. This is important as it is imperative that the cost-effectiveness model characterise and capture all benefits of each treatment.

Table 55. Treatment arm specific utilities by domain

		Week															
Domain	**Arm**	0	6	12	18	24	30	36	42	48	54	60	66	72	78	84	90
Mobility (%)	Nivolumab	[redacted]															
	Control	[redacted]															
	Difference	[redacted]															
Self-Care (%)	Nivolumab	[redacted]															
	Control	[redacted]															
	Difference	[redacted]															

Usual Activities (%)	Nivo	[redacted]															
	Control	[redacted]															
	Difference	[redacted]															
Pain Discomfort (%)	Nivolumab	[redacted]															
	Control	[redacted]															
	Difference	[redacted]															
Anxiety Depression (%)	Nivo	[redacted]															
	Control	[redacted]															
	Difference	[redacted]															

[redacted]

Patients who experienced the AEs of interest (Section B.2.10) were removed from the dataset for exploratory analysis. This was done to examine the appropriateness of using arm specific utilities with respect to capturing treatment differences and any disutility related to treatment. Far more patients in the control arm experienced these AEs than in the nivolumab arm and interpretation should take this into consideration.

Where the patients experiencing AEs were removed, before the mean time on treatment for the control arm, the utility was slightly higher (as expected) than in the complete dataset. This was slight (approximately 0.02) but accounts for the treatment specific differences that may be expected during this time. As there were so few patients removed in the nivolumab arm, the differences are negligible. However, given the clear differences between the domain scores of the control and nivolumab arm and the limited impact of AES to the nivolumab arm, it is completely appropriate to use arm specific data to inform the cost-effectiveness model.

B.3.5. Cost and healthcare resource use identification, measurement and valuation

B.3.5.1. Intervention and comparators' costs and resource use

B.3.5.1.1. *Nivolumab costs*

The costs of nivolumab, including drug procurement and administration, are applied each cycle, based on acquisition and administration costs detailed in Table 56 and Table 57.

Table 56. Nivolumab dosing and acquisition

Dosing	3mg/kg by intravenous infusion over 60 mins every 2 weeks
Dose per cycle	240mg
Cost (excluding PAS)	10 mg/mL concentration for solution for infusion in vial, 4mL = £439.00; 10mL - £1,097.00[2]
Cost per cycle (excluding PAS)	£2,633.00
Administration costs	£241.06 (derived from costs detailed in Table 57)
Total (excluding PAS)	£2,874.06

Table 57. Administration costs for nivolumab

Component	National cost collection for the NHS 2018/19[89]	Cost
Deliver Simple Parenteral Chemotherapy at First Attendance	Weighted average of SB12Z codes (DCRDN: Daycase and Regular Day/Night; OP: Outpatient; Oth: Other)	£241.06

B.3.5.1.1.1. Proportion of patients receiving doses

The model utilises the application of a treatment cost adjustment based on the proportion of patients receiving a dose during ATTRACTION-3. The proportion is determined by a ratio of the actual doses received by the expected doses received, as presented in Table 58.

Table 58. Proportion of patients receiving doses in patients receiving nivolumab

Treatment	Proportion of patients receiving doses
Nivolumab	0.952

B.3.5.1.2. *Patient Access Scheme*

A Patient Access Scheme (PAS) has been applied, comprising a discount of [redacted] from the nivolumab list price. In order to best replicate the true economic impact of a positive recommendation for nivolumab, the economic evaluation presented in this submission applies the PAS in the base case analysis (Table 59).

Table 59. Acquisition cost of nivolumab following application of PAS

	4 ml vial	10 ml vial	Cost per two-week cycle
No PAS	£439.00	£1,097.00	£2,633.00
PAS	[redacted]	[redacted]	[redacted]
PAS: patient access scheme			

B.3.5.1.3. *Comparators*

The costs of docetaxel and paclitaxel, including drug procurement and administration, are applied each cycle, based on acquisition costs detailed in Table 60 and Table 61. It is assumed that the body surface area for an oesophageal cancer patient is 1.79m^2.[90] The lowest possible acquisition costs were applied.

In order to accurately report costing for the taxane arm, a simple average of the costs applied for docetaxel and paclitaxel is applied.

Table 60. Docetaxel dosing and acquisition

Dose per cycle	75mg/m^2 by intravenous infusion administered every 3 weeks
Cost	10 mg/mL concentration for solution for infusion in vial; 2 mL; £162.75, 8 mL: £534.75, 16 mL: £1,069.50 20 mg/mL concentration for solution for infusion in vial, 1 mL: £145.80, 4 mL: £479.06, 7 mL: £900.00, 8 mL: £958.11 20mg, For solution in infusion in vial; 20mg, 1 mL: £153.47, 80mg, 4 mL:£504.27, 140mg, 7 mL: £720.10 ,160mg, 8 mL: £1,008.54[2]
Assumed dose	134.25 mg (based on average body surface are of 1.79m^2)[90]
Cost per cycle	£720.10
Administration costs	£241.06 (derived from costs detailed in Table 57)
Total	£961.16

Table 61. Paclitaxel dosing and acquisition

Dose per cycle	100mg/m^2 administered once weekly for 6 consecutive weeks followed by a 2-week washout period
Cost	6 mg/mL concentration for solution for infusion in vial; 5 mL: £66.85, 16.7 mL: £200.35, 25 mL: £300.52, 50 mL: £601.03[2]
Assumed dose	179 mg (based on average body surface are of 1.79m^2)[90]
Cost per cycle	£367.37
Administration costs	£241.06 (derived from costs detailed in Table 57)
Total	£608.43

B.3.5.1.3.1. Proportion of patients receiving doses

Similar to nivolumab, the proportion of patients receiving doses is applied in the model (Table 62). As treatment costs are applied as an average of docetaxel and paclitaxel, the proportion of patients receiving doses are applied separately in the model.

Table 62. Proportion of patients receiving doses in patients receiving docetaxel and paclitaxel

Treatment	**Proportion of patients receiving doses**
Docetaxel	0.960
Paclitaxel	0.938

B.3.5.1.4. *Best supportive/palliative care*

Within the final scope set out by NICE, BSC is specified as a comparator, with composition as including, but not limited to, anti-emetics, blood transfusion, oesophageal stents, palliative radiotherapy and palliative surgery.[91]

The composition of BSC is available from a previous NICE TA for a similar indication[64], and this comprises morphine, cognitive behavioural therapy, blood transfusions and radiotherapy. The composition of BSC from this TA (TA378[64]) was presented at a clinical advisory board meeting for the purposes of validating BSC management in the UK; however, clinical experts

noted that other forms of BSC that are commonly used for heavily pre-treated gastric/GOJ cancer patients were notably omitted from the list, particularly oesophageal stents and ascites drainage.

As clinicians agreed with the NICE scope, a clinician survey was initiated, where the survey was completed by practising oncologists and nurses in the UK based on their experience in treating UK-based gastric and GOJ cancer patients.[91] Hence, information obtained from the survey was used to inform the composition of BSC in the management of gastric and GOJ cancer in the UK, which has been used to inform the calculation of BSC costs.

Costs for the BSC components from the clinician survey are summarised in Table 64, with supporting information for the sources and calculations of individual BSC components shown in Table 63 to Table 66. .The resulting weekly costs for model inputs applied in base case analyses in the economic assessment are set out in Table 67.

Table 63. Costs for indicated components of BSC

Components	National cost collection for the NHS 2018/19[89]	Unit cost (£)
Radiotherapy	SC31Z, SC47Z [a]	£487.45
Blood transfusion	SA44A	£521.08
Procedures to control GI bleeds	FD03A, FD03B, FD03C, FD03D, FD03E [a]	£2,952.25
Cognitive behavioural therapy	A06A1	£83.17
Oesophageal stents	FE10A, FE10B, FE10C, FE10D	£3,058.96
[a] Weighted average based on activity per currency code		

Table 64. Costs comprising best supportive care (components from clinician survey)

BSC component	Cost per treatment (£) [a]	BSC + nivolumab/taxane					BSC (scenario analysis)				
		Proportion of patients [b]	Costs (£)				Proportion of patients [b]	Costs (£)			
			Costs/ week	Start week	Applied for no. of weeks	Weeks between cost		Costs/ week	Start week	Applied for no. of weeks	Weeks between cost
Ongoing costs											
Pain relief	£2.17	0.453	£0.98	1	NA	1	0.459	£1.00	1	NA	1
	£532.96	0.049	£26.29	1	NA	8	0.005	£26.62	1	NA	8
Blood transfusion	£521.08	0.202	£105.26	1	NA	4	0.213	£110.99	1	NA	4
Limited term costs											
Radiotherapy	£487.45	0.189	£184.25	1	7.5	1	0.213	£207.94	1	7.5	1
Procedures to control GI bleeds	£2,952.25	0.150	£441.36	1	1	NA	0.121	£357.22	1	1	NA
Drugs to control GI bleeds	£25.71	0.163	£4.18	1	4	4	0.157	£4.03	1	4	NA
Cognitive behavioural therapy	£83.17	0.149	£12.42	1	6	1	0.168	£14.00	1	4	NA
Oesophageal stents	£3,058.96	0.214	£653.09	1	1	NA	0.187	£571.41	1	1	NA
Ascites drainage	£3,404.20	0.111	£377.19	1	1	NA	0.137	£467.40	1	1	NA

[a] Costs per treatment as set out in Table 2
[b] From UK clinician survey for patients with gastric/GOJ cancer in third-line setting

Table 65. BSC components, frequency of administration and costs per treatment

BSC components		Schedule of administration	Cost per treatment (£)	Reference
Pain relief	Medication	Ongoing daily cost, based on SPCs and MIMS	£2.17	Table 66
	Nerve blocks	Once per 8 weeks, based on clinician opinion	£532.96	
Radiotherapy		Twice per week, for 7.5 weeks (total 15 visits)[92]	£487.45	Table 64
Blood transfusion		Once per month; assumption from TA378[64]	£521.08	Table 64
Procedures to control GI bleeds		One off cost; assumed that patients will receive only once, based on clinician opinion	£2,952.25	Table 64
Drugs to control GI bleeds [a]		Once per month for 2 months, based on clinician opinion	£25.71	-
Cognitive behavioural therapy		Once per week for 6 weeks; assumption from TA378[64]	£83.17	Table 64
Oesophageal stents		One off cost; assumed that patients will receive only once based on clinician opinion	£3,058.96	Table 64
Ascites drainage		One-off cost: assumed that patients will receive twice, based on clinician opinion	£3,404.20	White (2012)[93][b]

[a] Medications to control upper gastro-intestinal tract bleeds, 2012/2013 costs from Campbell et al 2017 (£23.76)[94],
[b] Reported cost for in-patient large volume paracentesis of £3,146, inflated to 2015/2016 costs[95]

Table 66. Pain relief - components

Pain relief components	Schedule of administration	Proportion of patients [a]	Source from MIMS (UK)[2] [c]	Cost per treatment	Weighted cost
Morphine	Daily	0.292	Zomorph	£0.42	£0.12
Morphine derivatives: fentanyl	Daily	0.192	PecFent	£0.68	£0.04
Morphine derivatives: tramadol	Daily	0.152	Tramadol	£0.11	£0.02
Morphine derivatives: codeine	Daily	0.164	Codeine Phosphate	£0.24	£0.04
Other morphine derivatives	Assumed zero costs due to low usage	0.041	-	£0.00	£0.00
Radiotherapy	Assumed zero costs [b]	0.203	-	£0.00	£0.00
NSAIDs	Daily	0.199	Ibuprofen	£0.10	£0.02
Paracetamol	Daily	0.364	Paracetamol	£0.19	£0.07
Other: Diet	Assumed zero costs	0.003	-	£0.00	£0.00
Other: Morphine derivatives: oxycodone	Daily	0.003	Abtard	£0.22	£0.00
Total costs for week 1					**£2.17**
Total ongoing daily cost					**£0.31**

[a] From UK clinician survey for patients with gastric/GOJ cancer in previously treated patients
[b] Excluded in this calculation as already included in BSC components in Table 1
[c] Drug costs obtained from MIMS (UK)[2], representing the lowest dose and recommended cost in each category

Table 67. Nivolumab and taxane costs from clinician market survey: model inputs

	Nivolumab* (£) with PAS	Taxane* (£)
Week 1	[redacted]	£2,589.82
Week 2	[redacted]	£501.87
Week 3	[redacted]	£501.87
Week 4	[redacted]	£982.45
Week 5	[redacted]	£611.31
Week 6	[redacted]	£501.87
Week 7	[redacted]	£665.81
Week 8	[redacted]	£93.11
Week 9	[redacted]	£436.75
Week 10	[redacted]	£785.78
Week 11	[redacted]	£305.20
Week 12	[redacted]	£305.20
Week 13	[redacted]	£891.04
Week 14	[redacted]	£305.20
Week 15	[redacted]	£0.98
Week 16	[redacted]	£481.56
Week 17	[redacted]	£436.75
Week 18	[redacted]	£305.20
Week 19	[redacted]	£785.78
Week 20	[redacted]	£305.20
Week 21	[redacted]	£410.46
Week 22	[redacted]	£785.78
Week 23	[redacted]	£0.98
Week 24	[redacted]	£0.98
Week 25 and beyond	The costs associated above for each respective treatment arm are repeated every 24 weeks from this point onward	

PAS: patient access scheme.
*Treatment continued until progression or discontinued, nivolumab costs applied with PAS but original costs shown for reference.
All therapies assume wastage.
Values rounded up to nearest £.

B.3.5.1.5. *Subsequent therapy*

The model incorporates treatment switching due to discontinuation. Patients on nivolumab and taxanes are switched to BSC upon discontinuation (the exception being in the scenario versus BSC, in which patients remain on BSC). Subsequent costs applied in post-progression are presented in Table 68.

Table 68. Costs comprising best supportive care as used in post-progression (components from clinician survey)

BSC component	Cost per treatment (£) [a]	BSC (post-progression)				
		Proportion of patients [b]	Costs (£)			
			Costs/ week	Start week	Applied for no. of weeks	Weeks between cost
Ongoing costs						
Pain relief	£2.17	0.7	£1.52	1	NA	1
Blood transfusion	£521.08	0.2	£104.22	1	NA	4
Limited term costs						
Radiotherapy	£487.45	0.2	£97.49	1	7.5	1
Procedures to control GI bleeds	£2,952.25	0.1	£295.23	1	1	NA
Drugs to control GI bleeds	£25.71	0.1	£2.57	1	4	4
Cognitive behavioural therapy	£83.17	0.1	£8.32	1	6	1
Oesophageal stents	£3,058.96	0.2	£611.79	1	1	NA

[a] Costs per treatment as set out in Table 2
[b] From UK clinician survey for patients with gastric/GOJ cancer in third-line setting

B.3.5.2. Health-state unit costs and resource use

During the clinician survey described in Section B.3.5.1.2, clinicians were asked to provide estimates of resource use associated with disease management. Within the base case analysis, it was assumed that this resource use would apply throughout the treatment period for both nivolumab and taxanes. The frequencies of resource use are described in Table 69 and the resource use estimates for pre- and post-progression state are described in Table 70.

Table 69. Disease management costs: frequency of resource use from clinician survey

		Consultations	Imaging scans	Blood tests	Liver function tests	Kidney function tests	Hospitalisations	Palliative care specialist nurse
Every 3 months	n	13	18	5	7	7	21	2
	%	33%	45%	13%	18%	18%	53%	5%
Monthly	n	17	8	16	20	20	9	10
	%	43%	20%	40%	50%	50%	23%	25%
Biweekly	n	8	4	4	3	3	3	14
	%	20%	10%	10%	8%	8%	8%	35%
Weekly	n	2	2	12	5	6	2	14
	%	0.050	5%	30%	13%	15%	5%	35%
Never	n	0	8	3	5	4	5	0
	%	0	20%	8%	13%	10%	13%	0
Mean frequency per week*		0.153	0.092	0.221	0.162	0.170	0.095	0.359

* The mean weekly frequency of each resource component was derived from the clinician survey and calculated in two steps:
1) Calculation of mean weekly frequency after removal of the 'Never' category^
2) Subsequent mean weekly frequency adjusted to account for the 'Never' component, where mean weekly frequency was multiplied by the total proportion of responses not in the 'Never' category
^ 'Never' category refers to the answer depicting that patients of the respective oncologist/nurse never used that particular resource for their patients on BSC

Table 70. Cyclic (weekly) health state resource use and costs

Resource	Unit cost (£)	Source	Weekly cost	
			Use	Cost (£)
Clinician consultation	£187.36	National cost collection for the NHS 2018/19: Medical Oncology (weighted average of consultant led and non-consultant led; WF01A, WF01B, WF01C, WF01D, WF02A, WF02B, WF02C, WF02D)[89]	0.153	£28.67
CT scan	£97.15	National cost collection for the NHS 2018/19: Computerised Tomography (weighted average of direct access, outpatient and other costs; RD20A, RD21A, RD22Z, RD23Z, RD24Z, RD25Z, RD26Z, RD27Z)[89]	0.092	£8.94
Full blood count	£2.79	National cost collection for the NHS 2019/19: Haematology; DAPS05[89]	0.221	£0.62
Renal function test	£1.10	National cost collection for the NHS 2019/19: Clinical Biochemistry; DAPS04[89]	0.162	£0.18
Hepatic function test	£1.10	National cost collection for the NHS 2019/19: Clinical Biochemistry; DAPS04[89]	0.170	£0.19
Hospitalisation	£534.07	National cost collection for the NHS 2018/19: Malignant Gastrointestinal Tract Disorders (weighted average of elective and non-elective long-stay FD11A, FD11B, FD11C, FD11D, FD11E, DF11F, F11G, FD11H, FD11J, FD11K[89]	0.095	£50.74
Palliative care specialist nurse	£76.74	National cost collection for the NHS 2018/19: Specialist Nursing, Palliative/Repsite Care, Adult (weighed average of N21AF, N21AN)[89]	0.359	£27.55
Sum			**£116.87**	
SE assumed to be 20% of the mean value				

End of life costs are detailed in Table 71, and were applied as a one-time cost in the cycle prior to death.

Table 71. End of life costs

	Costs	Inflated to
	Mean	**Mean (SE)**
End-of-life costs	£7,987.00[96]	£8,973.61 (£1,794.72)
SE: standard error SE assumed to be 20% of the mean value		

B.3.5.3. Adverse reaction unit costs and resource use

In order to provide an assessment of the costs associated with AEs, costs were sourced from recent literature and inflated to 2019 costs. These costs are summarised in Table 72.

Table 72. Adverse event costs

Adverse event	Mean	SE	Source
Anaemia	£1,592.39	£318.48	Copley-Merriman et al. (2018)[97]
Febrile neutropenia	£4,755.76	£951.15	
Leukopenia	£1,308.26	£261.65	
Lymphocyte count decreased	£1,308.26	£261.65	
Neutropenia	£1,308.26	£261.65	
Alanine aminotransferase increased	£268.61	£53.72	
Aspartate aminotransferase increased	£268.61	£53.72	
Diarrhoea	£2,426.57	£485.31	
Rash	£1,039.65	£207.93	
SE: standard error Cost for leukopenia and lymphocyte count decreased assumed to be equal to neutropenia All costs were inflated to 2019 values using PSSRU inflation factors[98] All SEs assumed to be 20% of mean value			

B.3.5.4. Miscellaneous unit costs and resource use

All costs and resource use has been detailed in Sections B.3.5.1 to B.3.5.3.

B.3.6. Summary of base-case analysis inputs and assumptions

B.3.6.1. Summary of base-case analysis inputs

Table 73 Summary of variables applied in the economic model

Variable	Value	Measurement of uncertainty and distribution	Section
Baseline parameters			
Baseline parameters	Table 38	SE (age: normal; sex: beta)	B.3.2.2
Survival and progression functions			
Overall survival	Table 40	Described in Section B.3.3.2	B.3.3.2
Progression-free survival			
All-cause mortality	Table 49	None	B.3.3.2.2
Clinical parameters			
Discontinuations	Table 51	Covariance (normal)	B.3.3.2.3
AE rates	Table 52	SE (beta)	B.3.3.2.4
Utilities			
Health state utilities	Table 53	SE (beta)	B.3.4.3
Costs			
Medication costs	Table 64	NA	B.3.5.1
Health state costs	Table 70	SE (gamma)	B.3.5.2B.3.5.2
AE costs	Table 72	SE (gamma)	B.3.5.3
AE: adverse events; NA: not applicable; SE: standard error			

B.3.6.2. Assumptions

A summary of the main assumptions applied within the economic model is provided with Table 74.

Table 74. Assumptions applied within the economic model

Assumption	Rationale	Section
Base case analysis patient parameters are derived from ATTRACTION-3, which is assumed to be reflective of patients seen in UK clinical practice.	There may be differences between baseline characteristics in ATTRACTION-3 patients and gastric cancer patients in UK clinical practice, it may still be considered representative of the types of patients who will be considered for treatment in clinical practice. Sensitivity analyses (probabilistic and deterministic) have been conducted to assess the impact of variability in these parameters, while scenarios assessed the impact of different efficacy sources on outcomes.	B.3.2.1
To reflect the nature of OC and available evidence, the model assumes that OC phases are consecutive, so that patients cannot revert to pre-progression from more advanced phases of the disease.	This assumption has been validated by clinicians and is line with other HTAs and economic analyses assessing the GC population.	B.3.2.2
Weekly cycle length	A previous gastric cancer evaluation assessed by NICE for a similar indication had applied weekly cycle lengths, which was considered appropriate by the ERG.[64] Weekly cycles were also considered appropriate for this evaluation because it enables the model to reflect the timings of drug administrations associated with both intervention and comparator therapies, and also captures a realistic minimum time during which the symptoms or response can change in UK clinical practice.	B.3.2.3
Identification of most appropriate survival curves describing PFS, OS and time on treatment	Extensive analyses have been undertaken to identify appropriate and conservative survival curves describing nivolumab efficacy, with reference to the guidance from the NICE Decision Support Unit (DSU) and Bagust and Beale (2014).[66, 67] The approach and identified survival extrapolations have been validated by clinical and health economic experts. However, to address the uncertainty around this parameter, scenario analyses have been conducted by applying alternative assumptions around extrapolations.	B.3.3.2
Treatment is assumed to continue until discontinuation due to AEs (derived from nivolumab patient-level data).	This assumption follows treatment guidelines and is therefore likely to reflect clinical practice in most patients and with most therapies. However, nivolumab use in clinical practice may vary, where treatment may be discontinued upon progression. Further, UK clinicians may wish to stop treatment in patients responding at two years. To assess the impact of these potential scenarios, alternative treatment duration assumptions have been examined as scenario analyses.	B.3.2.3.2 B.3.2.3.3
Medical resource use is derived from a clinician survey	Robust estimates of medical resource use for patients in this setting are not publicly available, given the lack of alternative treatments available for which evidence may have previously been gathered. In order to provide relevant economic resource use was derived based on expert clinical opinion (in the form of survey responses) was assessed. Additionally, a scenario analysis in which resource use was derived based on expert clinical opinion (in the form of survey responses) was assessed.	B.3.5.2

B.3.7. Base-case results

B.3.7.1. Base case incremental cost-effectiveness analysis results

The results of the base case analysis are summarised in Table 75.

In terms of comparator treatments (taxanes), the model predicts a median OS of 0.75 years, with an accrual of 0.62 discounted QALYs over the modelled time horizon. By comparison, it was predicted that the use of nivolumab will result in an additional 0.46 discounted QALYs (total: [redacted] discounted QALYs) and an additional 0.54 discounted life years (total: [redacted] discounted life years), respectively. It was estimated that patients receiving nivolumab would spend [redacted] years in the pre-progression health state (versus 0.41 for taxane), with a subsequent [redacted] years in the post-progression health state (versus 0.59 for taxane), indicating that nivolumab is associated with incremental benefit across all health states. Figure 31 demonstrates the short survival observed for taxane patients, based on modelled outcomes, and that most of the survival benefit for nivolumab over taxanes during ATTRACTION-3 is derived from the observed treatment period.

Total discounted costs associated with nivolumab (with PAS), accrued over the modelled time horizon, were predicted to be [redacted]. By comparison, total discounted costs associated with taxanes were notably lower. Incremental discounted costs were predicted to be £20,842 over taxanes, under base case assumptions. The result ICER estimate for nivolumab versus taxanes was £45,491 per QALY gain. Therefore, the base case ICER is below a £50,000 per QALY wiliness-to-pay threshold when the current nivolumab PAS discount is applied.

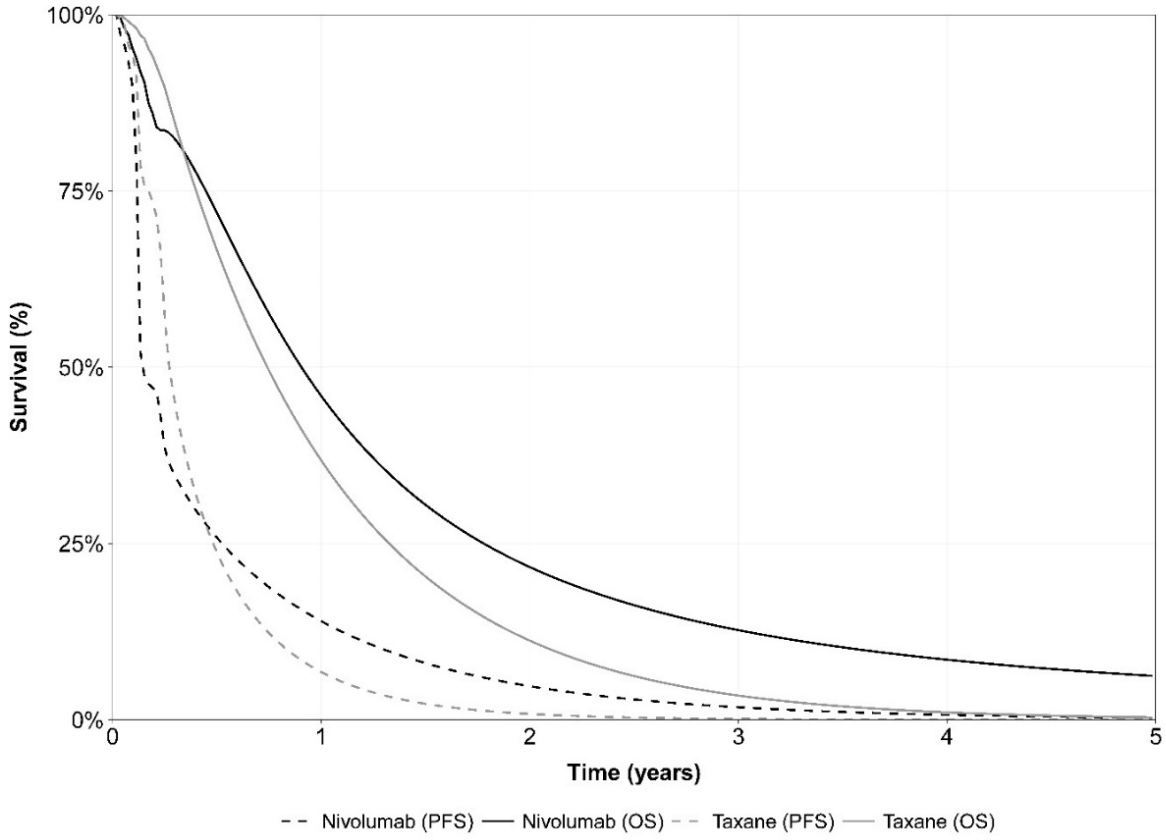


Figure 28. Overall survival and progression-free survival for nivolumab and taxanes

Table 75. Base case analysis results (with PAS, lifetime horizon)

	Nivolumab	Taxane
Patient-level survival (undiscounted)		
Median ToT (years)	0.230	0.211
Mean ToT (years)	0.496	0.291
Median PFS (years)	0.153	0.287
Mean PFS (years)	0.487	0.408
Median OS (years)	0.901	0.747
Mean OS (years)	1.650	0.997
Patient-level progression		
Time in pre-progression (years)	[redacted]	[redacted]
- Time initial therapy (years)	[redacted]	[redacted]
- Time in subsequent therapy (years)	[redacted]	[redacted]
Time in post-progression (years)	[redacted]	[redacted]
Costs (with PAS)		
HS costs	[redacted]	[redacted]
Treatment costs	[redacted]	[redacted]
BSC costs	[redacted]	[redacted]
Average AE costs per patient	[redacted]	[redacted]
Total costs	[redacted]	[redacted]
Health benefits		
Total QALYs	[redacted]	[redacted]
Total life years	[redacted]	[redacted]
Incremental results		
Incremental total costs	-	£20,842
Incremental QALYs	-	0.458
Incremental life years	-	0.536
ICER		
Cost/QALY	**-**	**£45,491**

AE: adverse event; BSC: best supportive care; HS: health state; ICER: incremental cost-effectiveness ratio; OS: overall survival; PAS: patient access scheme; PFS: progression-free survival; QALY: quality-adjusted life year; ToT: time on treatment

B.3.8. Sensitivity analyses

B.3.8.1. Probabilistic sensitivity analysis

In the probabilistic sensitivity analysis (PSA), a non-parametric bootstrapping approach was taken, sampling values from distributions around the means of input parameters in the model. Sampling utilises information of the mean and standard error of parameters to derive an estimated value using an appropriate distribution (costs: gamma, age and survival parameters: normal, utilities, probabilities and proportions: beta). These analyses are used to estimate the overall uncertainty that exists in the model results due to uncertainty in the chosen input parameters.

The majority of parameters included in the PSA are sampled independently, with the exception of survival estimates, where parameters associated with individual survival function are sampled using a common random number.

Several inputs are derived from sources where it has not been possible to ascertain standard errors. To assess uncertainty surrounding these inputs, the standard error has been assumed to be 20% of the mean value for the purposes of the PSA.

1,000 simulation of the model was deemed enough for the model results to converge to a sufficient degree of accuracy.

B.3.8.1.1. *PSA Results*

The ICER scatterplot for the base case analysis, arising from 1,000 simulations of the model with all parameters sampled is presented in Figure 32, while the cost-effectiveness acceptability curve (CEAC) is presented in Figure 33. Based on this analysis, the probability that nivolumab is cost-effective versus taxane is estimated to be [redacted] at a willingness-to-pay threshold of £50,000 per QALY. The base case results are presented in Table 76.

Table 76. Base case results (probabilistic)

Technology	Total costs (£)	Total life years	Total QALYs	Inc. costs (£)	Inc. life years	Inc. QALYs	ICER (£/QALY)
Nivolumab	[redacted]	[redacted]	[redacted]	-	-	-	
Taxane	[redacted]	[redacted]	[redacted]	£21,210	0.547	0.468	**£45,278**
ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year							

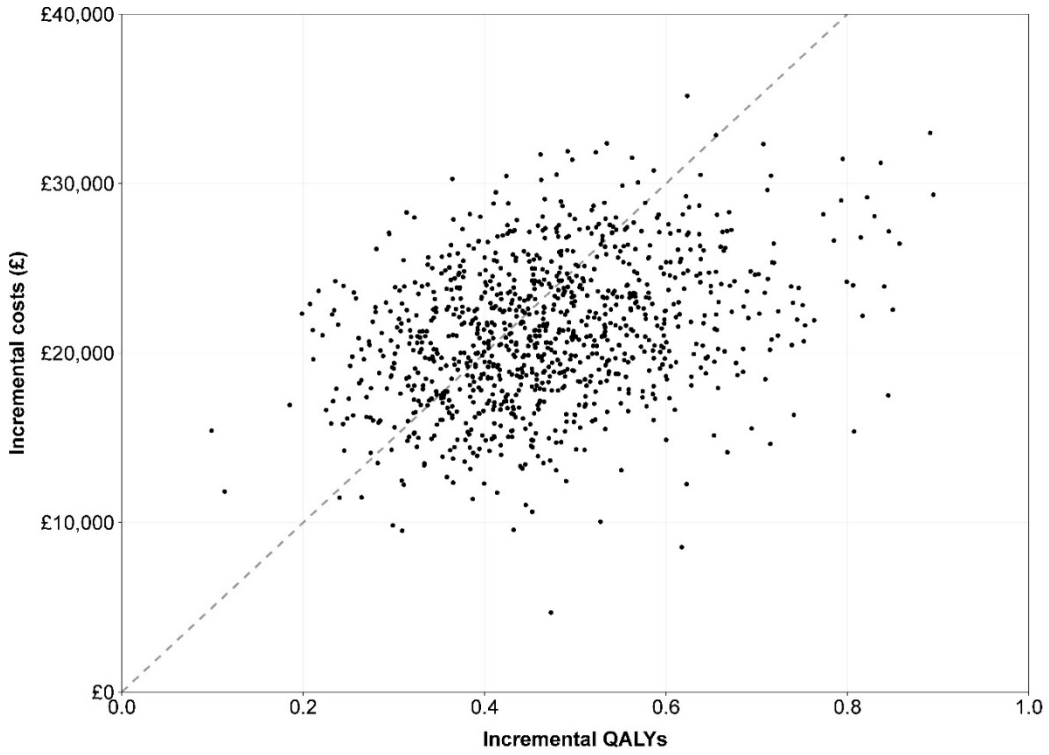


Figure 29. ICER scatterplot: Nivolumab versus taxanes

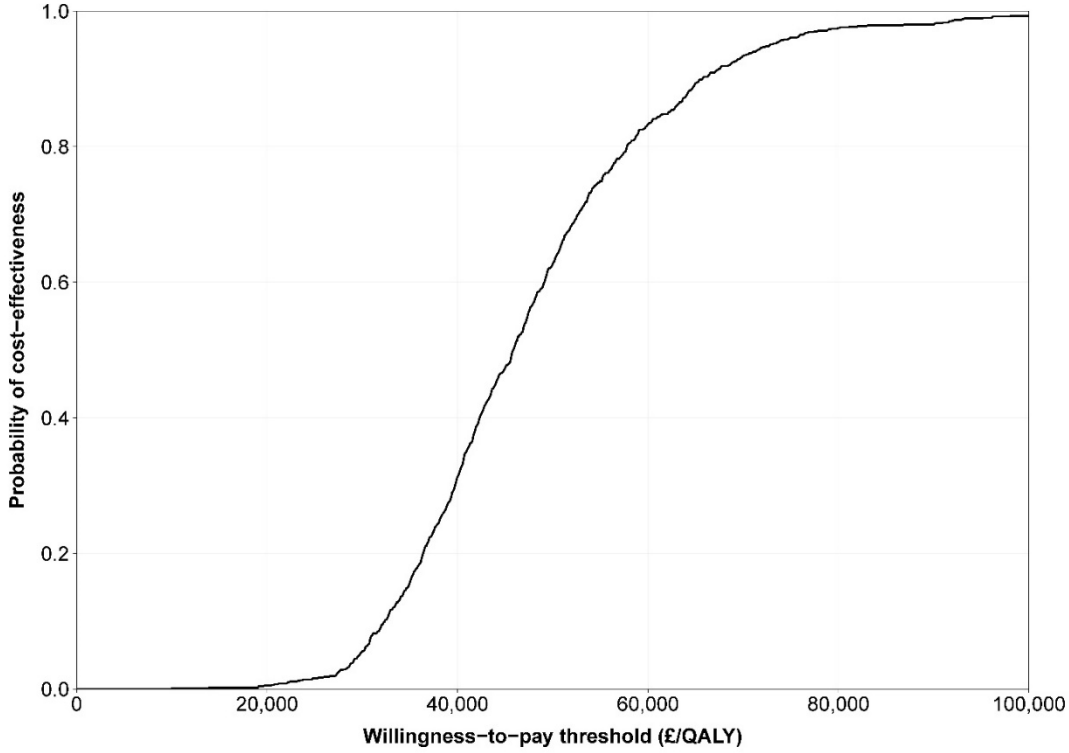


Figure 30. Cost-effectiveness acceptability curve: Nivolumab versus taxanes

B.3.8.2. Deterministic sensitivity analysis

A range of one-way (deterministic) sensitivity analyses have been conducted, regarding the following assumption and parameters:

- Time horizon (10 years)
- Discounting: costs (0% and 6%)
- Discounting: benefits (0% and 6%)
- Baseline characteristics: age (± 20%, impacting on all-cause mortality)
- Baseline characteristics: sex (0% and 100% male, impacting on all-cause mortality)
- Health state costs: pre-progression (± 20%)
- Health state costs: post-progression (± 20%)
- Health state costs: death (± 20%)
- Treatment costs: second line (± 20%)
- Treatment costs: second line BSC (± 20%)
- Treatment costs: third line BSC (± 20%)
- Adverse event costs (± 20%)
- Health state utility: pre-progression (± 20%)
- Health state utility: post-progression (± 20%)
- Proportion receiving dose (± 20%)
- Second line adverse event probability (± 20%)

Note; where (± 20%) is specified, the mean value is multiplied by 0.8 or 1.2 so to assess the impact of a 20% change in a value.

B.3.8.2.1. *Deterministic sensitivity analysis results*

Results of the deterministic sensitivity analysis is resented in Figure 34 and demonstrate the impact of specific parameters on ICER estimates. In the majority of scenarios, the ICER for nivolumab versus taxanes remained below the £50,000 per QALY willingness-to-pay threshold; scenarios where the ICER exceeded this threshold included increasing subsequent treatment costs, reducing post-progression health state utility and increasing the baseline age of patients.

Plausible alternative scenarios have been investigated further in Section 0, in order to assess the impact of the uncertainty in the analysis.

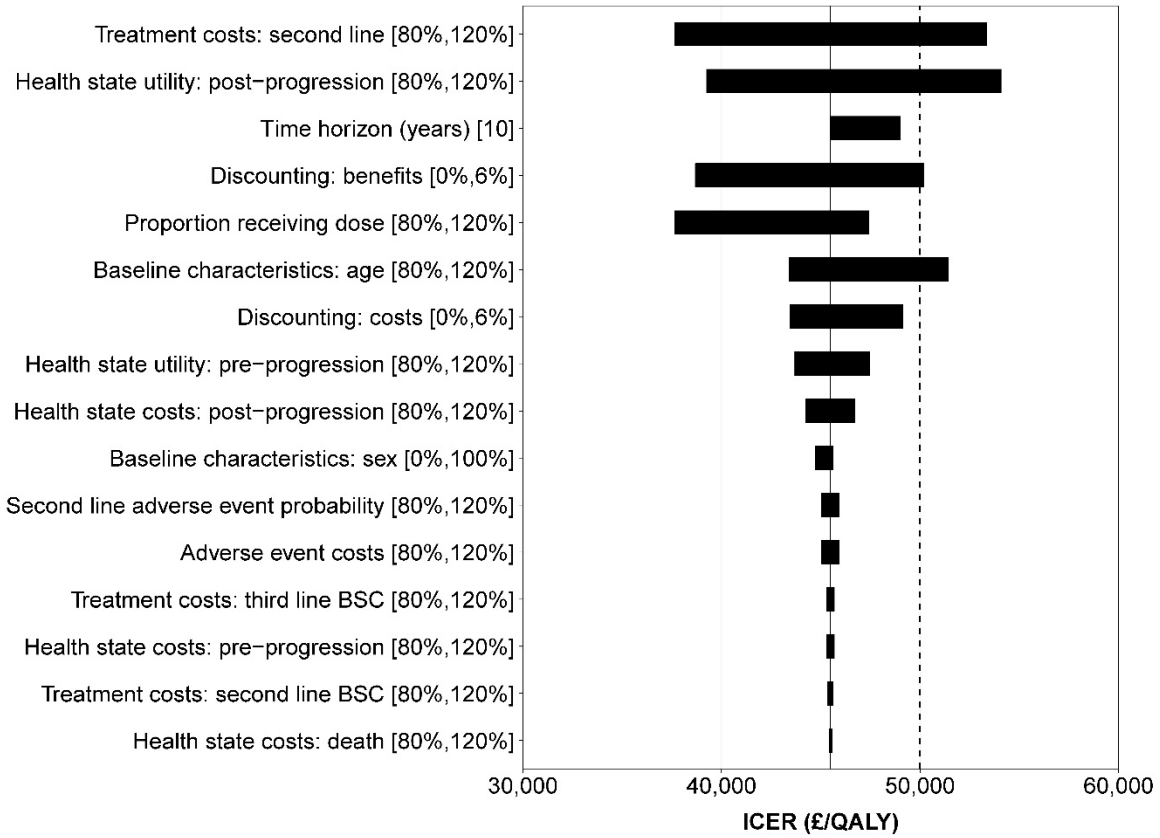


Figure 31. Deterministic sensitivity analysis: impact on ICER

B.3.8.3. Scenario analysis

B.3.8.3.1. *Alternative survival extrapolations*

Survival modelling using long-term extrapolation of parametric functions is subject to considerable uncertainty despite efforts to robustly and transparently provide survival curves that best represent patients in clinical practice. In order to assess the impact of alternative parametric fittings on the cost-effectiveness of nivolumab, survival curves described in the survival analysis report (Appendix M) have been applied within the model as scenario analyses.

This analysis should be viewed within the context of identifying the most appropriate survival extrapolation, as detailed in Section B.3.3.2. Parametric extrapolation of survival data from ATTRACTION-3 was undertaken with reference to the guidance from the NICE Decision Support Unit (DSU) and Bagust and Beale (2014).[67, 98] All extrapolations have been assessed for completeness. However, it should be noted that several of these extrapolations are not considered appropriate. Clinically implausible fits are presented in grey italics and are defined as extrapolations that exceed the 95% confidence intervals of the Kaplan-Meier data or provides mean survival that cannot be considered plausible. The impact of applying alternative survival extrapolations for the nivolumab arm (OS and PFS) in the base case analysis (ATTRACTION-3) is shown in Table 77. For OS and PFS, the majority of alternative extrapolations in grey italics were considered implausible because these extrapolations exceeded the 95% confidence intervals of the Kaplan-Meier data, with the exception of the parametric and semi-parametric gompertz curve for PFS, where the mean survival time didn't converge. The generalised gamma and semi-parametric log-logistic curves for PFS were also considered implausible due to the implausibly long mean survival time.

In addition to these clinically implausible fits, several of the clinically plausible extrapolations provided poor fits to the data, based on visual inspection of the observed Kaplan-Meier data and consideration of the cumulative hazard profile. These extrapolations did not reflect the clinical expectation of decreasing hazards over time for immunotherapy-treated patients. As demonstrated in Appendix M, the most appropriate survival extrapolations have been identified, rather than the most optimistic.

The impact of applying alternative clinically plausible extrapolations for the nivolumab arm (OS and PFS) in the base case analysis is shown in Table 77 and depicted in Figure 35. Predicted discounted incremental QALYs ranged from 0.260 to 0.508; while PFS extrapolations didn't greatly impact on the QALY gains, OS extrapolations had a large impact, with shorter extrapolations reducing survival benefit; conversely, longer extrapolations increasing QALY accrual. There was a similar variation in discounted incremental costs ranging from £18,543 to £21,426. This had an associated impact on ICERs versus taxanes, which ranged between £42,142 per QALY (when a log-logistic curve was applied for OS) and £71,434 per QALY (when a semi-parametric Weibull curve was applied for OS).

Similarly, alternative survival extrapolations were considered for the taxane arm, depicted in Table 78 and Figure 36. For OS and PFS, the majority of alternative extrapolations in grey italics were considered implausible because these extrapolations exceeded the 95% confidence intervals of the Kaplan-Meier data, with the exception of the Gompertz curves for PFS and OS, where the mean survival time didn't converge. The log-logistic curve for OS was also considered implausible due to the implausibly long mean survival time.

Predicted discounted incremental QALYs ranged from 0.324 to 0.4759, with variation in discounted incremental costs of £19,019 to £20,828. This had an associated impact on ICERs versus taxanes, which ranged between £45,308 per QALY (when a log-normal curve was applied for PFS) and £58,782 per QALY (when a semi-parametric log-logistic curve was applied for OS).

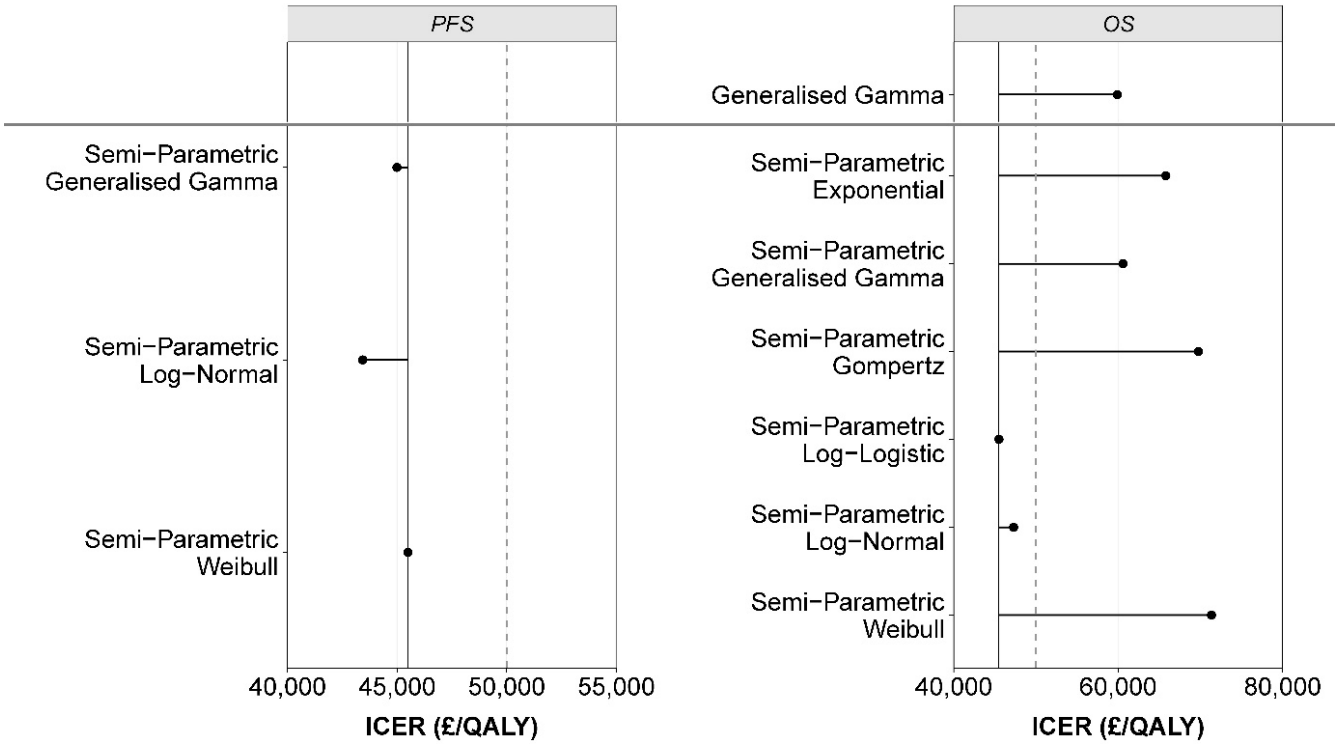


Figure 32. Scenario analysis: Impact of alternative clinically plausible survival curve extrapolation for nivolumab in the base case analysis (ATTRACTION-3)

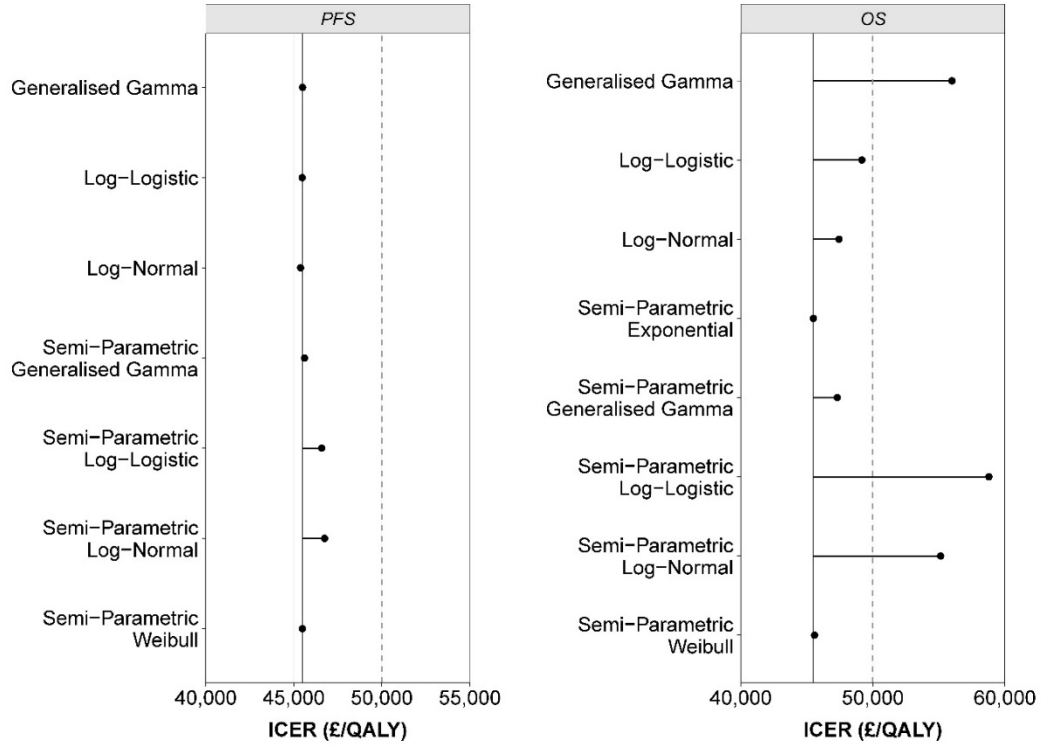


Figure 33. Scenario analysis: Impact of alternative clinically plausible survival curve extrapolation for taxanes in the base case analysis (ATTRACTION-3)

Table 77. Scenario analysis: impact of alternative nivolumab extrapolation in the base case analysis (discounted outcomes, ATTRACTION-3 nivolumab arm)

Scenario			Nivolumab		
			Inc. QALY	Inc. Cost (£)	ICER (£/ QALY)
PFS	Parametric	*Exponential*	*0.451*	*£20,842*	*£46,183*
		Generalised Gamma	*0.487*	*£20,842*	*£42,820*
		Gompertz	*0.506*	*£20,842*	*£41,151*
		Log-logistic	*0.449*	*£20,842*	*£46,416*
		Log-normal	*0.450*	*£20,842*	*£46,299*
		Weibull	*0.452*	*£20,842*	*£46,122*
	Semi-parametric with Kaplan-Meier to 2.99 months	Exponential	0.455	£20,842	£45,759
		Generalised Gamma	0.463	£20,842	£44,986
		Gompertz	*0.500*	*£20,842*	*£41,694*
		Log-logistic	*0.479*	*£20,842*	*£43,485*
		Log-normal	0.480	£20,842	£43,423
		Weibull	0.458	£20,842	£45,491
OS	Parametric	*Exponential*	*0.289*	*£18,871*	*£65,236*
		Generalised Gamma	0.322	£19,268	£59,873
		Gompertz	*0.263*	*£18,564*	*£70,649*
		Log-logistic	0.508	£21,426	£42,142
		Log-normal	0.455	£20,798	£45,736
		Weibull	*0.264*	*£18,595*	*£70,384*
	Semi-parametric with Kaplan-Meier to 2.99 months	Exponential	0.286	£18,848	£65,796
		Generalised Gamma	0.317	£19,214	£60,571
		Gompertz	0.267	£18,620	£69,743
		Log-logistic	0.458	£20,842	£45,491
		Log-normal	0.435	£20,577	£47,269
		Weibull	0.260	£18,543	£71,343

ICER: incremental cost-effectiveness ratio; Inc: incremental; OS: overall survival; PFS: progression-free survival; QALY: quality-adjusted life year
Grey italics denotes highly implausible extrapolations, defined as extrapolations that exceed the 95% confidence intervals of the Kaplan-Meier data or provides mean survival that cannot be considered plausible.

Table 78. Scenario analysis: Impact of alternative chemotherapy extrapolation in the base case analysis (discounted outcomes, ATTRACTION-3 taxane arm)

Scenario			Taxane		
			Inc. QALY	Inc. Cost (£)	ICER (£/ QALY)
PFS	Parametric	*Exponential*	*0.458*	*£20,842*	*£45,544*
		Generalised Gamma	0.458	£20,842	£45,503
		Gompertz	*0.458*	*£20,842*	*£45,476*
		Log-logistic	0.458	£20,842	£45,482
		Log-normal	0.459	£20,842	£45,380
		Weibull	*0.460*	*£20,842*	*£45,354*
	Semi-parametric with Kaplan-Meier to 2.99 months	*Exponential*	*0.460*	*£20,842*	*£45,352*
		Generalised Gamma	0.457	£20,842	£45,606
		Gompertz	*0.450*	*£20,842*	*£46,304*
		Log-logistic	0.447	£20,842	£46,584
		Log-normal	0.446	£20,842	£46,747
		Weibull	0.458	£20,842	£45,491
OS	Parametric	*Exponential*	*0.446*	*£20,761*	*£46,509*
		Generalised Gamma	0.345	£19,320	£55,989
		Gompertz	*0.473*	*£21,091*	*£44,595*
		Log-logistic	0.411	£20,210	£49,165
		Log-normal	0.432	£20,492	£47,434
		Weibull	*0.469*	*£21,000*	*£44,756*
	Semi-parametric with Kaplan-Meier to 2.99 months	Exponential	0.458	£20,842	£45,491
		Generalised Gamma	0.434	£20,515	£47,306
		Gompertz	*0.434*	*£20,524*	*£47,271*
		Log-logistic	0.324	£19,019	£58,782
		Log-normal	0.352	£19,415	£55,130
		Weibull	0.457	£20,828	£45,580

ICER: incremental cost-effectiveness ratio; Inc: incremental; OS: overall survival; PFS: progression-free survival; QALY: quality-adjusted life year
Grey italics denotes highly implausible extrapolations, defined as extrapolations that exceed the 95% confidence intervals of the Kaplan-Meier data or provides mean survival that cannot be considered plausible.

B.3.8.3.2. *Alternative comparators*

The base case analysis informed by ATTRACTION-3 compares nivolumab versus taxane, defined as the combination of docetaxel and paclitaxel chemotherapies. As outlined in Section B.3.2.4, this can be considered clinically appropriate based on current guidelines, clinical evidence and expert opinion. Further, the ATTRACTION-3 study was powered to show differences in efficacy for nivolumab against the combined taxane arm, as opposed to docetaxel and paclitaxel separately. Low patient numbers receiving individual treatments may impact on outcomes, particularly during later periods of follow-up. Hence, it is more appropriate to use the combined taxane arm as a comparator.

However, in order to inform decision-making, a comparison of nivolumab against docetaxel and paclitaxel separately has been provided as a scenario analysis. Survival extrapolation for each comparator arm was fitted using the same methodology as the base case, detailed in

Section B.3.3.2. Detailed explanation and rationales for extrapolation are provided in Appendix M. Outcomes for BSC OS are described in Appendix L; PFS outcomes for BSC are assumed based on the ratio of PFS:OS outcomes for docetaxel.

As described in Table 79, predicted discounted incremental QALYs ranged from 0.401 (versus docetaxel) to 0.414 (versus paclitaxel) to 0.630 (versus BSC), with variation in discounted incremental costs from £20,971 to £19,371 to £30,434 , versus paclitaxel, docetaxel and BSC, respectively. The resultant ICER estimate for nivolumab versus docetaxel was £52,340 per QALY and for nivolumab versus paclitaxel £46,764 per QALY and £48,298 versus BSC.

Table 79. Impact of alternative comparators

	Nivolumab	Docetaxel	Paclitaxel	BSC
Incremental QALYs	-	0.401	0.414	0.630
Incremental life years	-	0.425	0.482	0.819
Incremental costs	-	£20,971	£19,371	£30,434
ICER (£/QALY)	**-**	**£52,340**	**£46,764**	**£48,298**
BSC: best supportive care; ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year				

B.3.8.3.3. *Impact of alternative treatment stopping rules*

In the base case analysis, treatment cessation is based on time on treatment data from ATTRACTION-3. However, clinical practice may vary, where nivolumab treatment may cease upon progression or clinicians may wish to stop treatment in patients responding at two years. Hence, two scenario analyses were conducted, assessing each potential treatment strategy.

B.3.8.3.3.1. Stopping rule scenario

This scenario involves applying clinical efficacy inputs from the base case analysis where no stopping rules were applied to patients in the taxane arm, while patients receiving nivolumab and remaining on initial therapy at 24 months cease to receive nivolumab.

Results from the analysis is detailed in Table 80, where application of a 2-year stopping rule for nivolumab resulted in an ICER estimate of £40,909 per QALY, which signals a reduction in the estimate from the base case (£45,491 per QALY).

Table 80. Impact of applying a 2-year stopping rule for nivolumab

	Nivolumab	Taxane
Total QALYs	[redacted]	[redacted]
Total life years	[redacted]	[redacted]
Total costs	[redacted]	[redacted]
Incremental QALYs	-	£18,743
Incremental life years	-	0.458
Incremental costs	-	0.536
ICER (£/QALY)	**-**	**£40,909**
ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year		

B.3.8.3.3.2. No treatment beyond progression

This scenario involves the cessation of treatment upon progression in patients receiving nivolumab. Upon discontinuation, patients receiving nivolumab move to third line BSC.

Results from the analysis is detailed in Table 81, where patients discontinued nivolumab upon progression resulted in an ICER estimate of £45,455 per QALY, which is comparable to the base case ICER (£45,491 per QALY).

Table 81. Impact of removing treatment beyond progression for nivolumab

	Nivolumab	Taxane
Total QALYs	[redacted]	[redacted]
Total life years	[redacted]	[redacted]
Total costs	[redacted]	[redacted]
Incremental QALYs	-	£20,825
Incremental life years	-	0.536
Incremental costs	-	0.458
ICER (£/QALY)	**-**	**£45,455**
ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year		

B.3.8.3.4. *Summary of sensitivity analyses results*

Several sensitivity analyses have been undertaken, assessing the impact of variation in all variables and assumptions applied within the model, In the deterministic analysis, in the majority of scenarios nivolumab remained cost-effective at a willingness-to-pay threshold of £50,000 per QALY. Similarly, in the PSA, the probability that nivolumab of cost-effective versus taxanes is [redacted] at a willingness-to-pay threshold of £50,000 per QALY. Plausible alternative inputs and assumptions were assessed as scenario analyses within Section 0; again the majority of these scenarios resulting in cost-effective ICERs at the £50,000 per QALY threshold.

B.3.9. Subgroup analysis

All relevant subgroup analyses are presented in Section 0.

B.3.10. Validation

B.3.10.1. Validation of cost-effectiveness analysis

In the specific context of oesophageal cancer patients who already received prior therapy, patient numbers are low and survival outcomes are poor. Thus, there is a distinct paucity of evidence describing current treatment pathways, resource use and costs in UK clinical practice on which to base economic evaluation. In general, where no evidence has been identified, pragmatic assumptions have been made based on independent sources, such as

published literature, oesophageal cancer guidelines or previous NICE appraisals; as no previous NICE appraisals for oesophageal cancer therapies were identified, assessments of gastric cancer therapies have been used where relevant. These assumptions were then assessed for clinical plausibility; uncertainty has been characterised through the use of sensitivity analyses. Extensive sensitivity analyses were then undertaken, and the majority of ICERs remain below the £50,000/QALY threshold.

A technical review of the cost-effectiveness model was conducted by an independent consultant. Further, the relevance of the model structure and assumptions was validated at an Advisory Board. This allowed the model approach to be validated and permitted areas of disagreement to be resolved prior to generation of model results. In addition, quality control was undertaken, whereby a cell-by-cell verification process was conducted to allow checking of all input calculation, formulae and visual basic code.

B.3.10.2. Validation of nivolumab survival extrapolation

As described in B.3.3.2.1.3, there are no other studies with which to validate the results for extrapolation of the nivolumab arm other than ATTRACTION-3 and ATTRACTION-1, both of which have been compared against survival extrapolations.

Despite the lack of real-world data, it was possible to validate the survival extrapolation for nivolumab against longer-term survival data from studies evaluating other indications using immunotherapy agents. Available long-term data are presented in Table 82 for nivolumab in various other indications. As can be seen, there is typically an initial high rate of mortality followed by a lower rate of mortality over long-term follow-up. Long term survivorship without the need for prolonged treatment has been observed for immunotherapies in other indications. Long term survivorship without the need for prolonged treatment has been observed for immunotherapies in other indications. For example, ipilimumab therapy administered for four cycles at three-weekly intervals can lead to ten-year survival in 20-25% of melanoma patients, as presented in Table 82.[99]

Table 82. Comparison of OS outcomes and predicted survival extrapolations for nivolumab in melanoma, renal cell carcinoma and non-small cell lung cancer

	Melanoma			RCC			Non-Squamous NSCLC			OSCC		
	Modelled OS	Observed OS	Source	Modelled OS	Observed OS	Source	Modelled OS	Observed OS	Source	Modelled OS	Observed OS	Source
6 months	83%	85%	CheckMate 067; Wolchok 2017[100]	83.4%	89%	CheckMate 025; Motzer 2015[101]	-	-	-	71.9%	71.9%	ATTRACTION-3
12 months	68%	74%		74.5%	76%		47.8%	51%	CheckMate 057; Borghaei 2016[102]	46.9%	45.9%	
18 months	57%	65%		62.5%	63%		36.6%	39%		30.5%	-	
24 months	49%	59%		52.5%	52%		29%	29%		19.1%	21.6%	
36 months	40.1%	52%		38%	44%	CheckMate 003; McDermott 2015[103]	19%	18%	CheckMate 003, Gettinger 2015[104]	-	-	
48 months	35%	35%	CheckMate 003; Hodi 2016[112]	28%	38%	CheckMate 003; McDermott 2016[105]	14.4%	NA	NA	-	-	
60 months	32.8%	34%		20.9%	34%		11.5%	16%	CheckMate 003, Brahmer 2017[106]	-	-	

NA: not available; NSCLC: non-small cell lung cancer; OS: overall survival; RCC: renal cell carcinoma.

B.3.11. Interpretation and conclusions of economic evidence

Base case analysis

- In line with estimates of short life expectancy in patients receiving taxanes, the base case analysis predicts median OS of 0.75 years (mean 1.00 years), informed by a randomised-controlled trial.
- Use of nivolumab will result in an increased mean OS of 1.65 years, as well as additional discounted QALYs and life years of [redacted] and [redacted], respectively.
- Based on mean OS outcomes for patients treated with taxanes (12.0 months) and mean OS benefit associated with nivolumab (incremental 7.8 months), end of life criteria can be considered to be met.
- Discounted incremental costs were estimated to be £20,842 under base case assumptions and the resultant ICER was £45,491 per QALY, which is considered to be cost-effective at a willingness-to-pay threshold of £50,000 per QALY.

Sensitivity analysis

- In the deterministic and probabilistic sensitivity analyses, nivolumab was cost-effective in the majority of scenarios at a willingness-to-pay threshold of £50,000 per QALY.
- Extensive scenario analyses were undertaken, reflecting the assumptions required to undertake plausible, robust and transparent base case analysis.
- Within these scenario analyses, the majority of ICERs remain below the £50,000 per QALY threshold.

As previously noted, this analysis has been conducted where there is a paucity of evidence necessitating several pragmatic assumptions, which have been made based on independent sources, such as published literature, oesophageal cancer guidelines or previous NICE appraisals. These assumptions have been assessed through sensitivity analysis and scenario analysis in order to assess the impact of uncertainty. Further, the modelling approach has been chosen to reflect the most important treatment outcomes for most oesophageal cancer patients: survival (progression free and overall), side effects and quality of life.

In the base case analysis, it was estimated that nivolumab use would result in 1.073 discounted QALYs and 1.506 discounted LYs. Further, it was estimated that patients receiving nivolumab would spend [redacted] years in the pre-progression state (versus [redacted] years for patients receiving taxanes), with a subsequent [redacted] years in the post-progression state (versus [redacted] years for taxanes), indicating that nivolumab is associated with incremental benefit across all

health states, and also that the majority of survival benefit over taxanes is derived from the observed treatment period. Discounted incremental costs were expected to be £20,842 over taxanes under base case assumptions and the resultant ICER was £45,491, which can be considered cost-effective at a willingness-to-pay threshold of £50,000/QALY.

A large number of sensitivity analyses have been undertaken, assessing the impact of variation in all variables and assumptions applied within the model. In the deterministic analysis and PSA, nivolumab was cost-effective in the majority of scenarios at a WTP threshold of £50,000/QALY. Similarly, when plausible alternative inputs and assumptions were assessed as scenario analyses within Section B.3.8, the majority of ICERs remain below the £50,000/QALY threshold. This indicates that the ICER is relatively stable across analyses.

The availability of nivolumab for adults with previously treated oesophageal cancer would provide an opportunity to make a significant and substantial impact on health-related benefits, address a current unmet need, and would represent a further, significant advance in the management of this end of life condition.

B.3.11.1. Application of NICE end of life criteria to nivolumab use in OC

End of life criteria as applied by NICE are summarised as follows:

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

Current standard of care for previously treated oesophageal cancer are taxanes and BSC, both of which are associated with poor outcomes; median OS from ATTRACTION-3 was low (8.38 months) and survival at two years was 15.1%. These data are supported by SLR evidence described in Appendix D. Therefore, a high degree of unmet medical need remains for effective and tolerable treatments for this patient population.

Further, application of NICE end-of-life criteria to nivolumab use in previously treated oesophageal cancer should be set in the context of low patient numbers and the very high unmet need.

The case for application of NICE end-of-life criteria is set out in Section B.2.12.5 and based on this evidence, it can be considered that nivolumab meets both criteria for end-of-life.

B.4 References

1. Ohigashi Y, Sho M, Yamada Y, Tsurui Y, Hamada K, Ikeda N, et al. Clinical significance of programmed death-1 ligand-1 and programmed death-1 ligand-2 expression in human esophageal cancer. Clinical cancer research. 2005;11(8):2947-53.
2. Haymarket Media Group Ltd. Monthly Index of Medical Specialities.2017 1 September 2017. Available from: http://www.mims.co.uk.
3. Cancer Research UK. What is oesophageal cancer? 2019 [Available from: https://www.cancerresearchuk.org/about-cancer/oesophageal-cancer/about.
4. Cancer Research UK. Oesophageal cancer incidence statistics. 2019 [Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/oesophageal-cancer/incidence.
5. BMJ Best Practice. Oesophageal cancer 2018 [Available from: https://bestpractice.bmj.com/topics/en-gb/1029/pdf/1029.pdf.
6. National Cancer Institute. SEER Cancer Statistics Review (CSR) 1975-2016 2019 [Available from: https://seer.cancer.gov/csr/1975_2016/results_merged/sect_08_esophagus.pdf.
7. Cancer Research UK. Oesophageal cancer survival. 2016 [Available from: https://www.cancerresearchuk.org/about-cancer/oesophageal-cancer/survival.
8. National Institute for Health and Care Excellence. Oesophago-gastric cancer: assessment and management in adults. NICE guideline [NG83].2018 11 September 2018. Available from: https://www.nice.org.uk/guidance/ng83.
9. National Institute for Health and Care Excellence. Oesophageal and gastric cancer overview. NICE pathways.2018 12 September 2018. Available from: https://pathways.nice.org.uk/pathways/oesophageal-and-gastric-cancer#path=view%3A/pathways/oesophageal-and-gastric-cancer/oesophageal-and-gastric-cancer-overview.xml&content=view-index.
10. National Institute for Health and Care Excellence. Oesophago-gastric cancer. Quality standard [QS176] 2018 [updated 28 May 2019. Available from: https://www.nice.org.uk/guidance/QS176.
11. Lordick F, Mariette C, Haustermans K, Obermannová R, Arnold D. Oesophageal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of Oncology. 2016;27(suppl 5):v50-v7.
12. Cancer Research UK. Oesophageal cancer statistics. 2017 21 April 2017. Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/oesophageal-cancer.
13. Smyth EC, Lagergren J, Fitzgerald RC, Lordick F, Shah MA, Lagergren P, et al. Oesophageal cancer. Nature Reviews Disease Primers. 2017;3(1):17048.
14. Rice TW KD, Blackstone EH, et al. Esophagus and esophagogastric junction. In: Amin MB ES, Greene FL, et al., editor. AJCC Cancer Staging Manual. 8th ed. New York: Springer; 2017. p. 185-202.
15. Wu S-G, Zhang W-W, He Z-Y, Sun J-Y, Chen Y-X, Guo L. Sites of metastasis and overall survival in esophageal cancer: a population-based study. Cancer Manag Res. 2017;9:781-8.
16. Wu S-G, Zhang W-W, Sun J-Y, Li F-Y, Lin Q, He Z-Y. Patterns of Distant Metastasis Between Histological Types in Esophageal Cancer. Front Oncol. 2018;8(302).

17. Ai D, Zhu H, Ren W, Chen Y, Liu Q, Deng J, et al. Patterns of distant organ metastases in esophageal cancer: a population-based study. J Thorac Dis. 2017;9(9):3023-30.
18. Cancer Research UK. Oesophageal cancer mortality statistics 2018 [Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/oesophageal-cancer/mortality.
19. Cancer Research UK. Oesophageal cancer survival statistics. 2014 [updated 21 June 2019. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/oesophageal-cancer/survival.
20. Wu S-G, Xie W-H, Zhang Z-Q, Sun J-Y, Li F-Y, Lin H-X, et al. Surgery Combined with Radiotherapy Improved Survival in Metastatic Esophageal Cancer in a Surveillance Epidemiology and End Results Population-based Study. Sci Rep. 2016;6:28280-.
21. Moriwaki T, Kajiwara T, Matsumoto T, Suzuki H, Hiroshima Y, Matsuda K, et al. Survival analysis of platinum-refractory patients with advanced esophageal cancer treated with docetaxel or best supportive care alone: a retrospective study. Diseases of the esophagus : official journal of the International Society for Diseases of the Esophagus. 2014;27(8):737-43.
22. Nakatsumi H, Komatsu Y, Sawada K, Muranaka T, Kawamoto Y, Yuki S, et al., editors. Retrospective comparison of efficacy and safety of docetaxel and weekly-paclitaxel as 2nd-line chemotherapy for patients with unresectable or recurrent esophageal cancer. ANNALS OF ONCOLOGY; 2016: OXFORD UNIV PRESS GREAT CLARENDON ST, OXFORD OX2 6DP, ENGLAND.
23. Nomura M, Iwasa S, Tsushima T, Kato K, Yasui H, Boku N, et al. Active salvage chemotherapy versus best supportive care for patients with recurrent or metastatic squamous cell carcinoma of the esophagus refractory or intolerable to fluorouracil, platinum, and taxane. Cancer chemotherapy and pharmacology. 2016;78(6):1209-16.
24. Tsushima T, Nomura M, Iwasa S, Kato K, Yasui H, Muro K, et al. Salvage chemotherapy versus best supportive care in patients with recurrent or metastatic squamous cell carcinoma of the esophagus refractory or intolerable to fluorouracil, platinum, and taxane. Journal of Clinical Oncology. 2015;33(15_suppl):e15101-e.
25. Mellman I, Coukos G, Dranoff G. Cancer immunotherapy comes of age. Nature. 2011;480(7378):480-9.
26. Brahmer JR, Drake CG, Wollner I, Powderly JD, Picus J, Sharfman WH, et al. Phase I study of single-agent anti-programmed death-1 (MDX-1106) in refractory solid tumors: safety, clinical activity, pharmacodynamics, and immunologic correlates. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2010;28(19):3167-75.
27. Freeman GJ, Long AJ, Iwai Y, Bourque K, Chernova T, Nishimura H, et al. Engagement of the PD-1 immunoinhibitory receptor by a novel B7 family member leads to negative regulation of lymphocyte activation. The Journal of experimental medicine. 2000;192(7):1027-34.
28. Blank C, Gajewski TF, Mackensen A. Interaction of PD-L1 on tumor cells with PD-1 on tumor-specific T cells as a mechanism of immune evasion: implications for tumor immunotherapy. Cancer immunology, immunotherapy : CII. 2005;54(4):307-14.

29. Latchman Y, Wood CR, Chernova T, Chaudhary D, Borde M, Chernova I, et al. PD-L2 is a second ligand for PD-1 and inhibits T cell activation. Nature immunology. 2001;2(3):261-8.
30. Ansell SM. Targeting immune checkpoints in lymphoma. Current opinion in hematology. 2015;22(4):337-42.
31. McDermott DF, Atkins MB. PD-1 as a potential target in cancer therapy. Cancer medicine. 2013;2(5):662-73.
32. West HJ. Immune checkpoint inhibitors. JAMA oncology. 2015;1(1):115-.
33. Chiou VL, Burotto M. Pseudoprogression and immune-related response in solid tumors. Journal of Clinical Oncology. 2015;33(31):3541-3.
34. Borcoman E, Nandikolla A, Long G, Goel S, Le Tourneau C. Patterns of Response and Progression to Immunotherapy. American Society of Clinical Oncology Educational Book. 2018(38):169-78.
35. NHS England. Improving care for older people 2019 [Available from: https://www.england.nhs.uk/ourwork/clinical-policy/older-people/improving-care-for-older-people/.
36. The Kings Fund. We must end ageism and age discrimination in health and social care 2013 [Available from: https://www.kingsfund.org.uk/blog/2013/05/we-must-end-ageism-and-age-discrimination-health-and-social-care.
37. Ono Pharmaceutical Co. Ltd. ONO-4538 Phase III Study. A Multicenter, Randomized, Open-label Study in Patients with esophageal Cancer refractory or intolerant to Combination Therapy with Fluoropyrimidine and Platinum-based Drugs. Protocol (ONO-4538-24). 2016.
38. Ono Pharmaceutical Co. Ltd. ONO-4538 Phase III Study. A multicenter, randomized, open-label study in patients with esophageal cancer refractory or intolerant to combination therapy with fluoropyrimidine- and platinum-based drugs - study report. 2019.
39. Ono Pharmaceutical Co. Ltd. ONO-4538 Phase II Study. A multicenter, open-label, uncontrolled study in esophageal cancer. Protocol (ONO-4538-07). 2015.
40. Kudo T, Hamamoto Y, Kato K, Ura T, Kojima T, Tsushima T, et al. Nivolumab treatment for oesophageal squamous-cell carcinoma: an open-label, multicentre, phase 2 trial. The Lancet Oncology. 2017;18(5):631-9.
41. Ono Pharmaceutical Co. Ltd. ONO-4538 Phase II Study A Multicenter, Open-Label, Uncontrolled Study in Patients with Esophageal Cancer - study report. 2016.
42. ClinicalTrials.gov. Study of Nivolumab in Unresectable Advanced or Recurrent Esophageal Cancer (ONO-4538-24).2017 22 October 2018. Available from: https://clinicaltrials.gov/ct2/show/study/NCT02569242?show_locs=Y#locn.
43. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). European journal of cancer (Oxford, England : 1990). 2009;45(2):228-47.
44. Wolchok JD, Hoos A, O'Day S, Weber JS, Hamid O, Lebbe C, et al. Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. Clinical cancer research : an official journal of the American Association for Cancer Research. 2009;15(23):7412-20.
45. Zhang B, Wang X, Li Q, Mo H, Wang X, Song Y, et al. Efficacy of irinotecan-based chemotherapy after exposure to an anti-PD-1 antibody in patients with

advanced esophageal squamous cell carcinoma. Chin J Cancer Res. 2019;31(6):910-7.
46. Auzolle C, Dubreuil O, Pozet A, Coriat R, Dhooge M, Ducreux M, et al. 2316 Efficacy and toxicity of second-line chemotherapy in patients with advanced oesophageal squamous cell carcinoma progressing after a first line of 5-fluorouracil and platinum-based therapy: An AGEO retrospective multicentric study. European Journal of Cancer. 2015;51:S438.
47. Shirakawa T, Kato K, Nagashima K, Nishikawa A, Sawada R, Takahashi N, et al. A retrospective study of docetaxel or paclitaxel in patients with advanced or recurrent esophageal squamous cell carcinoma who previously received fluoropyrimidine- and platinum-based chemotherapy. Cancer chemotherapy and pharmacology. 2014;74(6):1207-15.
48. Song Z, Zhang Y. Second-line docetaxel-based chemotherapy after failure of fluorouracil-based first-line treatment for advanced esophageal squamous cell carcinoma. OncoTargets and therapy. 2014;7:1875-81.
49. Yamazaki K, Hironaka S, Boku N, Yasui H, Fukutomi A, Yoshino T, et al. A retrospective study of second-line chemotherapy for unresectable or recurrent squamous cell carcinoma of the esophagus refractory to chemotherapy with 5-fluorouracil plus platinum. International journal of clinical oncology. 2008;13(2):150-5.
50. Kim EJ, Kim JH, Song HJ, Park SR, Kim YH, Kim HR, et al. Impact of sequential lines of palliative chemotherapy in patients with recurrent/metastatic esophageal squamous cell carcinoma: A retrospective analysis of 107 patients at a single center. Asia Pac J Clin Oncol. 2019.
51. Sakamoto T, Takegawa N, Kushida S, Tsumura H, Mimura T, Tobimatsu K, et al. P2-6-3A RETROSPECTIVE STUDY OF WEEKLY PACLITAXEL AS SECOND-LINE TREATMENT FOR ADVANCED OR RECURRENT ESOPHAGEAL CANCER. Annals of Oncology. 2014;25(suppl_5):v93-v.
52. Kitagawa Y, Doki Y, Kato K, Ura T, Kojima T, Tsushima T. Two-year survival and safety update for esophageal squamous cell carcinoma treated with nivolumab (ATTRACTION-01/ONO-4538-07). European Society for Medical Oncology; Presented at ESMO Madrid (P638).2017.
53. Huang J, Xu B, Liu Y, Huang J, Lu P, Ba Y, et al. Irinotecan plus S-1 versus S-1 in patients with previously treated recurrent or metastatic esophageal cancer (ESWN 01): a prospective randomized, multicenter, open-labeled phase 3 trial. Cancer Commun (Lond). 2019;39(1):16.
54. Kato K, Cho BC, Takahashi M, Okada M, Lin CY, Chin K, et al. Nivolumab versus chemotherapy in patients with advanced oesophageal squamous cell carcinoma refractory or intolerant to previous chemotherapy (ATTRACTION-3): a multicentre, randomised, open-label, phase 3 trial. Lancet Oncol. 2019;20(11):1506-17.
55. Auzolle C, Landi B, Pozet A, Coriat R, Dhooge M, Ducreux M, et al. Efficacy and toxicity of second-line chemotherapy in patients with advanced oesophageal squamous cell carcinoma progressing after a first line of 5-fluorouracil and platinumbased therapy: An ageo retrospective multicentric study. United European Gastroenterology Journal. 2015;3(5):A482.
56. Dias S, Welton, N.J., Sutton, A.J., Ades. A.E. NICE DSU TECHNICAL SUPPORT DOCUMENT 2:A GENERALISED LINEAR MODELLING

FRAMEWORKFOR PAIRWISE AND NETWORK META-ANALYSIS OF RANDOMISED CONTROLLED TRIALS. 2011.
57. Dias S, Sutton, A.J., Welton, N.J., Ades, A.E. ,. NICE DSU Technical Support Document 3: Heterogeneity: subgroups, meta-regression, bias and bias-adjustment. . 2011; last updated April 2012.
58. Bristol-Myers Squibb Pharmaceuticals limited. OPDIVO 10 mg/mL concentrate for solution for infusion 2019 [Available from: https://www.medicines.org.uk/emc/product/6888/smpc#UNDESIRABLE_EFFECTS.
59. Bristol-Myers Squibb Pharmaceutical Limited. Summary of Product Characteristics. OPDIVO 10 mg/mL concentrate for solution for infusion 2017 [Available from: https://www.medicines.org.uk/emc/medicine/30476.
60. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013 2013 [Available from: https://www.nice.org.uk/article/pmg9.
61. Janmaat VT, Bruno MJ, Polinder S, Lorenzen S, Lordick F, Peppelenbosch MP, et al. Cost-Effectiveness of Cetuximab for Advanced Esophageal Squamous Cell Carcinoma. PloS one. 2016;11(4):e0153943.
62. National institute for Health and Care Excellence Decision Support Unit. Technical Support Document 18: Methods for Population-Adjusted Indirect Comparisons in Submissions to NICE.2016 30 January 2017. Available from: http://www.nicedsu.org.uk/Populationadjusted-ICs-TSD(3026862).htm.
63. Bristol-Myers Squibb Company. Nivolumab in advanced gastric cancer. Clinical advisory board meeting minutes (17 May 2017). 2017.
64. National Institute for Health and Care Excellence. Ramucirumab for treating advanced gastric cancer or gastro–oesophageal junction adenocarcinoma previously treated with chemotherapy (TA378). 2016.
65. National Institute for Health and Care Excellence. Trastuzumab emtansine for treating HER2-positive advanced breast cancer after trastuzumab and a taxane. Technology appraisal guidance [TA458].2015 22 April 2020. Available from: https://www.nice.org.uk/guidance/TA458.
66. National Institute for Health and Care Excellence Decision Support Unit. Technical Support Document 14: Survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data.2013 7 June 2016. Available from: http://www.nicedsu.org.uk/NICE%20DSU%20TSD%20Survival%20analysis.updated%20March%202013.v2.pdf.
67. Bagust A, Beale S. Survival Analysis and Extrapolation Modeling of Time-to-Event Clinical Trial Data for Economic Evaluation An Alternative Approach. Medical Decision Making. 2014;34(3):343-51.
68. National Institute for Health and Care Excellence. Technology appraisal guidance [TA428]. Pembrolizumab for treating PD-L1-positive non-small-cell lung cancer after chemotherapy.2017 14 September 2017. Available from: https://www.nice.org.uk/guidance/ta428.
69. National Institute for Health and Care Excellence. Pembrolizumab for untreated PD-L1-positive metastatic non-small-cell lung cancer. Technology appraisal [TA531]2018 22 April 2020. Available from: https://www.nice.org.uk/guidance/ta531.

70. Institute for Clinical Immuno-Oncology. Assessing Immunotherapy Response—Why irRC Matters.2015 27 September 2017. Available from: https://accc-iclio.org/resources/assessing-immunotherapy-response-why-irrc-matters/.
71. Office for National Statistics. Cancer Registration Statistics, England, 20152016 August 2017. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancerregistrationstatisticsengland/2015.
72. Office for National Statistics. National life tables:UK 2019 [Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesunitedkingdomreferencetables.
73. Mislevy RJ. Journal of Educational Statistics. 1991;16(2):150-5.
74. Hardin J. Prais-Winsten regression. Stata Technical Bulletin. 1996;5.
75. Heckman JJ. The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a simple Estimator for Such Models Annalsof Economic and Social Measurement. 1976;5(4):475-92.
76. Galimard J-E, Chevret S, Curis E, Resche-Rigon M. Heckman imputation models for binary or continuous MNAR outcomes and MAR predictors. BMC Medical Research Methodology. 2018;18(1):90.
77. Hatswell AJ, Pennington B, Pericleous L, Rowen D, Lebmeier M, Lee D. Patient-reported utilities in advanced or metastatic melanoma, including analysis of utilities by time to death. Health and quality of life outcomes. 2014;12:140.
78. Harel O, Hofer SM, Hoffman L, Pedersen NL, Johansson B. Population inference with mortality and attrition in longitudinal studies on aging: a two-stage multiple imputation method. Experimental aging research. 2007;33(2):187-203.
79. Cunningham D, Allum WH, Stenning SP, Thompson JN, Van de Velde CJ, Nicolson M, et al. Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer. The New England journal of medicine. 2006;355(1):11-20.
80. Shen ZJ. Nested Multiple Imputation . Phd . Dissertation. Cambridge, MA: Harvard Univeristy; 2000.
81. Grothey A, Van Cutsem E, Sobrero A, Siena S, Falcone A, Ychou M, et al. Regorafenib monotherapy for previously treated metastatic colorectal cancer (CORRECT): an international, multicentre, randomised, placebo-controlled, phase 3 trial. Lancet. 2013;381(9863):303-12.
82. National Institute for Health and Care Excellence. Nivolumab for previously treated locally advanced or metastatic non-squamous non-small-cell lung cancer [ID900].In development [GID-TAG524].2017 1 September 2017 Available from: https://www.nice.org.uk/guidance/indevelopment/gid-tag524.
83. National Institute for Health and Care Excellence. Nivolumab for previously treated locally advanced or metastatic squamous non-small cell lung cancer [ID811]. Single Technology Appraisal: Evidence Review Group Report. In development [GID-TAG506].2015 14 September 2017. Available from: https://www.nice.org.uk/guidance/gid-tag506/documents/committee-papers
84. National Institute for Health and Care Excellence. Technology appraisal guidance [TA462]. Nivolumab for treating relapsed or refractory classical Hodgkin lymphoma. 2017 14 September 2017. Available from: https://www.nice.org.uk/guidance/ta462.

85. National Institute for Health and Care Excellence. Technology appraisal guidance [TA384]. Nivolumab for treating advanced (unresectable or metastatic) melanoma.2016 14 September 2017. Available from: https://www.nice.org.uk/guidance/ta384.
86. National Institute for Health and Care Excellence. Technology appraisal guidance [TA400]. Nivolumab in combination with ipilimumab for treating advanced melanoma. 2016 14 September 2017. Available from: https://www.nice.org.uk/guidance/ta400.
87. National Institute for Health and Care Excellence. Technology appraisal guidance [TA417]. Nivolumab for previously treated advanced renal cell carcinoma. 2016 14 September 2017. Available from: https://www.nice.org.uk/guidance/ta417.
88. National Institute for Health and Care Excellence. Nivolumab for treating metastatic, squamous, non-small-cell lung cancer after chemotherapy [ID811]. In development [GID-TAG506].2017 14 November 2017. Available from: https://www.nice.org.uk/guidance/indevelopment/gid-tag506.
89. NHS Improvement. 2018/19 National Cost Collection data 2020 [Available from: https://improvement.nhs.uk/resources/national-cost-collection/#ncc1819.
90. Sacco JJ, Botten J, Macbeth F, Bagust A, Clark P. The average body surface area of adult cancer patients in the UK: a multicentre retrospective study. PloS one. 2010;5(1):e8933-e.
91. National Institute for Health and Care Excellence. Pembrolizumab for previously treated oesophageal or gastro-oesophageal junction cancer ID1357. In development [GID-TA10322].2018 12 September 2018. Available from: https://www.nice.org.uk/guidance/indevelopment/gid-ta10322.
92. Tey J, Soon YY, Koh WY, Leong CN, Choo BA, Ho F, et al. Palliative radiotherapy for gastric cancer: a systematic review and meta-analysis. Oncotarget. 2017;8(15):25797.
93. White J, Carolan-Rees G. PleurX peritoneal catheter drainage system for vacuum-assisted drainage of treatment-resistant, recurrent malignant ascites. Applied health economics and health policy. 2012;10(5):299-308.
94. Campbell H, Stokes E, Bargo D, Logan R, Mora A, Hodge R, et al. Costs and quality of life associated with acute upper gastrointestinal bleeding in the UK: cohort analysis of patients in a cluster randomised trial. BMJ open. 2015;5(4):e007230.
95. Personal Social Services Research Unit. Unit Costs of Health and Social Care 2016.2016 13 April 2017. Available from: http://www.pssru.ac.uk/project-pages/unit-costs/2016/index.php.
96. Georghiou T, Bardsley M. Exploring the cost of care at the end of life. London: Nuffield Trust Research Report. 2014.
97. Copley-Merriman C, Stevinson K, Liu FX, Wang J, Mauskopf J, Zimovetz EA, et al. Direct costs associated with adverse events of systemic therapies for advanced melanoma: Systematic literature review. Medicine. 2018;97(31):e11736.
98. Personal Social Services Research Unit. Unit Costs of Health and Social Care 2019 2020 [Available from: https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2019/.
99. Schadendorf D, Hodi FS, Robert C, Weber JS, Margolin K, Hamid O, et al. Pooled Analysis of Long-Term Survival Data From Phase II and Phase III Trials of Ipilimumab in Unresectable or Metastatic Melanoma. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2015;33(17):1889-94.

100. Wolchok JD, Chiarion-Sileni V, Gonzalez R, Rutkowski P, Grob J-J, Cowey CL, et al. Overall survival with combined nivolumab and ipilimumab in advanced melanoma. New England Journal of Medicine. 2017;377(14):1345-56.
101. Motzer RJ, Escudier B, McDermott DF, George S, Hammers HJ, Srinivas S, et al. Nivolumab versus everolimus in advanced renal-cell carcinoma. New England Journal of Medicine. 2015;373(19):1803-13.
102. Borghaei H, Brahmer JR, Horn L, Ready N, Steins M, Felip E, et al. Nivolumab (nivo) vs docetaxel (doc) in patients (pts) with advanced NSCLC: CheckMate 017/057 2-y update and exploratory cytokine profile analyses. American Society of Clinical Oncology; 2016.
103. McDermott DF, Drake CG, Sznol M, Choueiri TK, Powderly JD, Smith DC, et al. Survival, durable response, and long-term safety in patients with previously treated advanced renal cell carcinoma receiving nivolumab. Journal of Clinical Oncology. 2015;33(18):2013-20.
104. Gettinger SN, Horn L, Gandhi L, Spigel DR, Antonia SJ, Rizvi NA, et al. Overall survival and long-term safety of nivolumab (anti–programmed death 1 antibody, BMS-936558, ONO-4538) in patients with previously treated advanced non–small-cell lung cancer. Journal of clinical oncology. 2015;33(18):2004-12.
105. McDermott DF, Motzer RJ, Atkins MB, Plimack ER, Sznol M, George S, et al. Long-term overall survival (OS) with nivolumab in previously treated patients with advanced renal cell carcinoma (aRCC) from phase I and II studies. American Society of Clinical Oncology; 2016.
106. Brahmer J, Horn L, Jackman D, Spigel D, Antonia S, Hellmann M, et al. Abstract CT077: Five-year follow-up from the CA209-003 study of nivolumab in previously treated advanced non-small cell lung cancer (NSCLC): Clinical characteristics of long-term survivors. AACR; 2017.

List of appendices

Appendix number	Appendix Title
C	Nivolumab SmPC
D	Identification, selection and synthesis of clinical evidence: systematic literature review report
E	Subgroup analysis
	E.1: ATTRACTION-3 Clinical Study Report
	E.2: ATTRACTION-1 Clinical Study Report
F	Adverse reactions
G	Published cost-effectiveness studies: systematic literature review
H	Health-related quality-of-life studies: systematic literature review
I	Cost and healthcare resource identification:
J	Clinical outcomes and disaggregated results from the model
K	Checklist of confidential information
L	Indirect treatment comparison report
M	Survival analysis report
N	Utility analysis report

Appendix C: Summary of product characteristics (SmPC) and European public assessment report (EPAR)

SmPC

This document is provided as a separate document, labelled Appendix C.

EPAR

No EPAR is currently available describing nivolumab for the treatment of unresectable, advanced oesophageal cancer when standard chemotherapy has failed; the latest EPAR for nivolumab is available on the EMA website.

Appendix D: Identification, selection and synthesis of clinical evidence

The clinical systematic literature review is provided as separate document.

Appendix E: Subgroup analysis

Relevant subgroup analysis results for ATTRACTION-3 and ATTRACTION-1 are presented in Section 0 and Section B.2.6.2.10, respectively, and further described in in the CSRs, labelled as Appendix E.1 for ATTRACTION-3 (database lock on 30 November 2019) and Appendix E.2 for ATTRACTION-1 (database lock 17 November 2016).

Appendix F: Adverse reactions

All relevant information has been provided in Section B.2.10.

Appendix G: Published cost-effectiveness studies

This systematic literature review is provided as separate document.

Appendix H: Health-related quality-of-life studies

This systematic literature review is provided as separate document.

Appendix I: Cost and healthcare resource identification, measurement and valuation

This systematic literature review is provided as separate document.

Appendix J: Clinical outcomes and disaggregated results from the model

J1.1 Clinical outcomes from the model

	Nivolumab	Taxane	
		Absolute	Incremental
Median ToT (years)	0.230	0.211	0.019
Mean ToT (years)	0.496	0.291	0.205
Median PFS (years)	0.153	0.287	-0.134
Mean PFS (years)	0.487	0.408	0.080
Median OS (years)	0.901	0.747	0.153
Mean OS (years)	1.650	0.997	0.653
Time in pre-progression (years)	[redacted]	[redacted]	0.080
- Time initial therapy (years)	[redacted]	[redacted]	0.036
- Time in subsequent therapy (years)	[redacted]	[redacted]	0.043
Time in post-progression (years)	[redacted]	[redacted]	0.573
OS: overall survival; PFS: progression-free survival; ToT: time on treatment			

J1.2 Disaggregated results of the base-case incremental cost-effectiveness analysis

	Nivolumab	Taxanes		
		Absolute	Incremental	% Absolute Increment
Costs (with PAS)				
HS costs	[redacted]	[redacted]	£3,106	14.9%
Treatment costs	[redacted]	[redacted]	£17,994	86.3%
BSC costs	[redacted]	[redacted]	£741	3.6%
Average AE costs per patient	[redacted]	[redacted]	-£998	-4.8%
Total costs	[redacted]	[redacted]	£20,842	100%
Health benefits				
Total QALYs	[redacted]	[redacted]	0.458	100%
Total life years	[redacted]	[redacted]	0.536	100%
AE: adverse event; BSC: best supportive care; HS: health state; PAS: patient access scheme; QALY: quality-adjusted life year;				

Appendix K: Checklist of confidential information

This document is provided as a separate document.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal

Nivolumab for unresectable, advanced oesophageal cancer when standard chemotherapy has failed [ID1249]

Clarification questions

May 2020

File name	Version	Contains confidential information	Date
[ID1249] Nivolumab ERG clarification questions for company [CIC].docx	V1.	Yes	11 May 2020

Section A: Clarification on effectiveness data

Literature searches

A1. Were searches of trial registers (e.g. ClinicalTrials.gov, WHO ICTRP) completed?

As indicated in Appendix D under the methodology section, trial registries including ClinicalTrials.gov were not searched for primary studies. However, all relevant studies assessing nivolumab in this indication were identified. Further, searches conducted in the different electronic databases (PubMed, Embase, Cochrane CENTRAL) and review of HTAs were considered comprehensive enough to identify all relevant publications for comparators.

A2. Please confirm the platform used to search Embase? (Ovid, Embase.com?)

Embase was searched via Embase.com for the clinical, CEM, utilities and cost SLR.

A3. Were adverse event searches completed for nivolumab or comparators?

Adverse events were assessed as part of the clinical effectiveness SLR described in Appendix D of the company submission.

A4. On page 9, appendix D, it states that "A free text internet search was also conducted to identify any further studies…" .Please provide further information about this search.

The free text internet search was conducted via google scholar using simple disease-specific search terms.

A5. PRISMA flow diagram. For the 2018 update search (page 12, Appendix D), the text describing the flow of studies suggests 112 records were excluded at the full-text. The PRISMA diagram (page 13, Appendix D) indicates 122 records were excluded at this stage. Please confirm the correct number of excluded records.

The PRISMA flow diagram states that 100 records were excluded based on the population, 7 records were included based on the comparison, 5 records were excluded

based on the outcomes, 2 records were excluded based on study type and 8 records were excluded based on other reasons. Thus, 122 records is the correct number excluded at the second pass stage for the 2018 SLR update.

A6. PRISMA flow diagram. The PRISMA diagram (page 13, Appendix D) shows the total number of unique studies included in the qualitative synthesis as 54. However, there were 36 unique included studies from the 2017 search, plus 7 from 2018, and 14 from 2020. This would total 57 unique included studies in the qualitative synthesis. Please clarify.

The PRISMA flow diagram indicates how many unique studies were identified at the time of the conducted search for each update respectively. As stated above, there were 36 unique included studies from the 2017 search, plus 7 from 2018, and 14 from 2020. However, three of these studies were unique at the time of search, but were not unique across the overall SLR. Hence, 54 unique studies were included within qualitative synthesis.

A7. PRISMA flow diagram. The PRISMA diagram (page 13, Appendix D) shows 741 excluded references at the full text stage from the 3 SLR searches (540 from 2017; 122 from 2018; 79 from 2020). The linked spreadsheet on page 37 'list of excluded studies at second pass' only contains 718 records. Please confirm the number of excluded records at this stage.

Please find attached an updated list of excluded studies. A total of 741 studies were excluded at the second pass stage of the clinical SLR.

Trial data and design

A8. Please clarify if the decision to compare paclitaxel with docetaxel was undertaken pre-randomisation. If so, please provide stratified estimates for OS, PFS and ORR by 'paclitaxel-eligible' and 'docetaxel-eligible' subgroups.

Investigator choice between paclitaxel and docetaxel was declared and documented in the randomization system (IWRS) prior to randomization. However, the investigator's

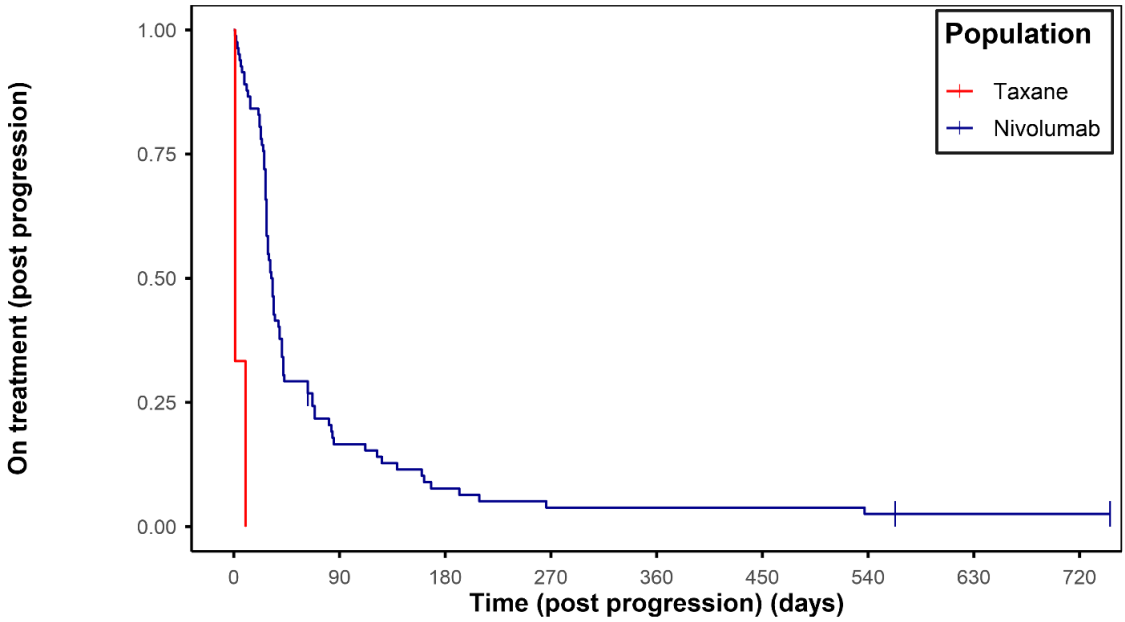
choice of taxane was not a stratification factor and so did not influence randomisation. The trial was not designed to compare paclitaxel with docetaxel.

A9. PRIORITY QUESTION: Please clarify what proportion of patients received post-progression treatment in ATTRACTION-3 by arm.

In the nivolumab arm, 82 of the 210 patients (39.0%) received treatment post-progression, with a median of 3 treatments (range: 1-52 treatments) and a median post-discontinuation time on treatment of 32.5 days (95% CI: 28-39 days).[1]

In the taxane arm, 3 of the 209 patients (1.4%) received treatment post-progression; all patients had one subsequent treatment and a median post-discontinuation time on treatment of 1 day.[1]

This pattern of post-progression treatment is in line with the known mechanism of action of nivolumab, so that patients may continue to receive treatment benefit following progression. This is reflected in the economic model, where patient time on treatment is directly modelled, based on ATTRACTION-3 time on treatment.



NAR (Cumulative Events)

Taxane	3 (0)	0 (3)	0 (3)	0 (3)	0 (3)	0 (3)	0 (3)	0 (3)	0 (3)
Nivolumab	82 (0)	13 (68)	6 (75)	3 (78)	3 (78)	3 (78)	2 (79)	1 (79)	1 (79)

Figure 1. ATTRACTION-3 post-progression time on treatment[1]

A10. Please clarify why the trial null hypothesis was a one-sided test of superiority as the statistical methods draw on two-sided tests.

Although the null hypothesis was a one-sided test of superiority, a two-sided test is more stringent. Hence, all significant and superior results under a two-sided test will remain significantly superior under a one-sided test[2].

A11. Please clarify what proportion of patients in ATTRACTION-1 would not have been included in ATTRACTION-3, and due to which inclusion and exclusion criteria.

ATTRACTION-1 included patients that were refractory or intolerant to fluoropyrimidine-based, platinum-based and taxane-based chemotherapy, whereas ATTRACTION-3 included patients refractory or intolerant to fluoropyrimidine-based and platinum-based chemotherapy and excluded patients refractory or intolerant to taxane-based

chemotherapy[1, 2]. This was necessary in order to enable recruitment of ATTRACTION-3 patients to a relevant comparator arm (i.e. taxane monotherapy).

Given this difference in eligibility criteria, patients enrolled in ATTRACTION-1 would not have been eligible for ATTRACTION-3. However, this gives confidence that the beneficial impact of nivolumab is observed in patients currently receiving taxane therapy in the UK and those currently receiving best supportive care. Hence, both studies are directly relevant to the UK setting.

A12. PRIORITY QUESTION: Please clarify what proportion of patients in ATTRACTION-3 had recurrent cancer.

As indicated in Table 23 (page 62) in the company submission, slightly more patients in the nivolumab arm (49.0%) had recurrent cancer versus the taxane arm (42.6%; 47.6% of patients receiving paclitaxel, 40.2% receiving docetaxel).[1]The number of patients in ATTRACTION-3 is further summarised in the table below.

Table 1. ATTRACTION-3: Proportion of patients with recurrent cancer[1]

	Nivolumab	Control group		
		Total	Docetaxel	Paclitaxel
Recurrent, n (%)				
No	107 (51.0)	120 (57.4)	34 (52.3)	86 (59.7)
Yes	103 (49.0)	89 (42.6)	31 (47.7)	58 (40.3)

A13. A higher proportion of patients died within the first 2.5 months in the nivolumab arm (32/210, 15.2%) as compared to the chemotherapy arm (15/209, 7.2%). Please can you provide an explanation for this finding, e.g. do you think it may be related to the mechanism of action of nivolumab?

As stated in Section 1.3.5.1 of the company submission, conventional anti-cancer therapies typically aim to reduce the tumour burden through direct disruption of tumour cell proliferation or induction of apoptosis. In contrast, there are key differences with immunotherapy agents such as nivolumab, as a result of their novel mechanism of

action. One of these differences is the varying patterns of response that can be observed with immunotherapy agents, compared with chemotherapy.

- Due to the indirect anti-tumour mechanism associated with immunotherapies, where host immune cells are recruited to the tumour site, the initial effect of immunotherapy may present as growth of existing lesions or formation of new lesions that result from the infiltration of tumour-specific immune cells and other inflammatory cells ("pseudo-progression").[3-5] This brief initial enlargement of the tumour may be followed by tumour shrinkage or eradication.[3, 4]

- Due to the delayed clinical responses observed in immunotherapies, the "time to response" from immunotherapy treatment may differ from that seen after conventional chemotherapy.[5]

- In addition, these differences in response patterns after immunotherapy may potentially be prematurely misclassified as disease progression under the WHO or RECIST criteria.[4, 5] For the same reasons, PFS may not be an adequate endpoint in immunotherapy trials and may not be considered a surrogate for OS for the achievement of clinical efficacy.

For this reason, in the oesophageal cancer setting, where there is short life expectancy and poor prognosis, Kaplan-Meier curves for patients receiving nivolumab monotherapy often demonstrate a high initial hazard, followed by decreasing hazard over time. By contrast, Kaplan-Meier data describing patients receiving conventional chemotherapies have a lower initial hazard followed by increasing hazard over time. This is reflected in the survival profiles applied in the economic model.

This pattern of response is observed in the ATTRACTION-3 study, where OS at two months is lower in the nivolumab arm, comparable at approximately four months and substantially improved by six months ([redacted]).[1]This improvement is maintained at all subsequent timepoints.

A14. Patients with extreme malnutrition were excluded from participating in the ATTRACTION-3 study. However, given that malnutrition impacts on patient fitness and so may be particularly relevant for this condition, please can you provide a summary of baseline patient data for BMI and body mass for patients in both arms of the ATTRACTION-3 study?

BMI and weight at baseline in ATTRACTION-3 are summarised for patients in the nivolumab arm and the control arm in the table below. As can be seen, patients in the nivolumab has slightly lower weight and BMI than those in the taxane arms, so that any bias would favour the control arm.[1]

Table 2. ATTRACTION-3: Weight and BMI[1]

	Nivolumab	Control group		
		Total	Docetaxel	Paclitaxel
Weight (kg)				
Mean (SD)	[redacted]	[redacted]	[redacted]	[redacted]
Median	[redacted]	[redacted]	[redacted]	[redacted]
Min-Max	[redacted]	[redacted]	[redacted]	[redacted]
BMI (kg/m^2)				
Mean (SD)	[redacted]	[redacted]	[redacted]	[redacted]
Median	[redacted]	[redacted]	[redacted]	[redacted]
Min-Max	[redacted]	[redacted]	[redacted]	[redacted]

A15. Please provide estimates of PFS assessed by iRECIST criteria, to address pseudo-progression in the ascertainment of trial outcomes.

PFS was not assessed by iRECIST criteria in ATTRACTION-3.[1]

A16. PRIORITY QUESTION: The OS and PFS curves for nivolumab in ATTRACTION-3 appears to have several 'bumps' that may be protocol driven, but this is not mirrored in the curves for taxanes. Please could the company clarify if there is an aspect of trial design that might account for this?

There are no differences in ATTRACTION-3 trial design between the nivolumab and taxane arms. PFS data are subject to initial "bumps" (several events or an absence of events during a short period) in both arms, due to the timing of assessments in ATTRACTION-3 (every six weeks from the start of cycle one until one year; subsequently, every 12 weeks from start of cycle 1). While ATTRACTION-3

investigators may determine that a patient has progressed at any date, progression is more likely to be diagnosed at an evaluation point, particularly in the early stages of the study. This is notable in both treatment arms, as outlined in Figure 2, so it would be incorrect and inappropriate to say that these PFS “bumps” are experienced in the nivolumab arm only.

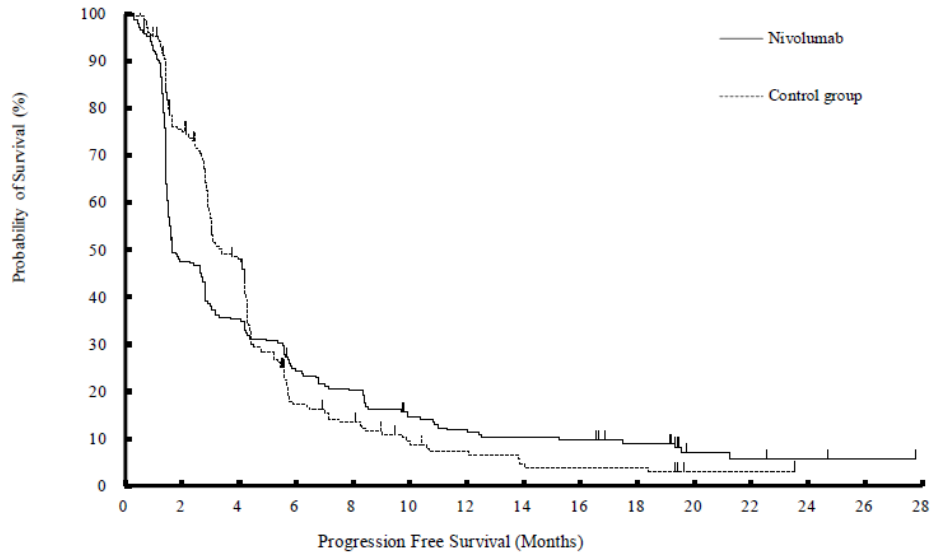


Figure 2. ATTRACTION-3: Kaplan-Meier plot of progression-free survival in patients receiving nivolumab or taxanes[1]

There are no aspects of trial design that may cause similar “bumps” for OS in the nivolumab arm only. As outlined above, all aspects of trial design are equivalent between the two treatment arms. Further, the baseline characteristics for the two arms are broadly comparable, as outlined in Section B.2.6.1.8 of Document B. It should also be noted that OS “bumps” are observed in both trial arms, although timing differs slightly.

In common with most randomised clinical trials for cancer indications, patients in ATTRACTION-3 had a predicted life expectancy that was greater than three months.[6] This may have resulted in a lower initial hazard observed in the Kaplan-Meier data, but would impact on both trial arms and would affect overall hazard, as opposed to creating “bumps”. Additionally, this would be reflected in baseline characteristics for patients, which are broadly comparable.

It is possible that post-discontinuation treatment may have impacted on OS curves. However, time on treatment was broadly comparable between arms in the initial months (as outlined in Figure 46 of Appendix M in the company submission), so that equivalent numbers of patients would be impacted across arms. In addition, the proportion of patients receiving subsequent treatment was similar across arms ([redacted]).

Further, the timing of these “bumps” (several events or an absence of events during a short period) is aligned to the mechanism of action of nivolumab. There is a high initial hazard from treatment initiation to around 2-2.5 months, which is aligned to the response profile for patients receiving immunotherapies, such as nivolumab (the median time to response was [redacted] months for patients receiving nivolumab during ATTRACTION-3). Hence, observing a “bump” at around this point can be expected, as the impact of nivolumab is reflected in the natural history of OSCC, which has extremely poor prognosis and short survival. Further, the time to response has a wide range in patients receiving nivolumab: although median time to response is 2.6 months, the mean time to response is [redacted] months (standard deviation: [redacted] months), with patients achieving up to [redacted] months. Hence, small fluctuations in the hazard profile can be anticipated and are observed. However, these fluctuations do not change the conclusions from ATTRACTION-3: nivolumab treatment significantly improves survival outcomes in OSCC, where patients would otherwise have short survival.

[redacted]

Figure 3. ATTRACTION-3: Kaplan-Meier plot of overall survival in patients receiving nivolumab (ONO-4524) or docetaxel/paclitaxel[1]

A17. Please provide four-way subgroup analyses for OS, PFS and ORR crossed by Japan vs rest of world and PD-L1 expression.

These analyses are not available.

Systematic review methods and indirect treatment comparison

A18. Please clarify which studies contributed to reconstructed individual participant data in the indirect treatment comparison.

Kaplan Meier (KM) data was reconstructed where available, only to validate the findings reported. These were not used as inputs to the NMA but were used to assess proportional hazards and allowed for validation of the published findings. Only the reported data was used as inputs to the NMA to avoid any discrepancies or additional uncertainty.

A19. Please clarify the basis on which post-progression survival was judged to be exchangeable with overall survival in the indirect treatment comparison.

Post-progression survival was not judged to be comparable with overall survival (OS) across all studies. For one study, (Moriwaki et al.[7]), post-progression survival was judged to be equivalent to OS due to the definition and measurement of this endpoint. Moriwaki et al.[7] was a retrospective study of patients treated with docetaxel or best supportive care (BSC). The study authors noted that while the time to death from initiation with docetaxel was measurable, it was not possible to quantify the time of initiation with BSC. Hence, OS was not available as it is typically defined within clinical studies (time from treatment initiation to date of death). Therefore, BSC start date was defined as the date of disease progression on platinum-based chemotherapy to death from any cause or to the last follow up (censored). This is comparable to the definition of OS applied in other studies and thus it was included within the current ITC as if it were a measure of OS.

While there are limitations associated with this assumption, this was the only study that was available to inform any estimate of the relative efficacy of BSC. As there were no

measures of post-progression survival for other comparators, this represented the only option to include BSC into the network.

A20. Please provide analysis files and summary effect estimates used in the indirect treatment comparison, suitable for replication in WinBUGS.

These files are provided with this response. It should be noted that all analyses have been conducted in line with TSD2 and use the recommended WinBUGS template provided in example 7 (a and b).

A21. PRIORITY: The report of the indirect treatment comparison makes reference to a 'base case' analysis. Please clarify any sensitivity analyses (i.e. 'non-base-case') undertaken and what the results of these analyses were.

The reference to the "base case" analysis refers to the mean values analyses in the economic model that include BSC as a comparator. Sensitivity analysis was not performed for the NMA due to the limited size of the network and sparsity of studies. Indeed, removing any studies for sensitivity analysis would result in a very small network, further adding uncertainty and reducing the usefulness of this analysis. As such, it was deemed inappropriate to exclude any studies to determine their influence. No other sensitivity analysis was considered.

A22. Please provide a model-generated indirect estimate of effectiveness for paclitaxel compared with best supportive care.

The indirect estimate of effectiveness for paclitaxel is shown in Table 3.

Please note that the company do not consider it appropriate to use a model generated indirect estimate of effectiveness for paclitaxel in this context. There is no evidence available to validate the assumption of proportional hazards between paclitaxel and BSC. Indeed, the comparison between paclitaxel and BSC is made via docetaxel; therefore using this result to compare to the paclitaxel arm in ATTRACTION-3 is not necessarily appropriate. As evidence, there are not proportional hazards between the combined taxane arm and BSC, as evidenced by Figure 8 from the response to question B8, which compared the combined taxane arm from ATTRACTION-3 with the

Moriwaki study BSC arm.[1, 7] Importantly, Figure 9 does show proportional hazards for Moriwaki study BSC arm versus ATTRACTION-3 docetaxel arm, indicating that it is inappropriate to apply any generated HR to the paclitaxel arm.

Table 3: Estimates of comparative efficacy from an indirect treatment comparison as log hazard ratios

	Fixed Effect				Random Effect			
	Mean	**SD**	**Median**	**95% CI**	**Mean**	**SD**	**Median**	**95% CI**
Docetaxel vs BSC log HR	0.4772	0.202	0.4771	0.0806, 0.8729	0.4798	1.226	0.4784	-2.08, 3.029
Docetaxel vs Paclitaxel log HR	-0.1165	0.129	-0.1162	-0.3695, 0.1366	-0.2189	0.7189	-0.193	-1.755, 1.239
BSC vs Paclitaxel log HR	-0.5937	0.24	-0.594	-1.065, -0.1245	-0.6987	1.42	-0.6721	-3.733, 2.197
σ	-	-	-	-	0.814	0.9005	0.4888	0.030, 3.6
Residual Deviance	5.477	1.992	4.868	3.539, 10.77	3.932	2.636	3.43	0.499, 10.5
pD	1.990	-	-	-	3.627	-	-	-
DIC	3.320	-	-	-	3.422	-	-	-

Section B: Clarification on cost-effectiveness data

Literature searches

B1. In the company's SLR of previous cost-effectiveness studies, one relevant study by Janmaat *et al.*, (2016) was identified. Please can the company explain how elements of this study factored into the development of the *de novo* model produced to inform its submission?

The SLR aimed to identify modelling types previously applied in this indication (or similar) and to assess their suitability for modelling unresectable, advanced oesophageal cancer when standard chemotherapy has failed.

In common with many cancer models, Janmaat et al. (2016)[8] performed a cost-utility analysis using a linear model, based around mean PFS and mean OS. The model accounted for quality of life for OC patients, treatment discontinuation in terms of costs and also medical resource use. These approaches were all applied within the current de novo model.

Of note, utility values within this study were derived from patients with Barrett's oesophagus, which may not be appropriate to comparisons of patients with oesophageal squamous cell carcinoma, due to the differences in aetiology.

B2. In the company's SLR of health effects, six studies were identified for inclusion within qualitative synthesis. Acknowledging that EQ-5D data are available from ATTRACTION-3, please can the company comment on the suitability of these studies to inform the economic model?

An overview of the suitability of quality of life evidence to inform economic modelling is provided in Table 4.

Table 4. Suitability of quality of life evidence to inform economic modelling

Study author (year, country)	Suitability of evidence to inform economic modelling
Bascoul-Mollevi (2017, France)[9]	Included non-metastatic patients only; did not report EQ-5D
Doherty (2018, Canada)[10]	Does not report values by progression status; however, does report EQ-5D values for patients using palliative chemotherapy (0.74)
Dutton (2014, UK)[11]	Does not report EQ-5D values; only reports outcomes at baseline and four weeks, so limited suitability for modelling long-term outcomes
Shenfine (2009, UK)[12]	Does not report EQ-5D values; only reports outcomes at baseline, week one and week six, so limited suitability for modelling long-term outcomes
Tian (2016, China)[13]	Only reports Ogilvie's dysphagia score
Xinopoulos (2005, Greece)[14]	Insufficient information reported to enable derivation of utility value

Doherty (2018)[10] did not report values by progression status, limiting the extent to which it could be applied in the economic model. However, these data were used for the validation of the utility estimates based on EQ-5D data available from ATTRACTION-3. Additionally, these data has been used to inform analyses detailed below.

B3. Please can the company confirm how the outputs from the SLR of costs and healthcare resource identification, measurement and valuation were used to inform the economic model?

Although several of these studies report costs and/or resource use associated with elements of standard of care for patients with advanced OC, none reported on the composition of standard of care. For this reason, these outputs did not inform modelled analyses.

B4. PRISMA flow diagram. For the 2020 update search for the cost-effectiveness SLR, the text (page 8, Appendix G) and PRISMA diagram (page 9, Appendix G) report that the bibliographic database searches identified 165 records. However, the search strategies (page 2-5, Appendix G) indicate 190 records were identified (Medline: 42; Embase: 101; Cochrane: 27; NHS EED: 20; EconLit: 0). Please confirm the correct number of records identified in the 2020 update search of bibliographic databases.

The search strategy identified 190 records, which resulted in 165 records when duplicates were removed. An updated PRISMA flow chart has been attached.

Model structure

B5. PRIORITY QUESTION: The submitted economic model does not include functionality to select alternative fully-parametric or semi-parametric models. Without this functionality, the ERG cannot re-produce the results of the company's sensitivity analyses concerning alternative survival models, thus impeding the ERG's ability to fully critique the CS.

Please can the company update the economic model to provide functionality in order for the ERG to review <u>all</u> alternative specifications of survival models (for the outcomes of OS, PFS, and ToT) provided within the CS? More specifically, please can the company load in the relevant "profiles" for each of the curve options discussed within the CS, including the variance-covariance matrices?

In addition, please can the company check/confirm that <u>all</u> other scenario analyses presented within the CS can also be reproduced within the economic model file (including scenario analyses where the comparator is set to BSC, docetaxel, or paclitaxel)?

The ERG requests that the response to this question be provided as a matter of urgency, as the submitted model should have had the functionality to reproduce any reported results at the time of submission.

Economic models have been provided for each cost-effectiveness scenario within the original submission, as well as those provided within this response. Additionally, each model includes instructions outlining how to adapt the base case to reflect the depicted analysis. Additional support can be provided if necessary.

B6. As an alternative approach to estimate outcomes for docetaxel and paclitaxel separately, please can the company provide sensitivity analyses for the relevant survival models wherein taxanes use is included as a covariate (as opposed to the approach previously taken wherein the comparator arm was separated into

two subgroups)? Please ensure the ability to reproduce the results of this sensitivity analysis is incorporated within the economic model file

Models of OS, PFS and ToT were regressed upon the combined ITT taxane arm of ATTRACTION-3.

For OS and PFS, non-parametric (Kaplan-Meier) estimates of the survival function were derived independently per assigned taxane (docetaxel or paclitaxel) and used to inform survival up to a cut point of 2.99 months. For modelling after this time, the data was subset conditional upon observation beyond this time and both taxane arms were combined. Parametric survival models were then fitted upon the time to event data beyond this time, with a covariate applied to the default scaling parameter of the distribution conditional upon the assigned taxane (reference level paclitaxel). As in the base case analysis, the time variable was reset such that *time=0* when study time was 2.99 months. These models are demonstrated in Figure 4 to Figure 7.

For ToT, fully parametric survival models were fitted upon the time to event data conditional upon the assigned taxane (reference level paclitaxel) from the study index day. These models are demonstrated in Figure 8 and Figure 9.

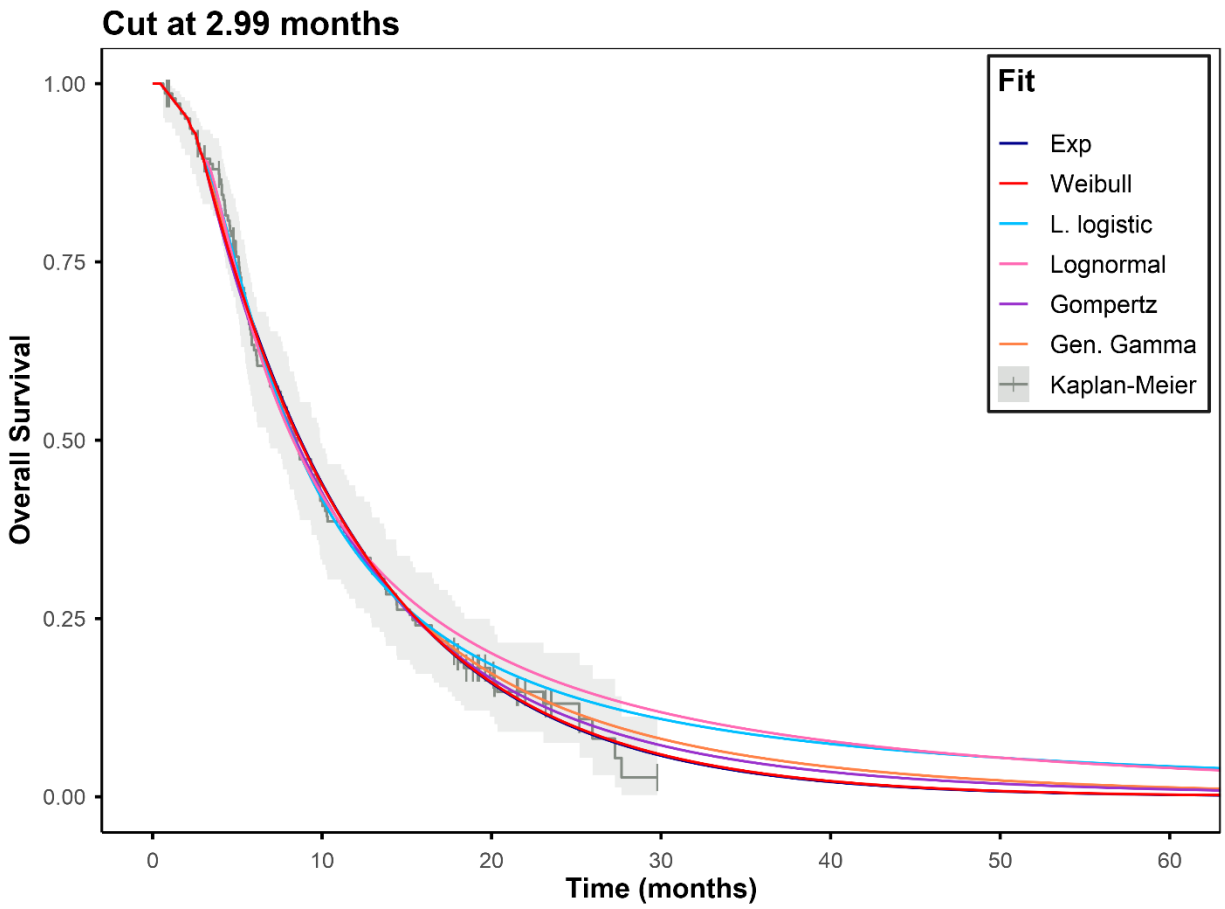
For OS, the exponential model was discarded as fitting the single rate parameter for each arm results in no difference to independent models, and so its use would not be in the spirit of the question. The log-logistic and lognormal fits were deemed poor fits in extrapolation for paclitaxel, whilst the gompertz failed to give a finite mean, implying cure for some patients. As a model that was conservatively positive for both taxanes with a reasonable goodness of fit, the generalised gamma model for OS after 2.99 months was chosen.

For PFS, as for OS, the exponential model was discarded as it forms two independent models and provides no new information. The gompertz fit, whilst having good fit statistics, was obviously incompatible with the OS model, and so was discarded. The conservatively positive log-logistic model was chosen by compromise due to its good fit statistics and the low proportion of the distribution that would be affected by interference

with the OS model. In practice, the PFS curve intercepted the OS curve within the economic model at 322 weeks for docetaxel and 292 weeks for paclitaxel.

For ToT, the differences between the models were minor and so the distinction between independent and conditional models was felt to be of less importance. Therefore, under the principal of parsimony, the exponential model was chosen due to its superlative fit statistics.

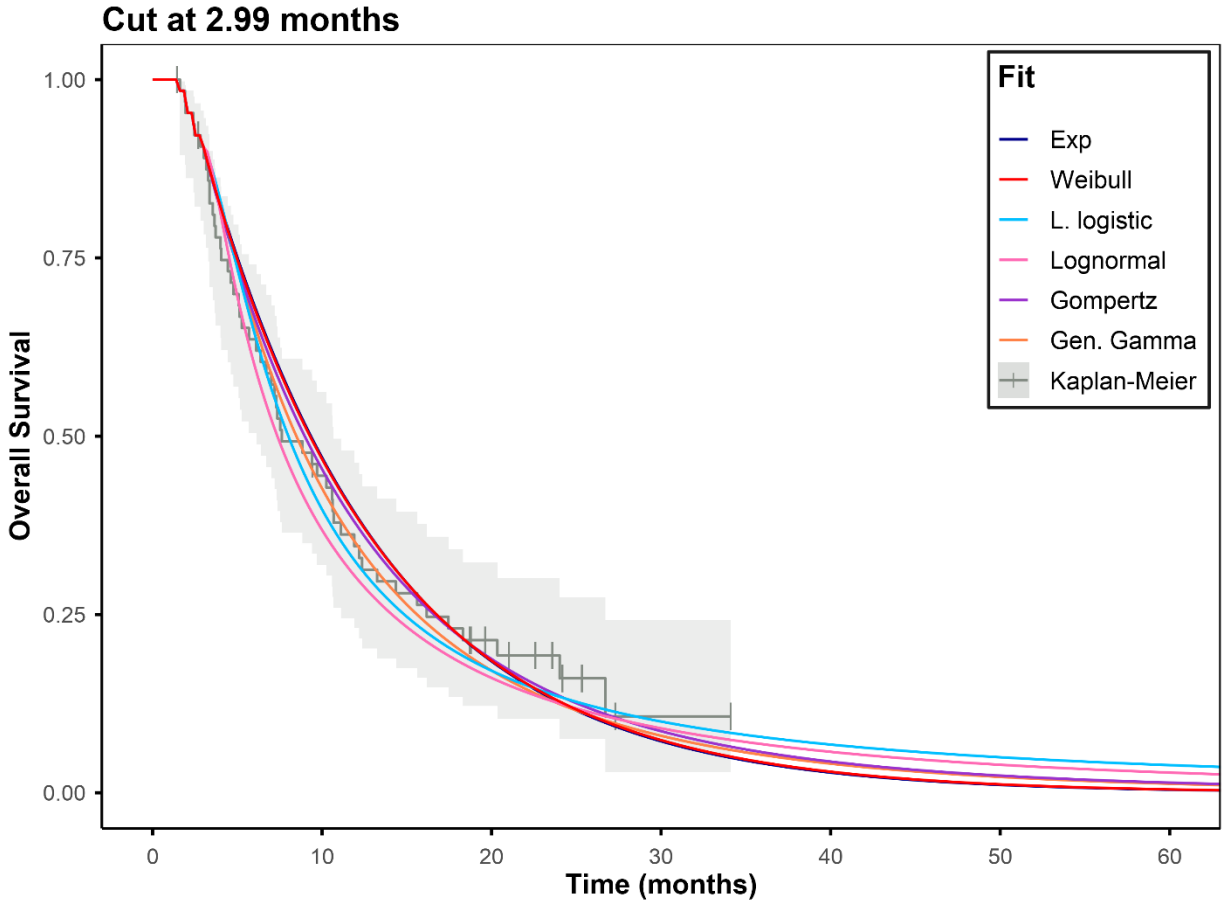
The company notes that the sum of the AICs/BICs for the independent models used in the submission are lower than that of the conditional models, indicating that the additional parameter(s) estimated in independent fitting are likely to provide a benefit in prediction, and that the nested conditional models requested are inferior.



Fit Name	AIC	BIC	Parameters	Median (Months)	Mean (Months)
Exp.	1010.81	1017.22	rate: 0.1015	8.72	11.69
Weibull	1012.77	1022.40	shape: 0.9880; scale: 9.8327	8.67	11.72
L. logistic	1011.54	1021.17	shape: 1.3637; scale: 6.3515	8.33	20.49
Lognormal	1014.70	1024.33	meanlog: 1.8592; sdlog: 1.2897	8.29	16.07
Gompertz	1011.99	1021.62	shape: -0.0131; rate: 0.1107	8.43	N/A
Gen. Gamma	1011.74	1024.58	mu: 2.0885; sigma: 1.1465; Q: 0.5400	8.44	12.65

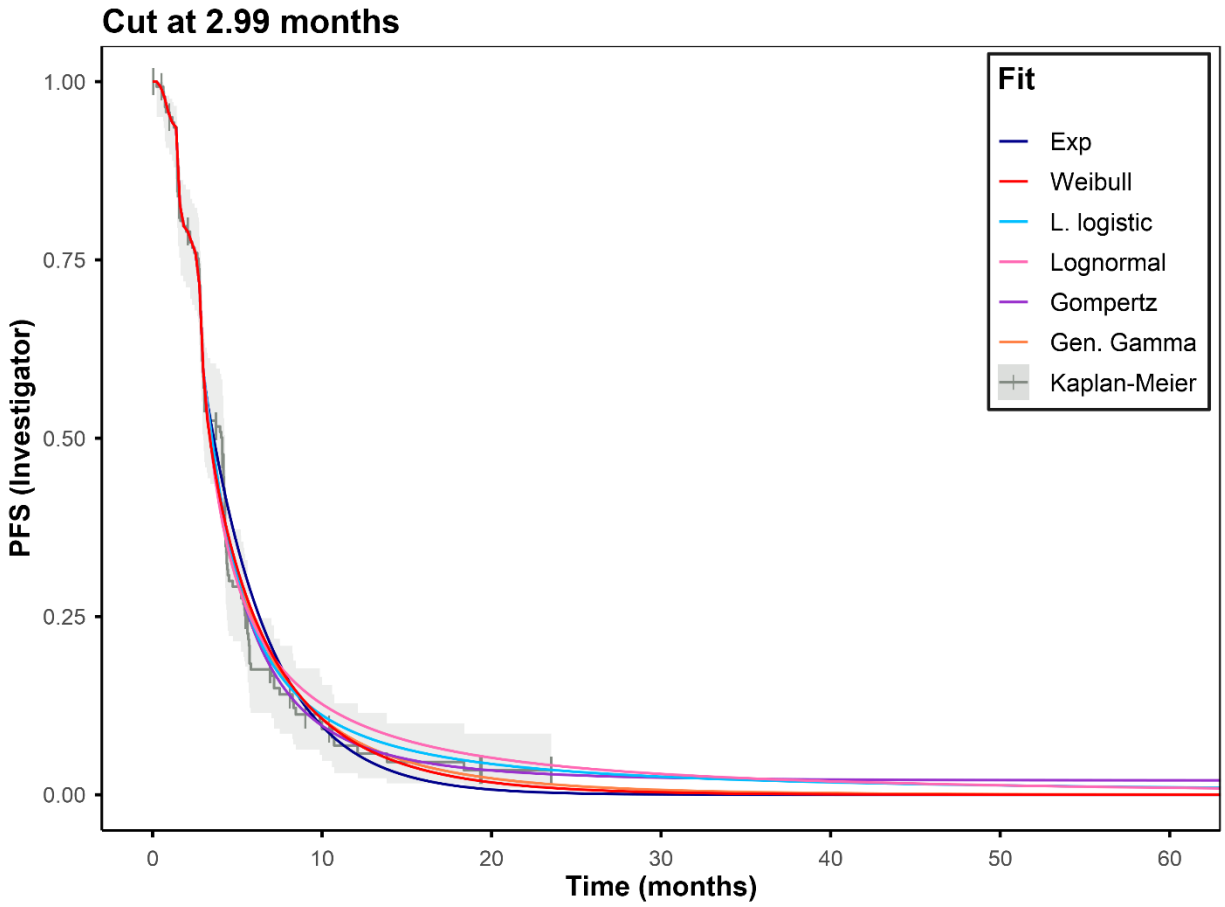
Figure 4: Semi-parametric (Gelber) models of Overall Survival conditional upon taxane assigned, evaluated for paclitaxel. NB: AIC/BIC apply for post 2.99 month period only,

and represent the AIC/BIC for both taxanes in the regressed data.



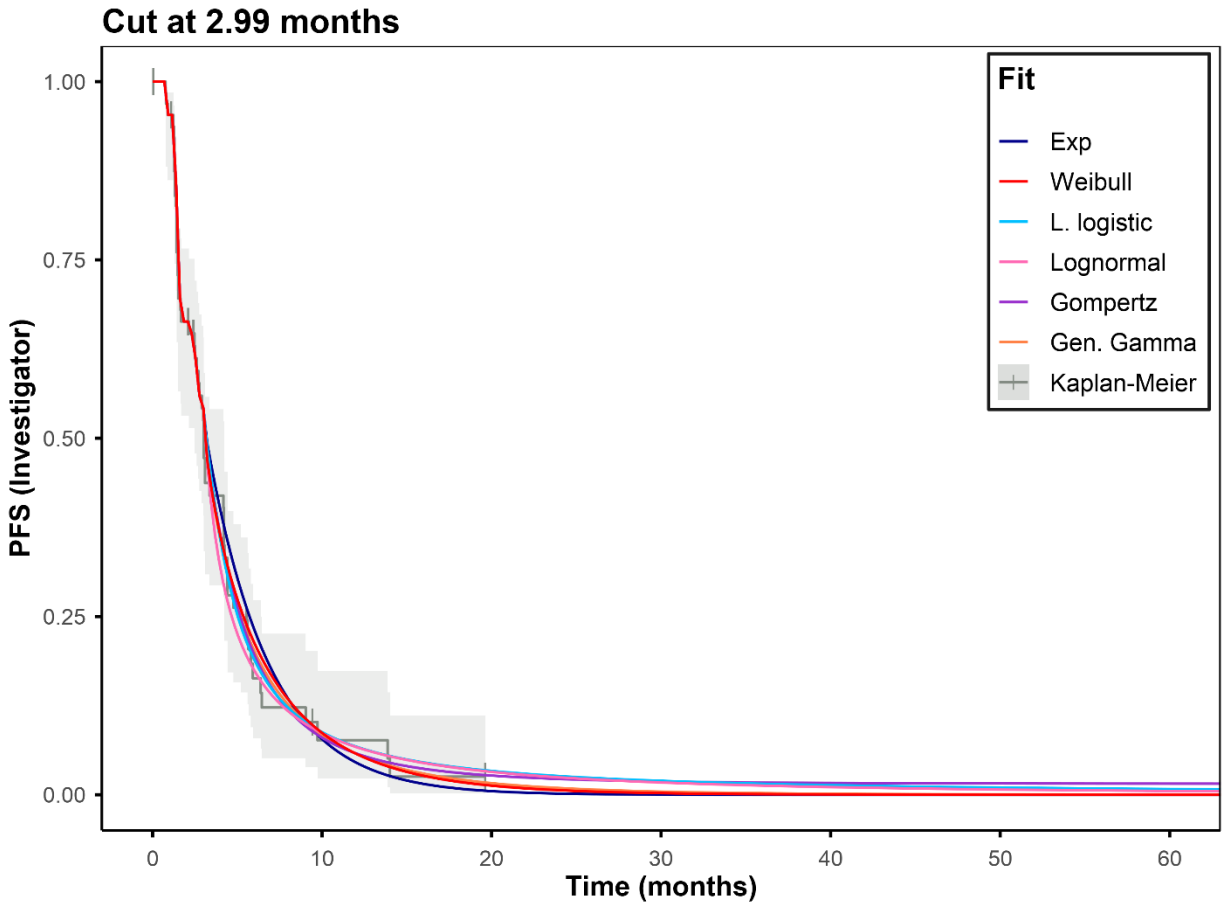
Fit Name	AIC	BIC	Parameters	Median (Months)	Mean (Months)
Exp.	1010.81	1017.22	rate: 0.0937	9.34	12.59
Weibull	1012.77	1022.40	shape: 0.9880; scale: 10.6533	9.28	12.62
L. logistic	1011.54	1021.17	shape: 1.3637; scale: 5.8421	8.00	19.32
Lognormal	1014.70	1024.33	meanlog: 1.6419; sdlog: 1.2897	7.36	13.66
Gompertz	1011.99	1021.62	shape: -0.0131; rate: 0.1033	8.97	N/A
Gen. Gamma	1011.74	1024.58	mu: 2.0730; sigma: 1.1465; Q: 0.5400	8.48	12.66

Figure 5: Semi-parametric (Gelber) models of Overall Survival conditional upon taxane assigned, evaluated for docetaxel. NB: AIC/BIC apply for post 2.99 month period only, and represent the AIC/BIC for both taxanes in the regressed data.



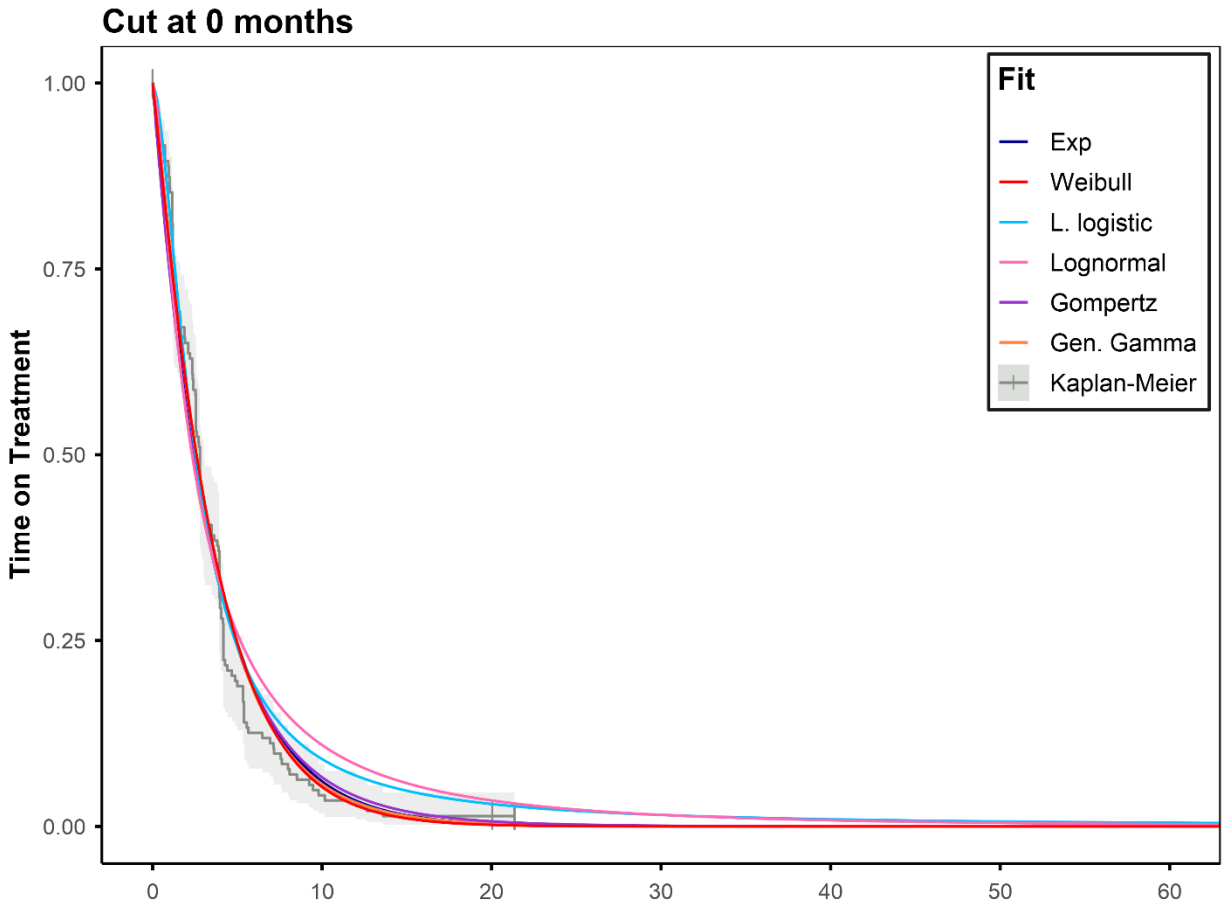
Fit Name	AIC	BIC	Parameters	Median (Months)	Mean (Months)
Exp.	429.90	435.24	rate: 0.2578	3.60	4.86
Weibull	425.06	433.08	shape: 0.8165; scale: 3.6451	3.37	4.97
L. logistic	423.17	431.18	shape: 1.2201; scale: 2.1409	3.49	8.60
Lognormal	431.50	439.51	meanlog: 0.7253; sdlog: 1.5623	3.39	6.69
Gompertz	419.88	427.90	shape: -0.1091; rate: 0.3675	3.43	N/A
Gen. Gamma	425.59	436.28	mu: 1.1264; sigma: 1.3068; Q: 0.7032	3.40	5.09

Figure 6: Semi-parametric (Gelber) models of Progression-Free Survival conditional upon taxane assigned, evaluated for paclitaxel. NB: AIC/BIC apply for post 2.99 month period only, and represent the AIC/BIC for both taxanes in the regressed data.



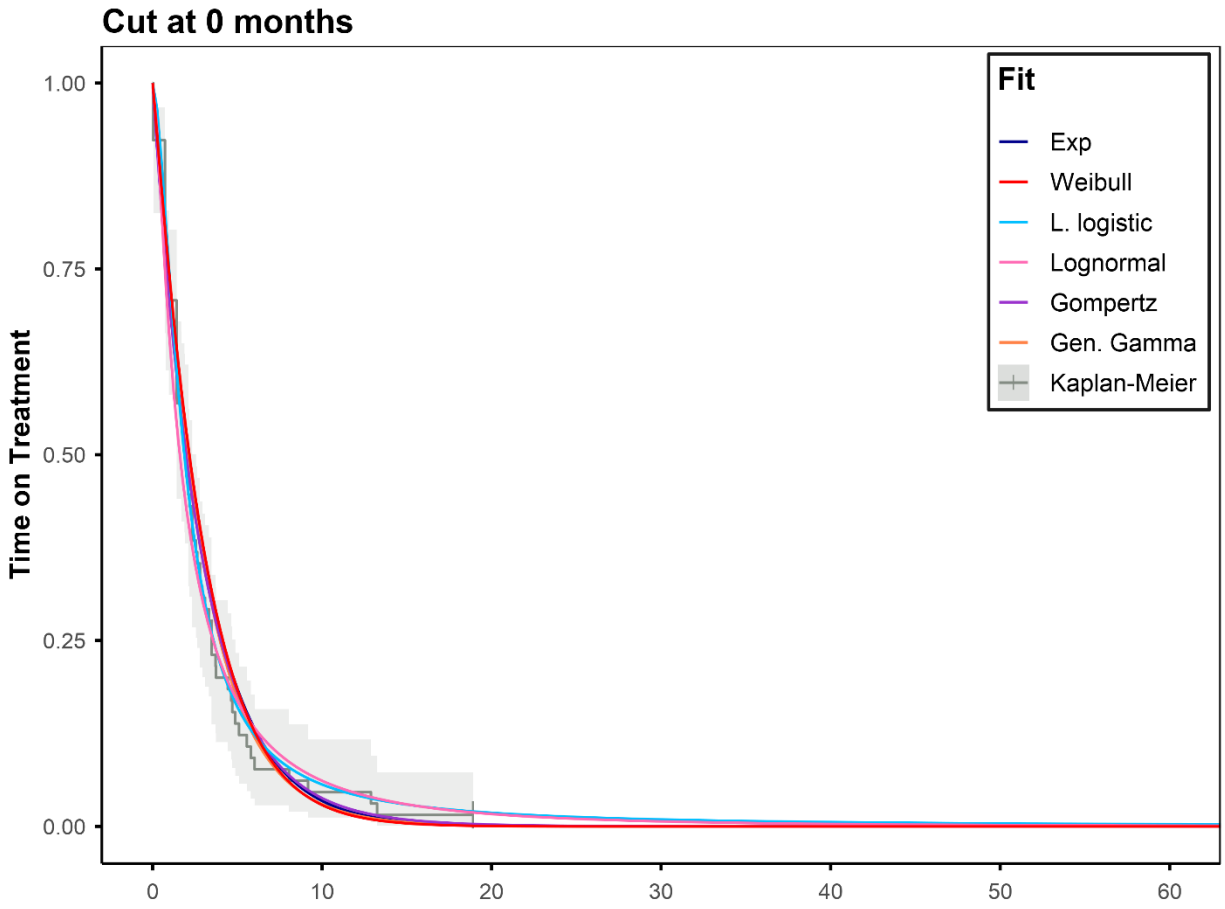
Fit Name	AIC	BIC	Parameters	Median (Months)	Mean (Months)
Exp.	429.90	435.24	rate: 0.2721	3.17	4.31
Weibull	425.06	433.08	shape: 0.8165; scale: 3.4117	3.07	4.38
L. logistic	423.17	431.18	shape: 1.2201; scale: 1.8809	3.15	7.11
Lognormal	431.50	439.51	meanlog: 0.4215; sdlog: 1.5623	3.10	5.09
Gompertz	419.88	427.90	shape: -0.1091; rate: 0.3838	3.12	N/A
Gen. Gamma	425.59	436.28	mu: 1.0245; sigma: 1.3068; Q: 0.7032	3.10	4.40

Figure 7: Semi-parametric (Gelber) models of Progression-Free Survival conditional upon taxane assigned, evaluated for docetaxel. NB: AIC/BIC apply for post 2.99 month period only, and represent the AIC/BIC for both taxanes in the regressed data.



Fit Name	AIC	BIC	Parameters	Median (Months)	Mean (Months)
Exp.	911.51	918.19	rate: 0.2805	2.47	3.56
Weibull	912.04	922.05	shape: 1.0675; scale: 3.6440	2.59	3.55
L. logistic	915.06	925.08	shape: 1.6770; scale: 2.5263	2.53	4.96
Lognormal	947.84	957.85	meanlog: 0.8422; sdlog: 1.1879	2.32	4.70
Gompertz	912.70	922.71	shape: -0.0182; rate: 0.2973	2.38	3.60
Gen. Gamma	911.77	925.12	mu: 1.2066; sigma: 0.9636; Q: 0.7929	2.55	3.59

Figure 8: Parametric models of Time on Treatment conditional upon taxane assigned, evaluated for paclitaxel. NB: AIC/BIC represents the AIC/BIC for both taxanes in the regressed data.



Fit Name	AIC	BIC	Parameters	Median (Months)	Mean (Months)
Exp.	911.51	918.19	rate: 0.3374	2.05	2.96
Weibull	912.04	922.05	shape: 1.0675; scale: 3.0537	2.17	2.98
L. logistic	915.06	925.08	shape: 1.6770; scale: 1.8643	1.86	3.66
Lognormal	947.84	957.85	meanlog: 0.4693; sdlog: 1.1879	1.60	3.24
Gompertz	912.70	922.71	shape: -0.0182; rate: 0.3576	1.97	2.96
Gen. Gamma	911.77	925.12	mu: 1.0001; sigma: 0.9636; Q: 0.7929	2.07	2.92

Figure 9: Parametric models of Time on Treatment conditional upon taxane assigned, evaluated for docetaxel. NB: AIC/BIC represents the AIC/BIC for both taxanes in the regressed data.

Table 5. Alternative survival approach for taxanes use

	Nivolumab	Docetaxel	Paclitaxel
Total QALYs	1.073	0.646	0.658
Total life years (undiscounted)	1.650	1.052	1.051
Total costs	£47,629	£26,175	£28,099
Incremental QALYs	-	0.428	0.415
Incremental life years (undiscounted)	-	0.598	0.599
Incremental costs	-	£21,454	£19,530
ICER (£/QALY)	**-**	**£50,176**	**£47,037**
ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year			

Results from the analysis are detailed in Table 5, where application of taxanes as a covariate resulted in ICER estimates of £50,176 per QALY versus docetaxel (docetaxel as a subgroup: £52,340 per QALY) and £47,037 per QALY versus paclitaxel (paclitaxel as a subgroup: £46,764 per QALY).

	Nivolumab	Docetaxel	Paclitaxel
Total QALYs	1.073	0.646	0.658
Total life years	1.506	1.015	1.014
Total costs	£47,629	£26,175	£28,099
Incremental QALYs	-	0.428	0.415
Incremental life years	-	0.491	0.492
Incremental costs	-	£21,454	£19,530
ICER (£/QALY)	**-**	**£50,176**	**£47,037**
ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year			

Results from the analysis are detailed in Table 5, where application of taxanes as a covariate resulted in ICER estimates of £50,176 per QALY versus docetaxel (docetaxel as a subgroup: £52,340 per QALY) and £47,037 per QALY versus paclitaxel (paclitaxel as a subgroup: £46,764 per QALY).

B7. Please can the company provide the following additional Kaplan-Meier plots:

- **Overlay of overall survival Kaplan-Meier plots from ATTRACTION-1 and ATTRACTION-3 (i.e. three arms: [1] nivolumab 3mg/kg, [2] nivolumab 240mg, and [3] taxane)**
- **As above, but for the outcome of PFS**

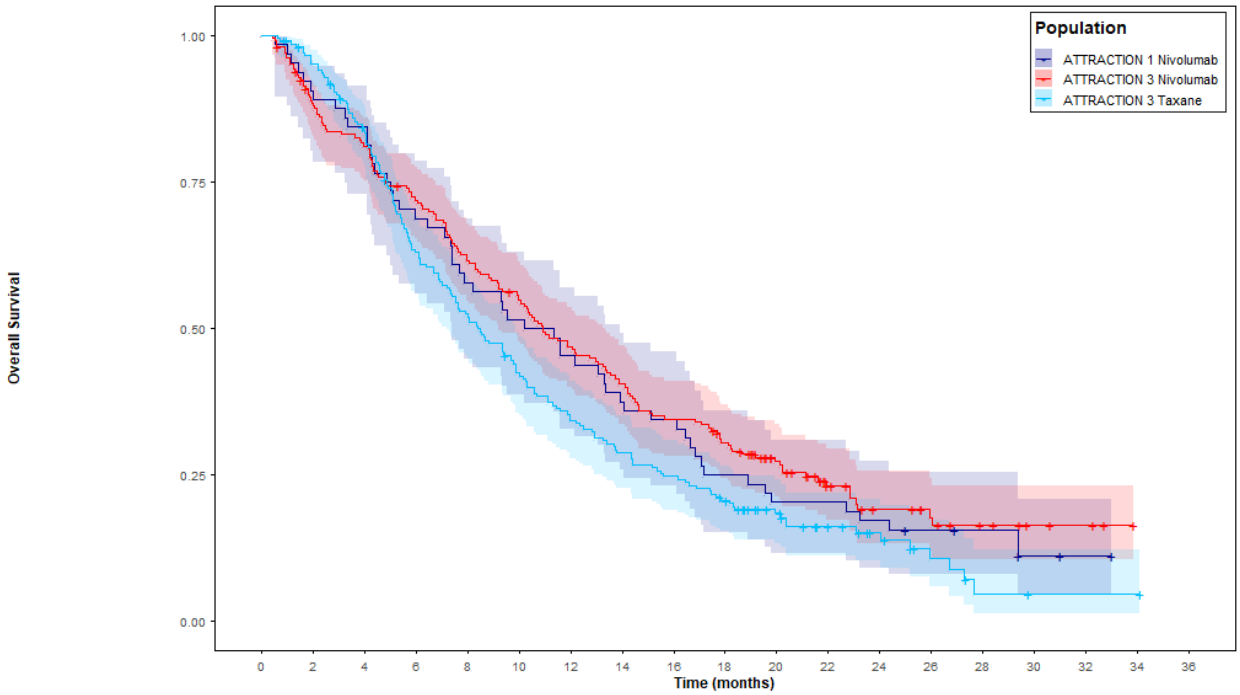
- **Re-creation of CS Figures 16 and 17 (ATTRACTION-1) <u>without</u> the presentation of parametric survival models**

When preparing these plots, the ERG requests that the following features are included/reflected:

  - **Time units are presented in 1- or 2-monthly increments (for ease of comparison with the figures produced for ATTRACTION-3)**
  - **The curve is a 'true' Kaplan-Meier plot, wherein lines are perpendicular and if the final observation was an event the curve should hit 0%**
  - **Numbers at risk are included**

An overlay of OS and PFS KM from ATTRACTION-1 and ATTRACTION-3 can be seen in Figure 4 and Figure 5 respectively. PLD was not available for ATTRACTION-1; therefore the PLD was recreated from digitised data. Data was digitised in DigitizeIt™ and then reconstructed using R version 3.6.2 and the Survival package (v3.1.8).

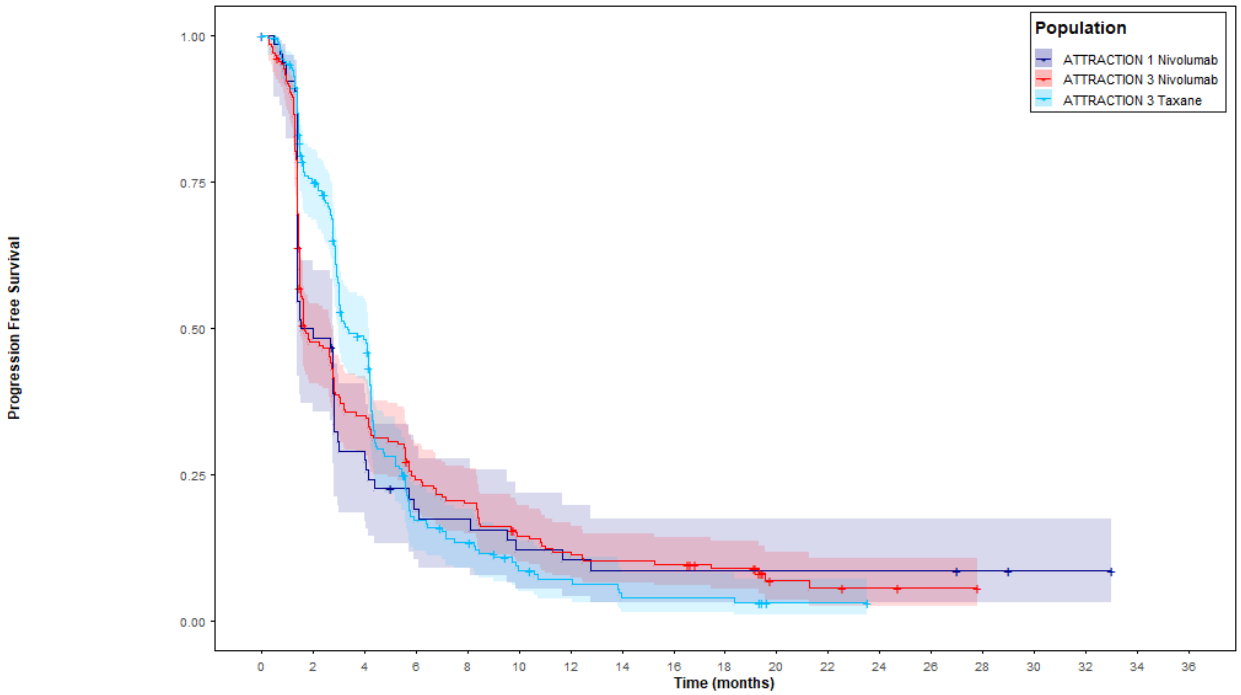
Figure 6 and Figure 7 show the ATTRACTION-1 KM without parametric models.



NAR (Cumulative Events)

ATTRACTION 1 Nivolumab	64 (0)	58 (6)	54 (10)	44 (20)	37 (27)	33 (31)	29 (35)	24 (40)	22 (42)	16 (48)	13 (51)	13 (51)	11 (53)	9 (54)	7 (54)	2 (56)	1 (56)	0 (56)	0 (56)
ATTRACTION 3 Nivolumab	210 (0)	182 (24)	167 (39)	147 (58)	126 (79)	111 (93)	95 (109)	82 (122)	70 (134)	60 (142)	43 (148)	25 (154)	17 (158)	13 (159)	7 (160)	4 (160)	3 (160)	0 (160)	0 (160)
ATTRACTION 3 Taxane	209 (0)	196 (10)	169 (33)	126 (75)	105 (96)	84 (116)	68 (132)	57 (143)	49 (151)	40 (159)	27 (163)	17 (166)	12 (167)	6 (170)	2 (173)	1 (173)	1 (173)	1 (173)	0 (173)

Figure 10: Overlay of Kaplan Meier curves from ATTRACTION-1 and ATTRACTION-3 for Overall Survival



NAR (Cumulative Events)

ATTRACTION 1 Nivolumab	64 (0)	31 (33)	18 (45)	11 (51)	10 (52)	7 (55)	6 (56)	5 (57)	5 (57)	5 (57)	5 (57)	5 (57)	5 (57)	5 (57)	3 (57)	2 (57)	2 (57)	0 (57)	0 (57)
ATTRACTION 3 Nivolumab	210 (0)	96 (108)	71 (133)	48 (155)	40 (163)	27 (174)	22 (179)	19 (182)	18 (183)	13 (184)	5 (186)	4 (187)	3 (187)	1 (187)	0 (187)	0 (187)	0 (187)	0 (187)	0 (187)
ATTRACTION 3 Taxane	209 (0)	147 (50)	89 (101)	29 (156)	22 (162)	12 (169)	9 (171)	5 (175)	5 (175)	5 (175)	1 (176)	1 (176)	0 (176)	0 (176)	0 (176)	0 (176)	0 (176)	0 (176)	0 (176)

Figure 11: Overlay of Kaplan Meier curves from ATTRACTION-1 and ATTRACTION-3 for Progression Free Survival

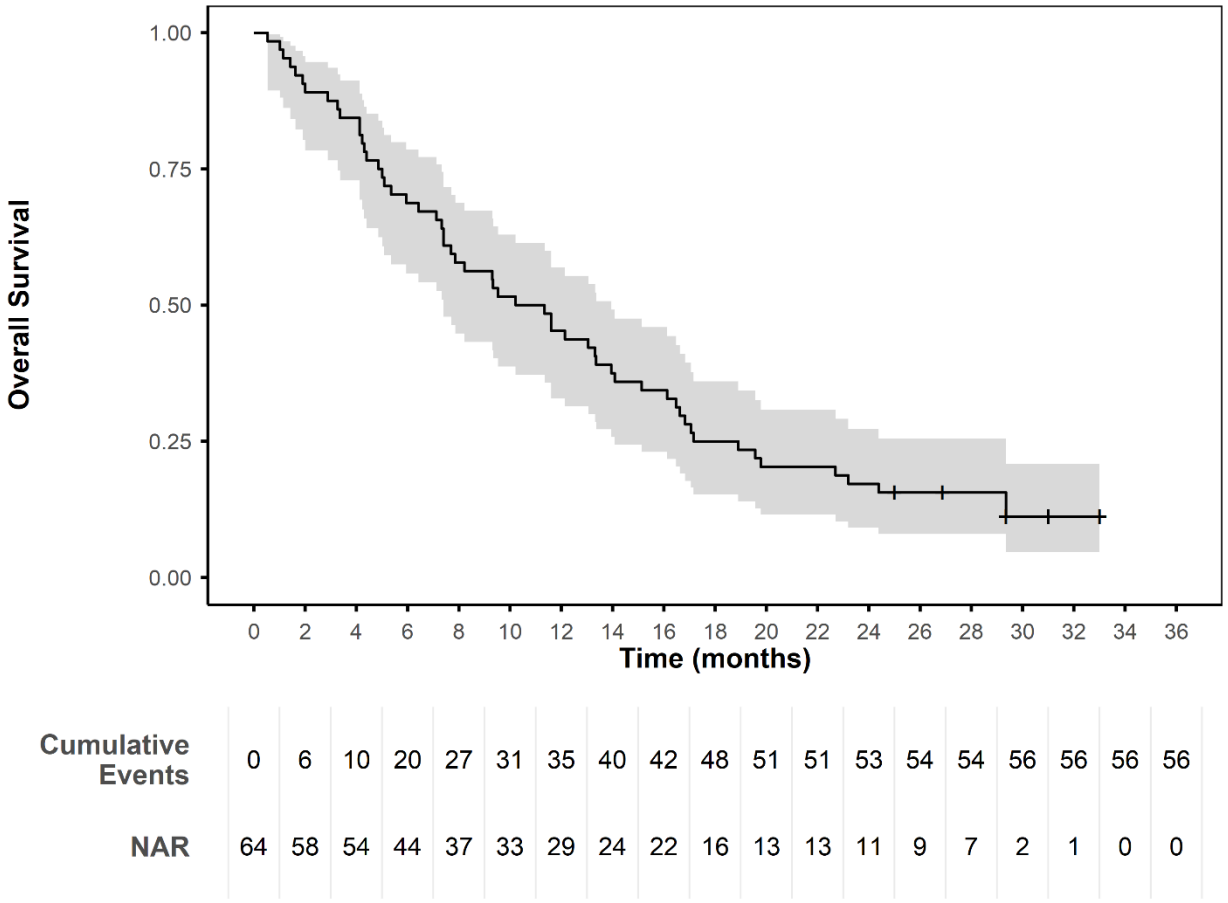


Figure 12: Recreation of the Overall Survival figure from the ATTRACTION-1 Clinical Study Report

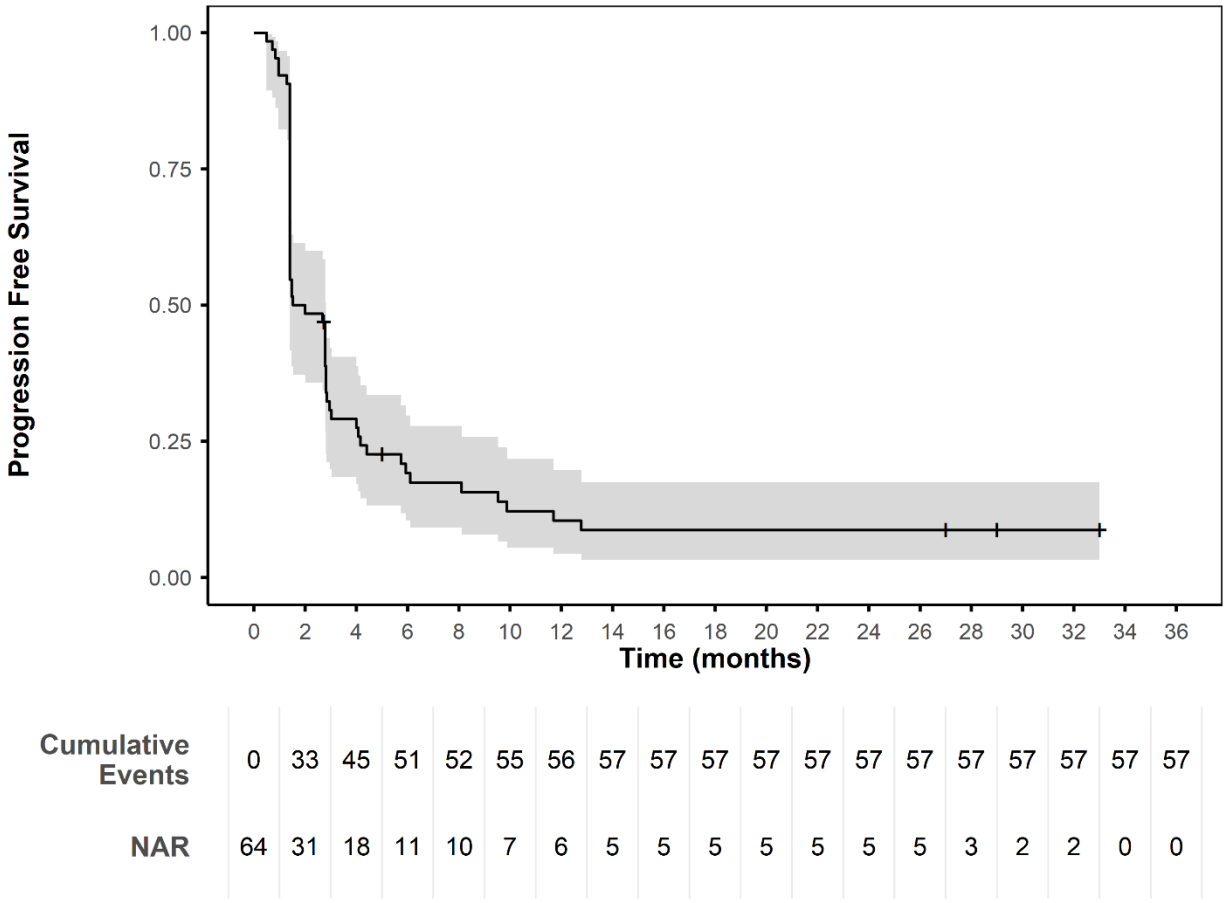


Figure 13: Recreation of the Progression Free Survival figure from the ATTRACTION-1 Clinical Study Report

B8. In Section B.2.9.3 of the CS, a log-cumulative hazard plot is presented comparing docetaxel and BSC. Please can the company clarify that this is based on the study by Moriwaki *et al.* only. If this is correct, please can the company provide the following additional plots:

- **Including an additional arm for the taxane arm in ATTRACTION-3**

Document B Figure 19 of the company submission does depict the Moriwaki et al study only.[1, 7] Figure 8 shows the log cumulative hazard plot for the arms in the Moriwaki et al study, presented with the combined taxane arm from ATTRACTION-3. This shows that the assumption of proportional hazards between the combined taxane arm and BSC or docetaxel alone is violated. It would therefore not be considered appropriate to apply a

generated hazard ratio (HR) to the combined taxane arm from ATTRACTION-3 in order to derive an estimate for BSC.

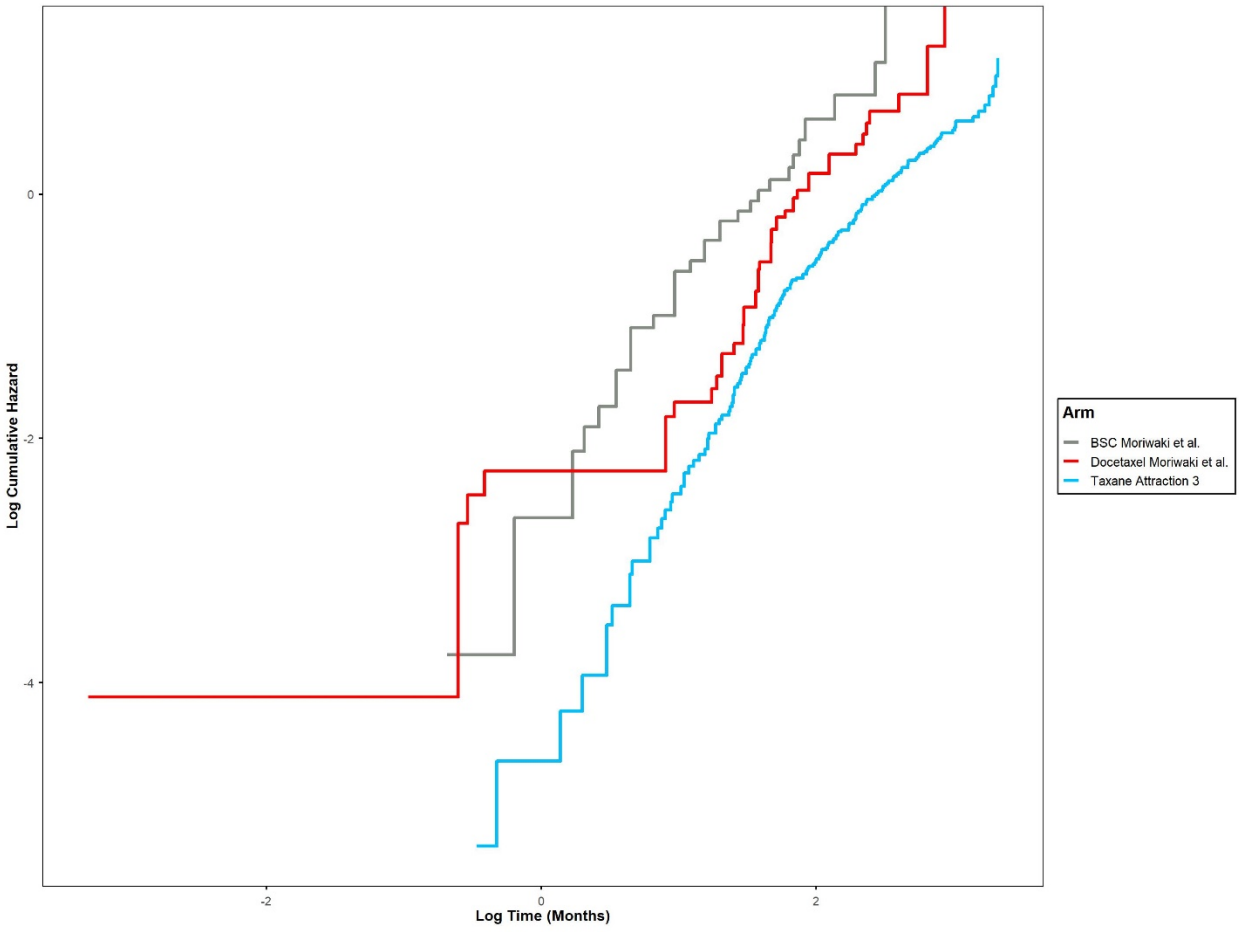


Figure 14: Log Cumulative Hazard for BSC and docetaxel arms from Moriwaki et al. and the ATTRACTION-3 taxane arm

- **Including an additional arm for the docetaxel group in ATTRACTION-3**

Figure 9 shows the log cumulative hazard plot for the arms in the Moriwaki et al. study, presented with the docetaxel arm from ATTARCTION-3. This shows that the assumption of proportional hazards between the docetaxel arm and BSC is intact. It would therefore be considered appropriate to apply a generated hazard ratio (HR) to the docetaxel arm from ATTRACTION-3 in order to derive an estimate for BSC. This was the rationale for the analysis in the CS where BSC was a considered comparator.

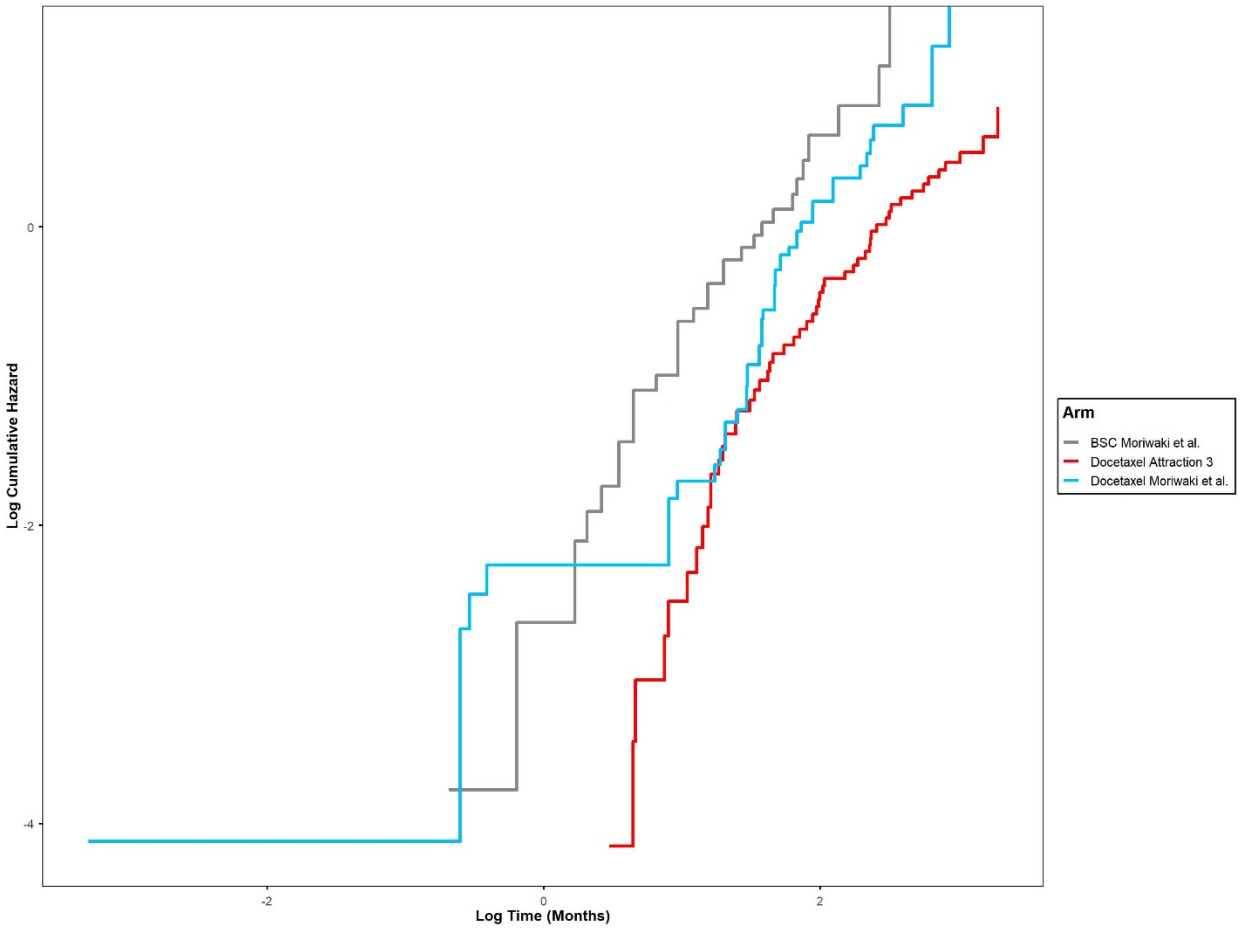


Figure 15: Log Cumulative Hazard for BSC and docetaxel arms from Moriwaki et al. and the ATTRACTION-3 docetaxel arm

B9. In Section B.3.2.2 of the CS, it is stated: *"Applying Kaplan-Meier data until 2.99 months followed by parametric extrapolation enabled the initial hazard to be modelled appropriately and captured the high rate of events between study entry and second assessment, which was scheduled for 12 weeks."* Please can the company confirm how many days were assumed to be in each month for the purpose of this analysis? If the value is not 84 days (12 weeks x 7 = 84 days), please can the company provide a sensitivity analysis wherein this cut point is specified instead of 2.99 months? Please ensure the ability to reproduce the results of this sensitivity analysis is incorporated within the economic model file

Trial analysis describes one month as 30.436 day. Consequently, 2.99 months (as presented for base case analysis) is equivalent to 91.0 days.

When selecting cut points, it is best practice to avoid dates where assessments are scheduled, which reflect a short period where several events occur and hence may bias the extrapolation. Whilst the effect may not be large, it is considered to be more scientifically robust. This approach was applied within the survival analysis provided within the company submission and has continued to be used within this response.

The Clinical Study Protocol for ATTRACTION-3 details that the target assessment date at “12 weeks” should be 84 days (12 weeks x 7 days = 84 days), but this assessment can be made in the 7 days prior or post the target date (i.e. 77-91 days).[1] When conducting survival analysis, cut points were deliberately placed outside assessment dates with a 7 day buffer to accommodate any events that will appear in this window. Hence, the 2.99 month (i.e. 91 day) cut point was identified, so that all events occurring after 91.0 days will inform the survival extrapolation.

Figure 10 and Figure 11 show OS and PFS, respectively, from ATTRACTION-3 with dashed lines representing 84 and 91 days. As can be expected, Figure 11 clearly depicts a large number of PFS events occurring within this window, which may bias extrapolations. Therefore, the company consider that any analysis where the cut point is placed at the exact assessment date, amidst events which should be considered for this assessment time, is not appropriate. Analyses where the cut point is placed just before or after the assessment period are the most appropriate and therefore, the cut point at 2.99 months where this is 91 days is more robust than placing a cut point at 84 days.

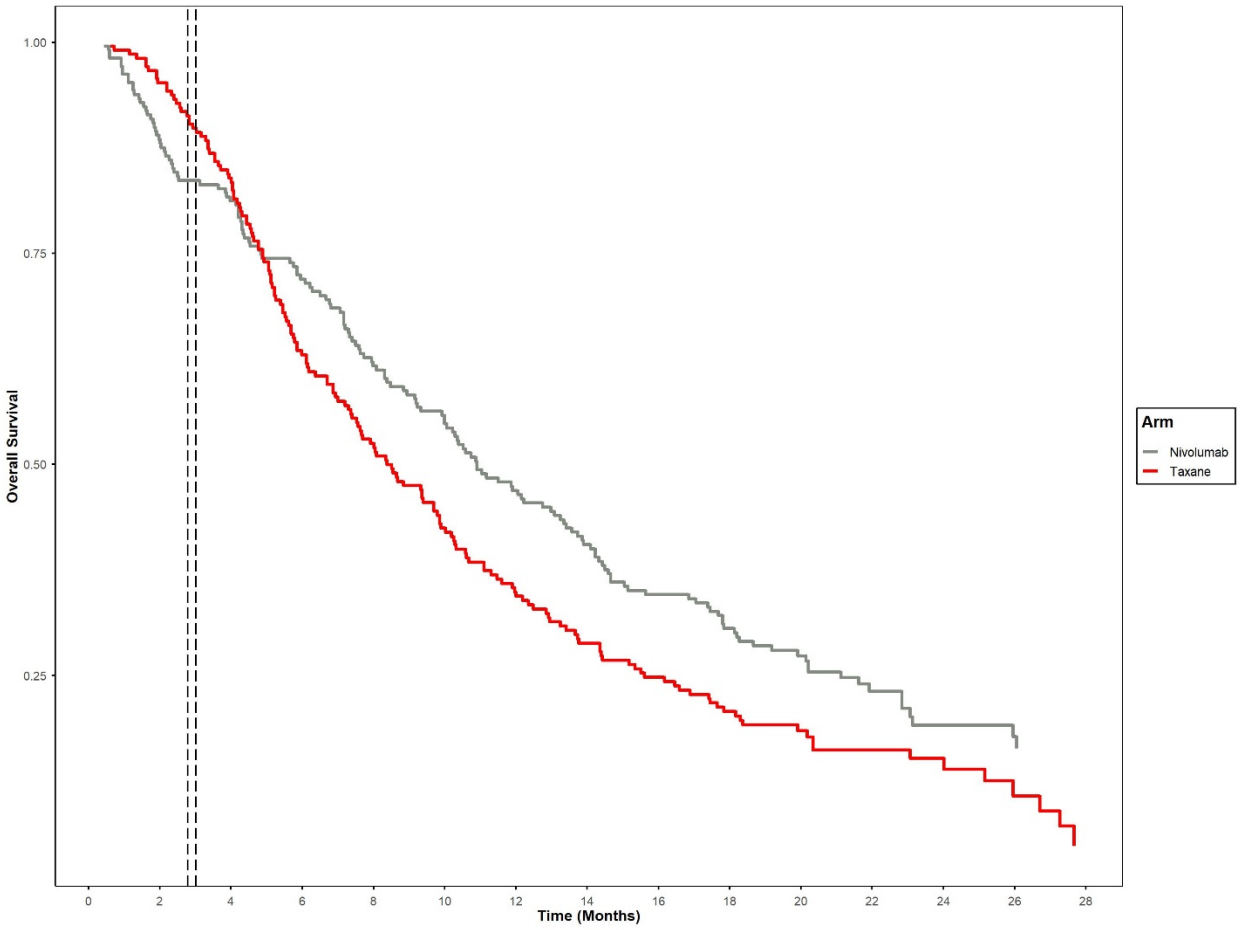


Figure 16: Overall Survival for ATTRACTION-3 with indicators for assessment times + 7 days

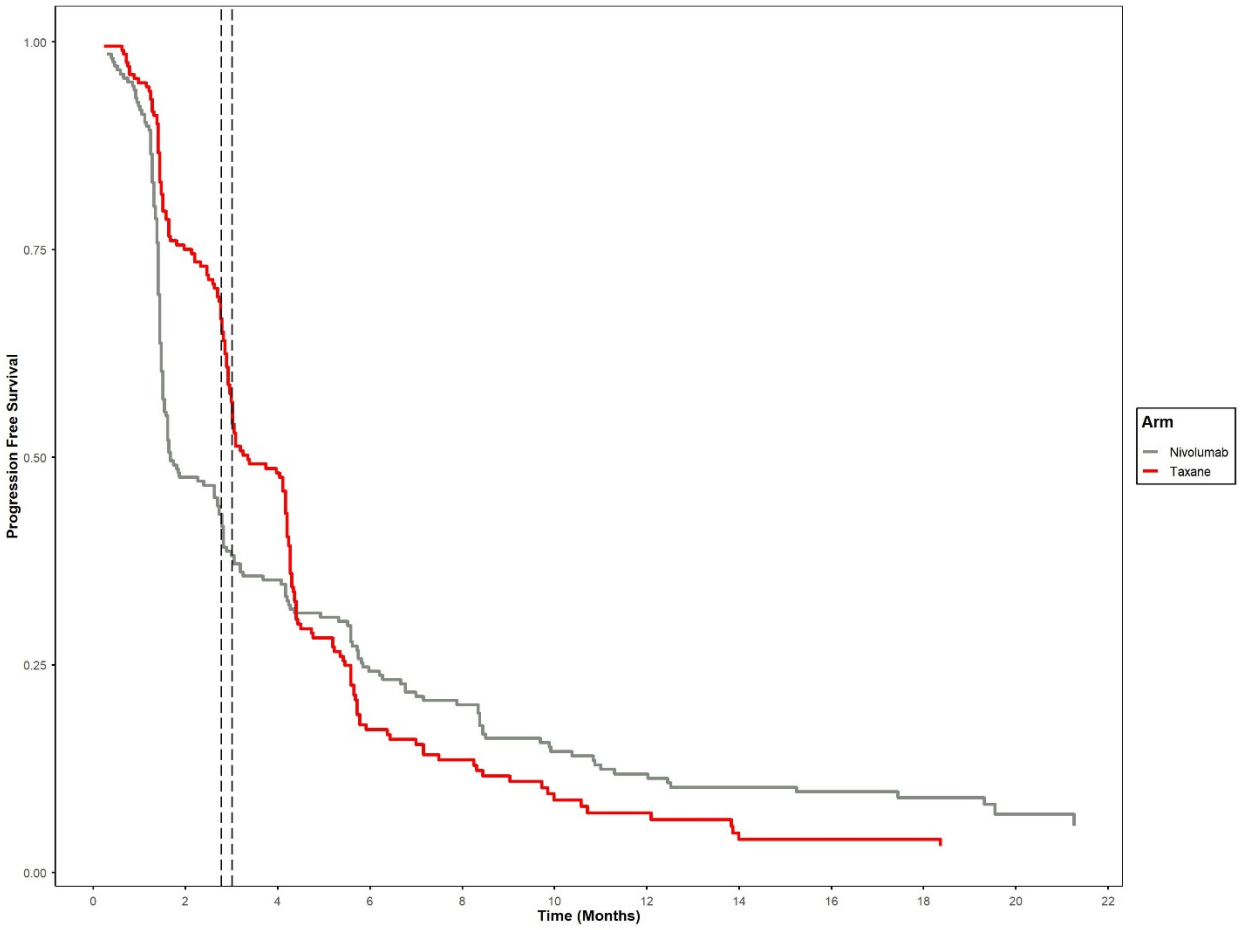


Figure 17: Progression Free Survival for ATTRACTION-3 with indicators for assessment times + 7 days

B10. PRIORITY QUESTION: Please can the company provide the following options to estimate OS, PFS, and ToT?

- **Spline-based models (which are discussed in addition to the fully-parametric and semi-parametric in document M) with 1 internal knot (or more, if deemed suitable)**

Spline models

Models with 1 internal knot are presented where the internal knot is placed at 2.99 months (akin to the semi parametric models used in the base case analysis) and where they are arbitrarily placed at the median of the log time of events.

The company would caution against using Royston-Parmar splines to model data with clear structural discontinuities, as this imposes conditions of smoothness on the hazard that are observed to be inconsistent with the data and can result in oscillation of the predictions and poor long-term performance due to the rapidly changing derivative of the hazard function. This is particularly true when a knot is placed near to such a discontinuity. The “Gelber” semi-parametric piecewise approach does not impose this smoothness upon the model, and so pieces may begin immediately after a rapid change in hazard.

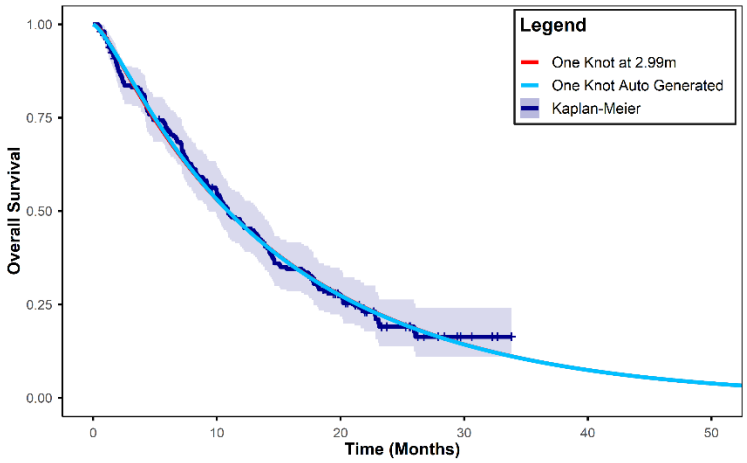
In this case, the models proved insensitive to the position of the internal knot, indicating that the average rate of change of gradient of log-hazard with log time was consistent throughout the trial period.

Where these have been provided, hazard spline models were chosen as they are most easily interpreted. The gradient of the curve is fixed on the hazard and so it degenerates to a Weibull model. This is one reason why the company does not consider the spline models to be appropriate for profiles that require extrapolation; this estimate is made at the point of the last observation as opposed to using all data as a standard parametric fit.

As can be seen, the provided spline models do not have an improved fit over the semi-parametric approach. Some models may even be considered inappropriate, due to the poor fit to the available data. In particular, the spline models for PFS in both arms and OS in the nivolumab arm exceeded the 95% confidence intervals for the observed data, indicating the implausibility of the fit. Where appropriate models were available, these did not provide improved fit over the base case analysis models provided in the company submission.

Cost-effectiveness outcomes are provided in Table 6 and Table 7. As can be seen, the majority of spline fits provided comparable ICERs to the base case analysis. The exception is spline models of nivolumab OS, which visibly overestimates hazard in the latter period of the study, exceeding the 95% confidence intervals of the observed data.

Nivolumab spline models



Fit Name	AIC	BIC	Mean (Months)	Median (Months)
One knot at 2.99m	1673.91	1700.00	15.66 (13.37, 19.32)	10.90 (9.24, 12.81)
One Knot Auto Generated	1674.75	1700.83	15.67 (13.36, 19.46)	10.91 (9.27, 12.81)

Figure 18: Spline model for Overall Survival for ATTRACTION-3 Nivolumab arm (Kaplan-Meier plot)

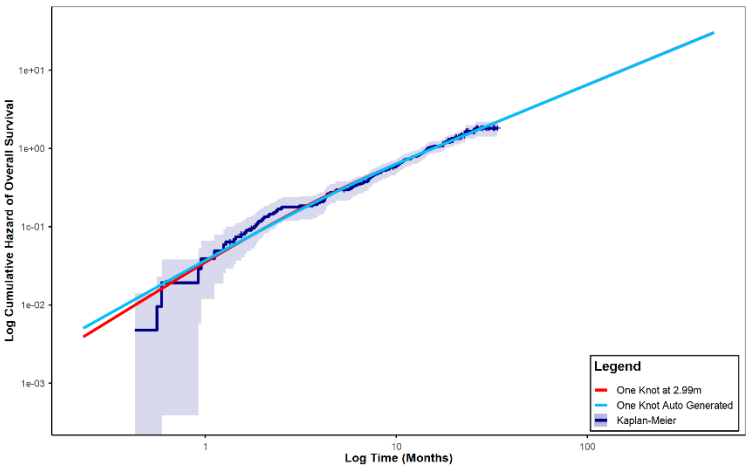
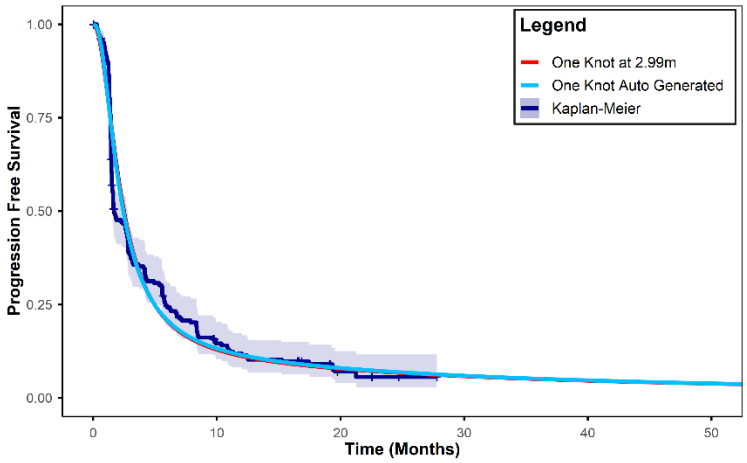


Figure 19: Spline model for Overall Survival for ATTRACTION-3 Nivolumab arm (cumulative hazard plot)



Fit Name	AIC	BIC	Mean (Months)	Median (Months)
One knot at 2.99m	1458.48	1484.56	11.37 (5.83, 83.91)	2.49 (2.19, 2.88)
One Knot Auto Generated	1449.22	1475.30	11.23 (5.80, 56.82)	2.43 (2.13, 2.81)

Figure 20: Spline models of Progression-Free Survival for ATTRACTION-3 Nivolumab arm (Kaplan-Meier plot)

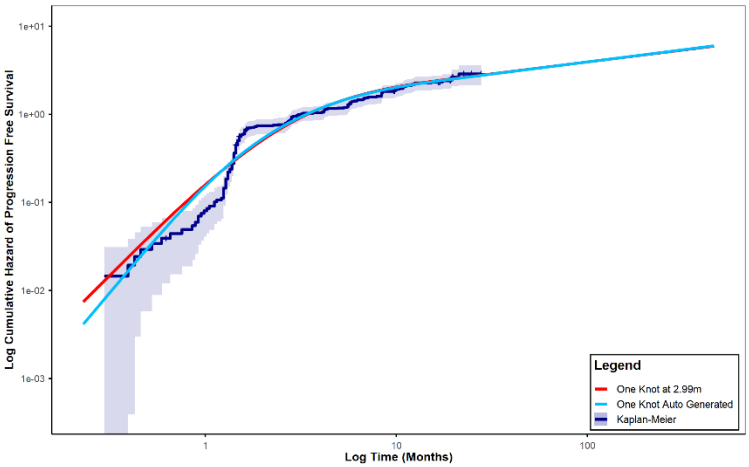
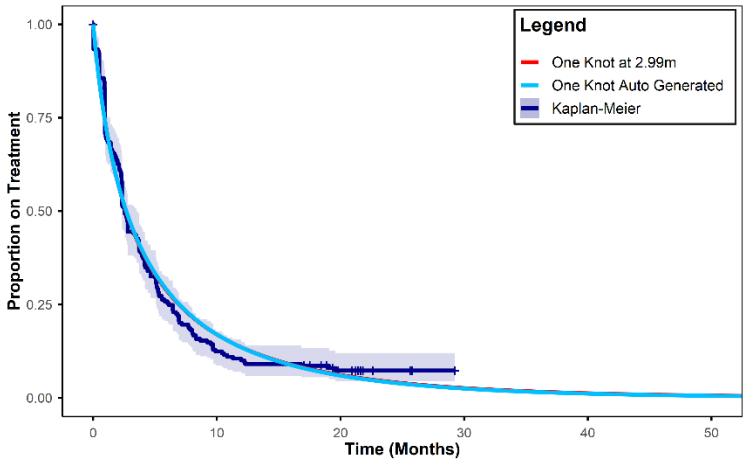


Figure 21: Spline models of Progression-Free Survival for ATTRACTION-3 Nivolumab arm (cumulative hazard plot)



Fit Name	AIC	BIC	Mean (Months)	Median (Months)
One knot at 2.99m	1573.09	1599.17	5.82 (4.56, 8.26)	2.66 (2.20, 3.22)
One Knot Auto Generated	1573.71	1599.79	5.78 (4.53, 7.95)	2.68 (2.21, 3.24)

Figure 22: Spline models of Time on Treatment for ATTRACTION-3 Nivolumab arm (Kaplan-Meier plot)

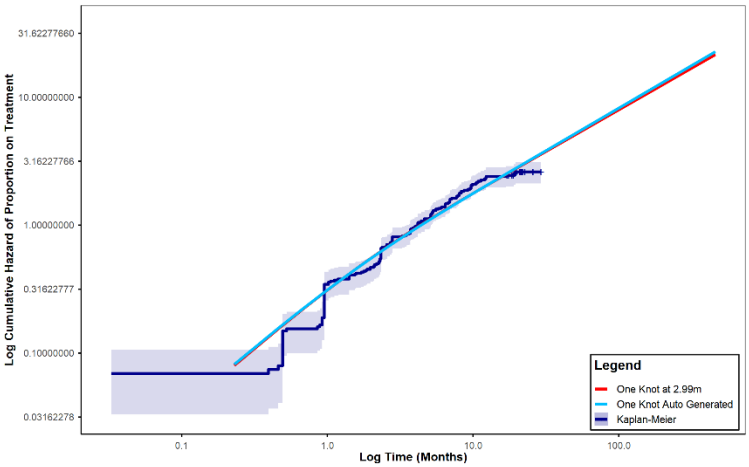
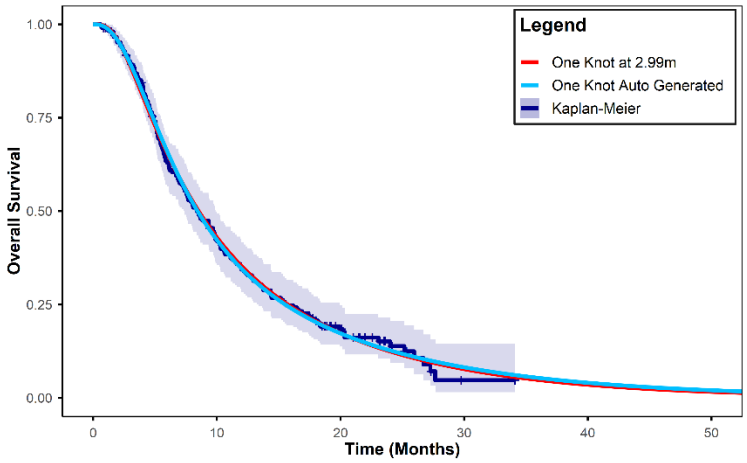


Figure 23: Spline models of Time on Treatment for ATTRACTION-3 Nivolumab arm (cumulative hazard plot)

Taxane spline models



Fit Name	AIC	BIC	Mean (Months)	Median (Months)
One knot at 2.99m	1691.31	1717.36	12.31 (10.77, 14.67)	8.53 (7.41, 9.76)
One Knot Auto Generated	1690.58	1716.63	12.50 (10.84, 15.14)	8.45 (7.34, 9.64)

Figure 24: Spline models of Overall Survival for ATTRACTION-3 pooled taxane arms (Kaplan-Meier plot)

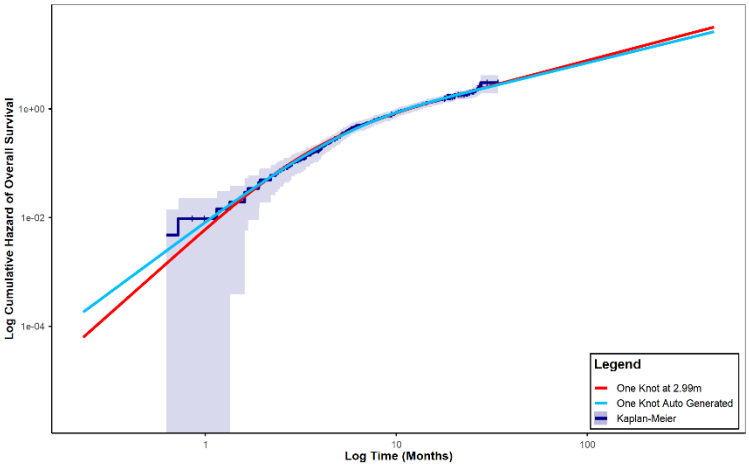
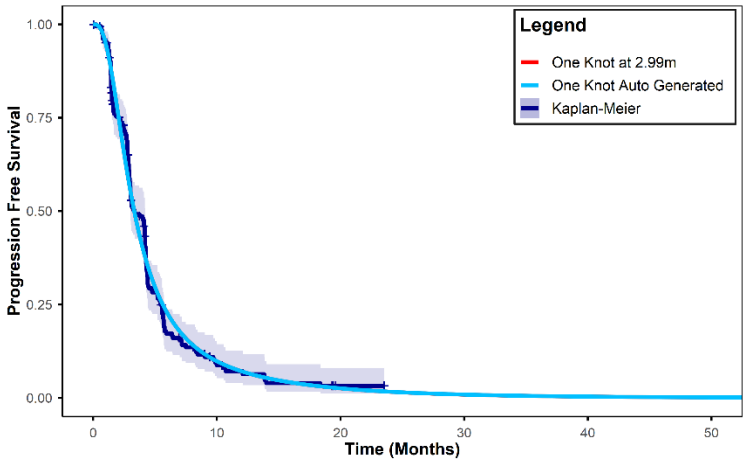


Figure 25: Spline models of Overall Survival for ATTRACTION-3 pooled taxane arms (cumulative hazard plot)



Fit Name	AIC	BIC	Mean (Months)	Median (Months)
One knot at 2.99m	1369.11	1395.16	4.96 (4.15, 6.82)	3.32 (2.97, 3.66)
One Knot Auto Generated	1369.06	1395.12	4.96 (4.15, 6.52)	3.32 (2.97, 3.67)

Figure 26: Spline models of Progression-Free Survival for ATTRACTION-3 pooled taxane arms (Kaplan-Meier plot)

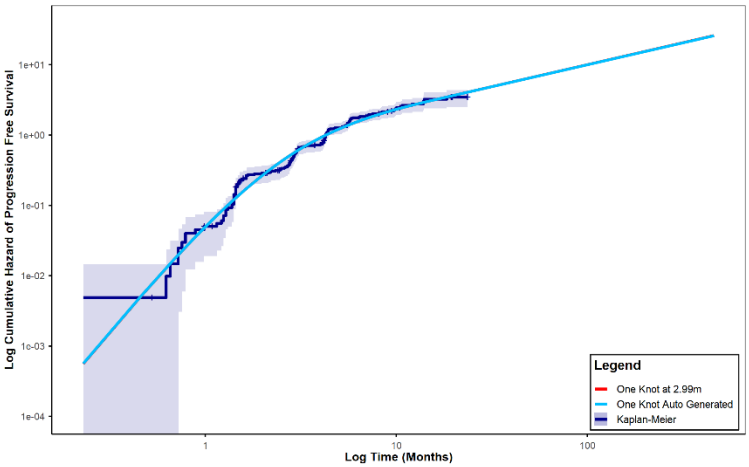
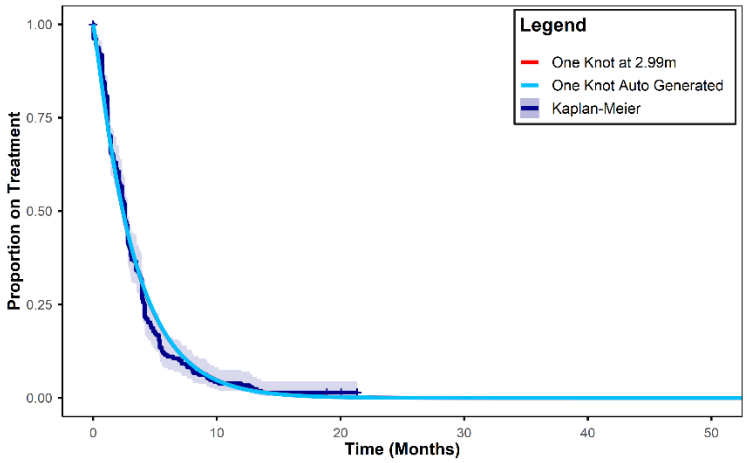


Figure 27: Spline models of Progression-Free Survival for ATTRACTION-3 pooled taxane arms (cumulative hazard plot)



Fit Name	AIC	BIC	Mean (Months)	Median (Months)
One knot at 2.99m	1515.41	1541.46	3.37 (2.90, 3.90)	2.39 (2.06, 2.72)
One Knot Auto Generated	1515.62	1541.67	3.37 (2.91, 3.89)	2.40 (2.07, 2.73)

Figure 28: Spline models of Time on Treatment for ATTRACTION-3 pooled taxane arms (Kaplan-Meier plot)

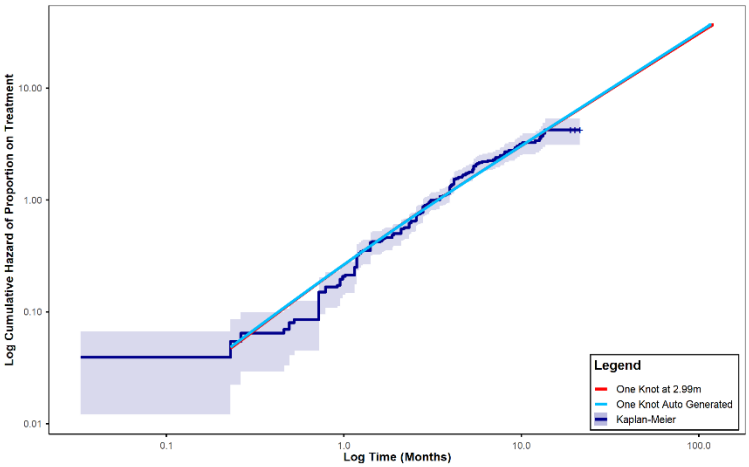


Figure 29: Spline models of Time on Treatment for ATTRACTION-3 pooled taxane arms (cumulative hazard plot)

- **Semi-parametric approach using an alternative cut-point of approximately 4 months (i.e. a time point avoiding the short period for the outcome of OS in the nivolumab arm where no events are experienced)**

Semi-parametric (4.37 month cut point)

A cut point of 4 months would relate to approximately 122 days, which would be within the time frame for assessment at week 18 (127 days). As outlined in the response to question B9, it is best practice to avoid dates where assessments are scheduled, which

reflect a short period where several events occur and hence may bias the extrapolation. This approach has been consistently applied within the survival analyses provided within the company submission and those used within this response.

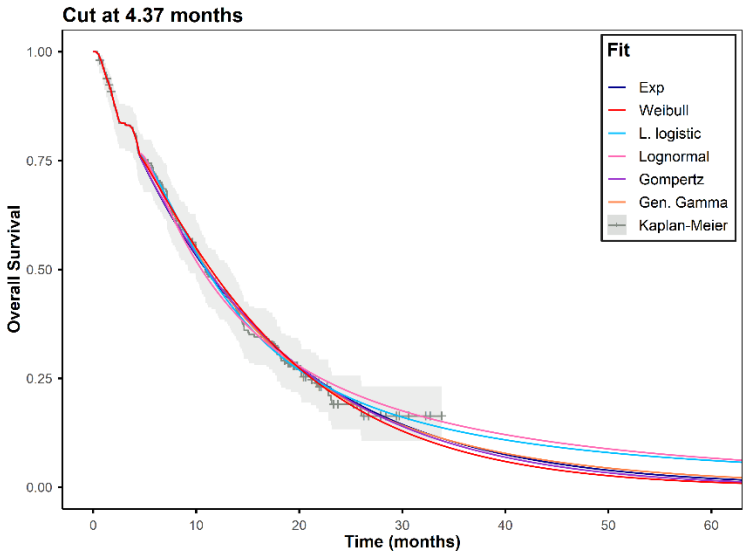
Thus, there are two potential options to addressing this response: provide a cut point before or after the assessment window. In line with the perceived request from the ERG, the decision was taken to ensure at least 4 months of Kaplan-Meier data were modelled directly. Hence, a cut-point of 4.37 months was applied, as this is outside the assessment window and directly models at least 4 months of Kaplan-Meier data.

As can be seen, the provided semi-parametric models typically have improved fit over the spline models; however, it is arguable whether there is improved fit over the semi-parametric approach where the cut point is provided at 2.99 months. Some models may be considered inappropriate, due to the poor fit to the available data or the implausible predicted mean time to event, but this can be considered in line with semi-parametric models with cut point at 2.99 months.

It should be noted that applying a cut point at 4.37 months, as opposed to 2.99 months, has a number of disadvantages. Although it allows for a greater number of people to demonstrate response in the nivolumab arm, it reduces the number of data points available to inform the extrapolation period. This particularly impacts the analysis of PFS, where median is 1.68 months in the nivolumab arm and 3.35 months in the taxane arm, and time on treatment, where the majority of patients have discontinued by 4.37 months.

Cost-effectiveness outcomes are provided in Table 6 and Table 7. As can be seen, the majority of plausible semi-parametric models provided comparable ICERs to the base case analysis, with several exceptions: nivolumab OS generalised gamma model (£60,571); nivolumab time on treatment lognormal (£61,853) and generalised gamma (£58,878). Overall, ICERs ranged from £43,412 to £63,418 for plausible semi-parametric models.

Nivolumab semi-parametric models (4.37 month cut point)



Fit Name	AIC	BIC	Parameters	Median (Months)	Mean (Months)
Exp.	837.45	840.52	lambda: 0.0652	10.95 (9.43, 12.76)	15.66 (13.52, 18.55)
Weibull	838.05	844.17	shape: 1.1032; scale: 15.1857	11.42 (9.76, 13.32)	15.13 (13.11, 18.02)
L. logistic	836.02	842.14	shape: 1.4387; scale: 10.1866	10.97 (9.44, 12.75)	24.77 (18.00, 40.51)
Lognormal	842.24	848.37	mu: 2.3057; sigma: 1.2611	10.52 (9.03, 12.34)	20.95 (16.13, 29.62)
Gompertz	839.41	845.53	shape: 0.0035; rate: 0.0635	11.05 (9.29, 13.09)	15.39 (12.96, 17.93)
Gen. Gamma	838.60	847.79	mu: 2.5846; sigma: 1.0297; Q: 0.6416	11.14 (9.45, 12.84)	16.01 (13.78, 22.09)

Figure 30. Semi-parametric model (4.37 month cut point) of overall survival: ATTRACTION-3 nivolumab arm (Kaplan-Meier plot)

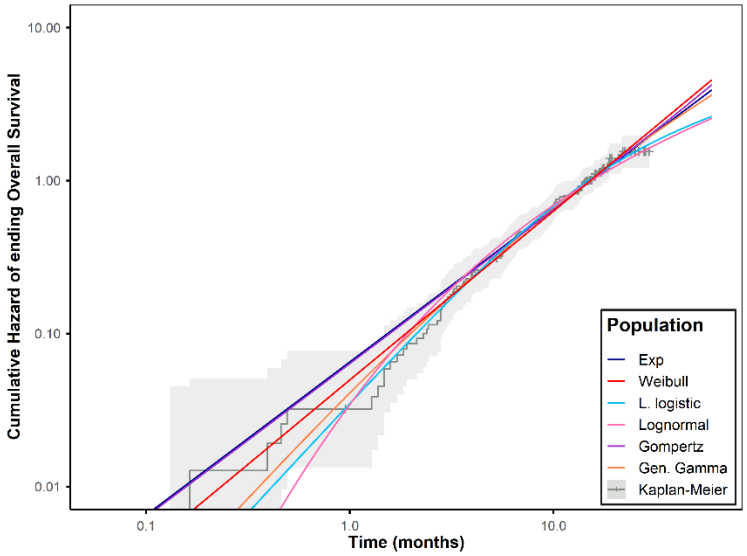
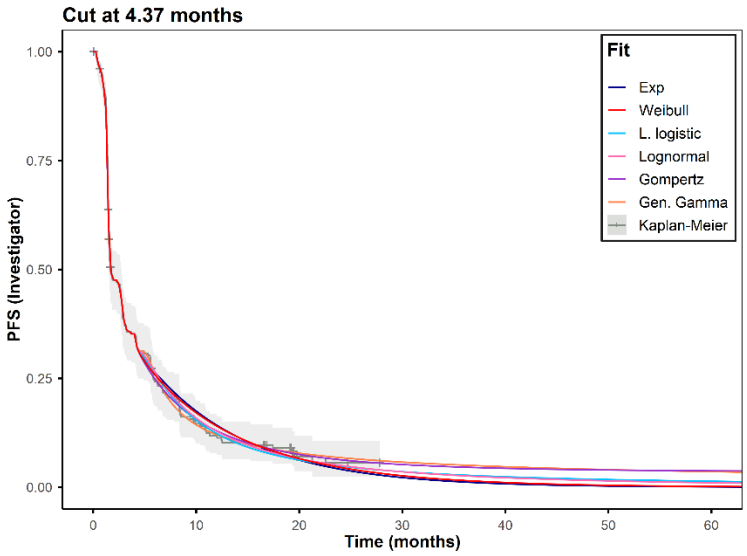


Figure 31. Semi-parametric model (4.37 month cut point) of overall survival: ATTRACTION-3 nivolumab arm (cumulative hazard plot)



Fit Name	AIC	BIC	Parameters	Median (Months)	Mean (Months)
Exp.	303.14	305.29	lambda: 0.1030	1.68 (1.54, 2.79)	5.61 (4.49, 7.01)
Weibull	304.77	309.06	shape: 0.9300; scale: 9.7022	1.68 (1.54, 2.79)	5.71 (4.48, 7.41)
L. logistic	297.38	301.66	shape: 1.3313; scale: 5.5216	1.68 (1.54, 2.79)	8.35 (5.23, 22.28)
Lognormal	294.82	299.10	mu: 1.7525; sigma: 1.2341	1.68 (1.54, 2.79)	6.43 (4.74, 9.65)
Gompertz	300.78	305.07	shape: -0.0688; rate: 0.1485	1.68 (1.54, 2.79)	N/A (4.26, N/A)
Gen. Gamma	290.30	296.73	mu: 0.9252; sigma: 1.0358; Q: -1.5560	1.68 (1.54, 2.79)	N/A (5.30, N/A)

Figure 32. Semi-parametric model (4.37 month cut point) of progression-free survival: ATTRACTION-3 nivolumab arm (Kaplan-Meier plot)

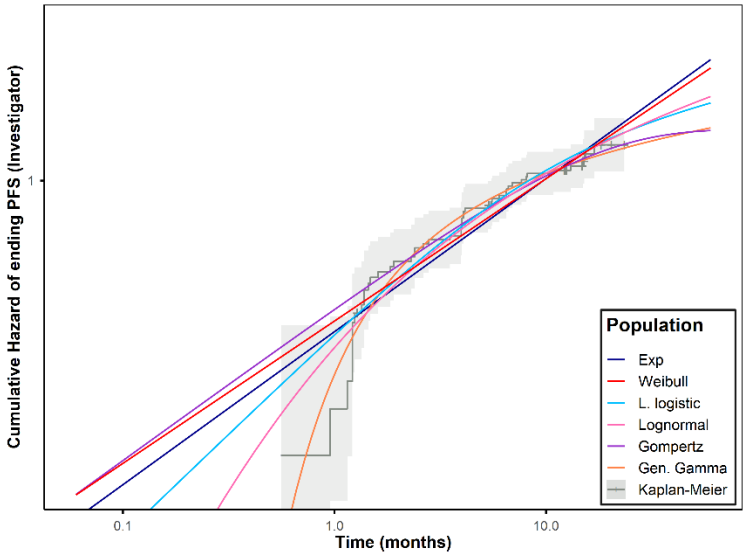
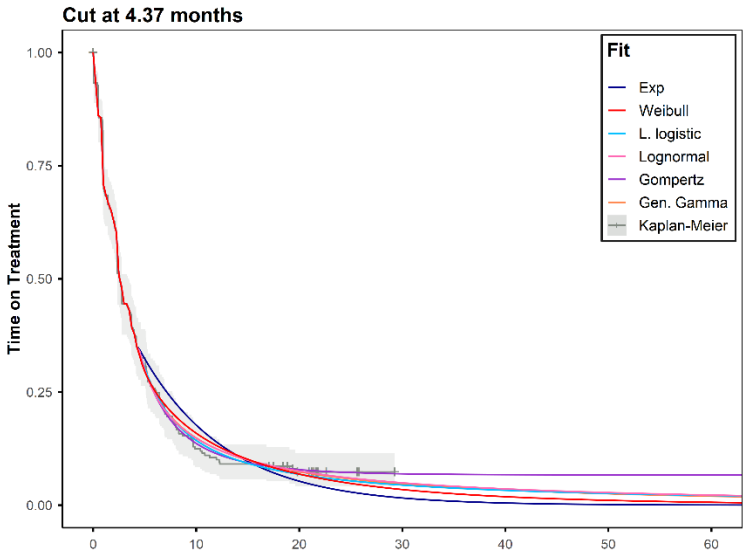


Figure 33. Semi-parametric model (4.37 month cut point) of progression-free survival: ATTRACTION-3 nivolumab arm (cumulative hazard plot)



Fit Name	AIC	BIC	Parameters	Median (Months)	Mean (Months)
Exp.	357.05	359.34	lambda: 0.1207	2.56 (2.33, 3.48)	5.53 (4.45, 6.96)
Weibull	347.92	352.50	shape: 0.7117; scale: 7.8784	2.56 (2.33, 3.48)	6.06 (4.65, 8.46)
L. logistic	340.32	344.90	shape: 1.0313; scale: 3.9490	2.56 (2.33, 3.48)	46.72 (6.29, 131.67)
Lognormal	342.43	347.01	mu: 1.3930; sigma: 1.7234	2.56 (2.33, 3.48)	8.84 (5.44, 18.60)
Gompertz	336.94	341.53	shape: -0.1494; rate: 0.2470	2.56 (2.33, 3.48)	N/A
Gen. Gamma	344.34	351.21	mu: 1.4693; sigma: 1.6919; Q: 0.1068	2.56 (2.33, 3.48)	8.03 (5.04, 34.82)

Figure 34. Semi-parametric model (4.37 month cut point) of time on treatment: ATTRACTION-3 nivolumab arm (Kaplan-Meier plot)

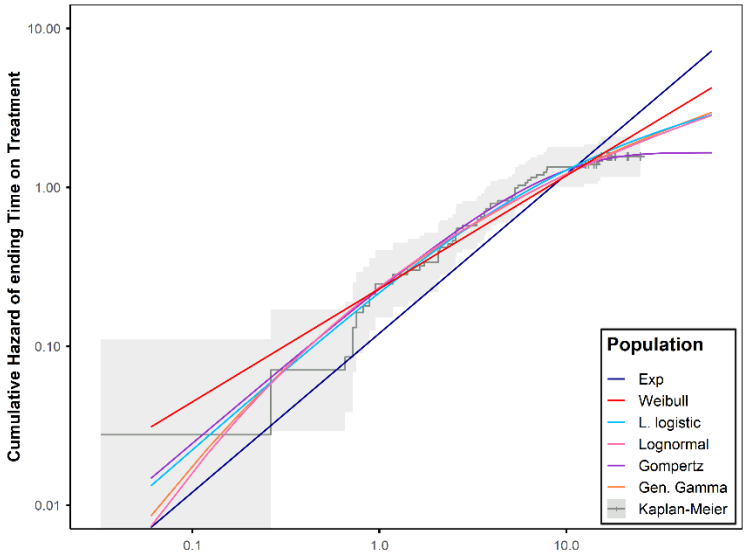
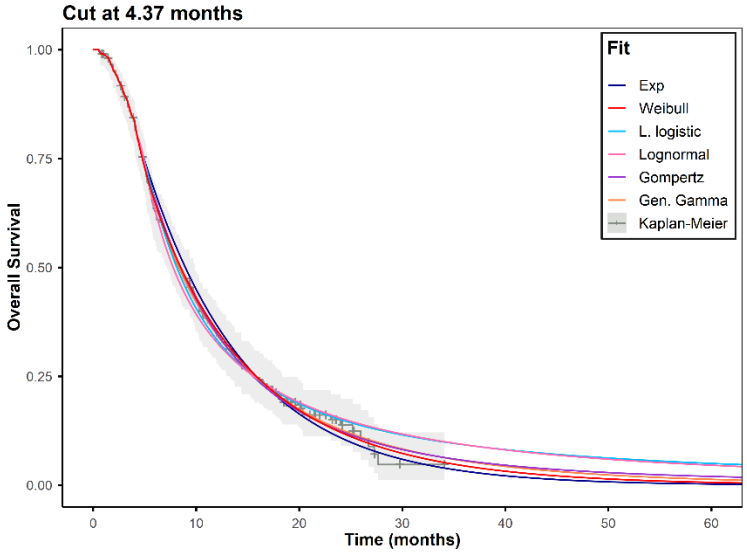


Figure 35. Semi-parametric model (4.37 month cut point) of time on treatment: ATTRACTION-3 nivolumab arm (cumulative hazard plot)

Taxane semi-parametric models (4.37 month cut point)



Fit Name	AIC	BIC	Parameters	Median (Months)	Mean (Months)
Exp.	864.70	867.77	lambda: 0.1010	8.95 (7.91, 10.01)	11.93 (10.49, 13.53)
Weibull	864.46	870.61	shape: 0.8983; scale: 9.7497	8.50 (7.37, 9.70)	12.22 (10.64, 14.10)
L. logistic	868.17	874.32	shape: 1.1972; scale: 5.8496	8.13 (7.10, 9.27)	28.71 (17.30, 81.19)
Lognormal	870.11	876.26	mu: 1.7091; sigma: 1.4693	7.77 (6.82, 8.79)	16.98 (13.20, 22.63)
Gompertz	864.70	870.85	shape: -0.0233; rate: 0.1177	8.49 (7.36, 9.80)	N/A (10.69, N/A)
Gen. Gamma	865.70	874.92	mu: 2.1339; sigma: 1.2171; Q: 0.7236	8.31 (7.32, 9.63)	12.69 (11.00, 16.47)

Figure 36. Semi-parametric model (4.37 month cut point) of overall survival: ATTRACTION-3 taxane arm (Kaplan-Meier plot)

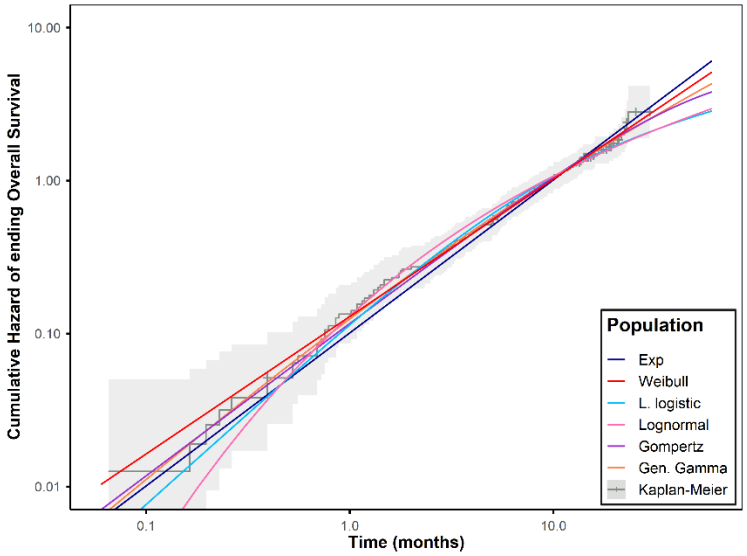
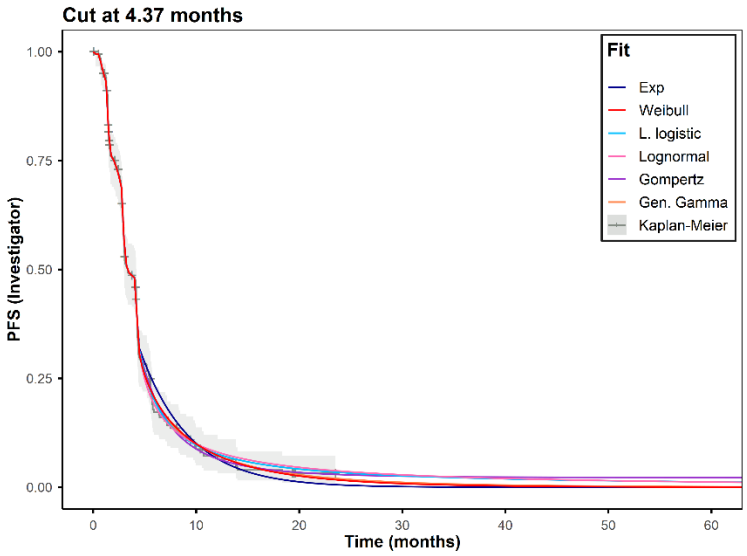


Figure 37. Semi-parametric model (4.37 month cut point) of overall survival: ATTRACTION-3 taxane arm (cumulative hazard plot)



Fit Name	AIC	BIC	Parameters	Median (Months)	Mean (Months)
Exp.	243.18	245.26	lambda: 0.2089	3.35 (2.99, 4.17)	4.75 (4.12, 5.58)
Weibull	237.63	241.79	shape: 0.7405; scale: 4.4573	3.35 (2.99, 4.17)	4.94 (4.18, 6.07)
L. logistic	238.12	242.27	shape: 1.0328; scale: 2.4317	3.35 (2.99, 4.17)	27.44 (4.90, 79.62)
Lognormal	241.22	245.37	mu: 0.8102; sigma: 1.7949	3.35 (2.99, 4.17)	6.87 (4.68, 12.16)
Gompertz	237.40	241.55	shape: -0.1196; rate: 0.3201	3.35 (2.99, 4.17)	N/A (3.90, N/A)
Gen. Gamma	239.31	245.55	mu: 1.3539; sigma: 1.4400; Q: 0.7770	3.35 (2.99, 4.17)	5.04 (4.22, 7.35)

Figure 38. Semi-parametric model (4.37 month cut point) of progression-free survival: ATTRACTION-3 taxane arm (Kaplan-Meier plot)

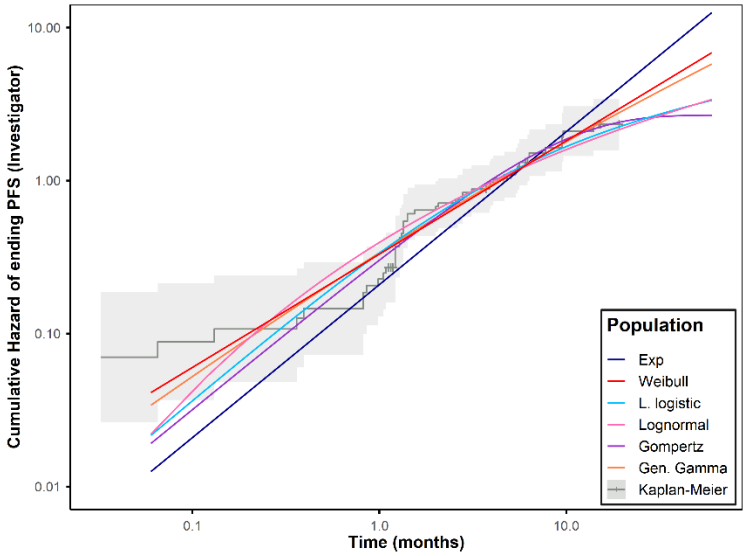
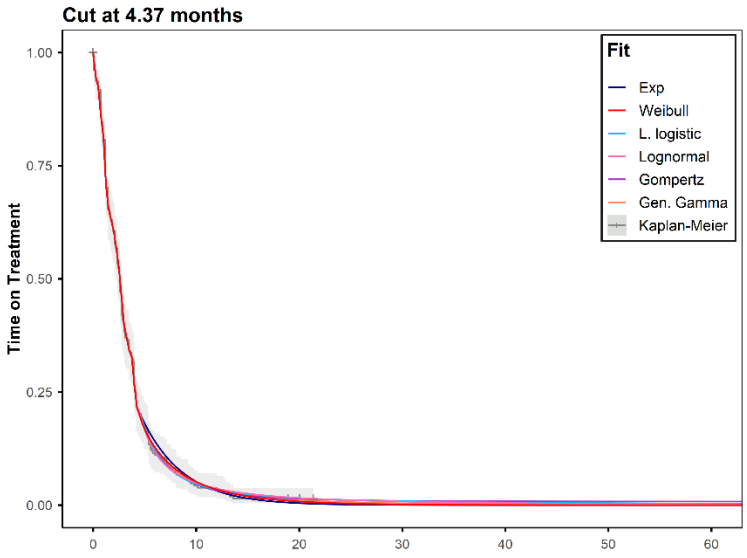


Figure 39. Semi-parametric model (4.37 month cut point) of progression-free survival: ATTRACTION-3 taxane arm (cumulative hazard plot)



Fit Name	AIC	BIC	Parameters	Median (Months)	Mean (Months)
Exp.	197.87	199.65	lambda: 0.2494	2.56 (2.33, 2.79)	3.39 (2.91, 3.92)
Weibull	197.25	200.82	shape: 0.8263; scale: 3.7080	2.56 (2.33, 2.79)	3.41 (2.92, 3.98)
L. logistic	195.19	198.76	shape: 1.2466; scale: 2.0582	2.56 (2.33, 2.79)	4.42 (3.19, 14.19)
Lognormal	196.61	200.18	mu: 0.6770; sigma: 1.4355	2.56 (2.33, 2.79)	3.70 (3.05, 4.82)
Gompertz	194.52	198.09	shape: -0.1113; rate: 0.3660	2.56 (2.33, 2.79)	N/A (2.86, N/A)
Gen. Gamma	197.40	202.76	mu: 0.9622; sigma: 1.3375; Q: 0.4334	2.56 (2.20, 2.79)	3.47 (2.97, 5.93)

Figure 40. Semi-parametric model (4.37 month cut point) of time on treatment: ATTRACTION-3 taxane arm (Kaplan-Meier plot)

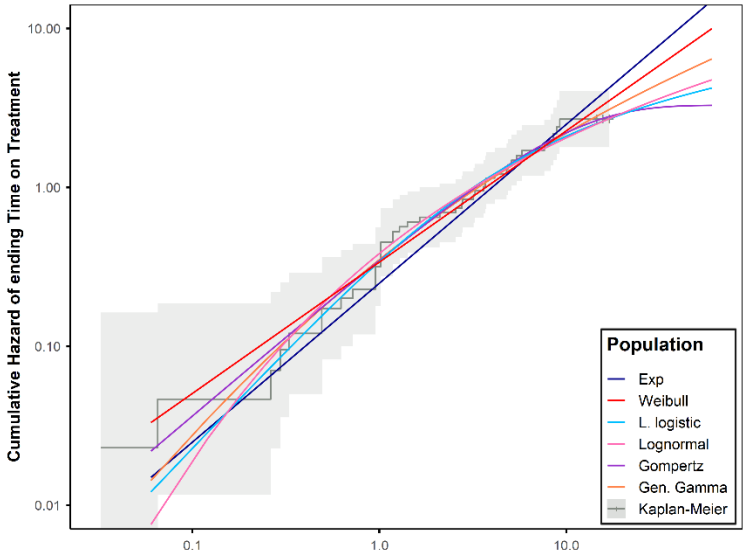


Figure 41. Semi-parametric model (4.37 month cut point) of time on treatment: ATTRACTION-3 taxane arm (cumulative hazard plot)

- **Semi-parametric approach using an alternative cut-point of approximately 5.5 months (i.e. a time point after the period over which the curves for each arm cross)**

Semi-parametric models (5.75 months)

A cut point of 5.5 months would relate to approximately 167 days, which would be within the time frame for assessment at week 24 (169 days). As outlined in the response to question B9, it is best practice to avoid dates where assessments are scheduled, which reflect a short period where several events occur and hence may bias the extrapolation. This approach has been consistently applied within the survival analyses provided within the company submission and those used within this response.

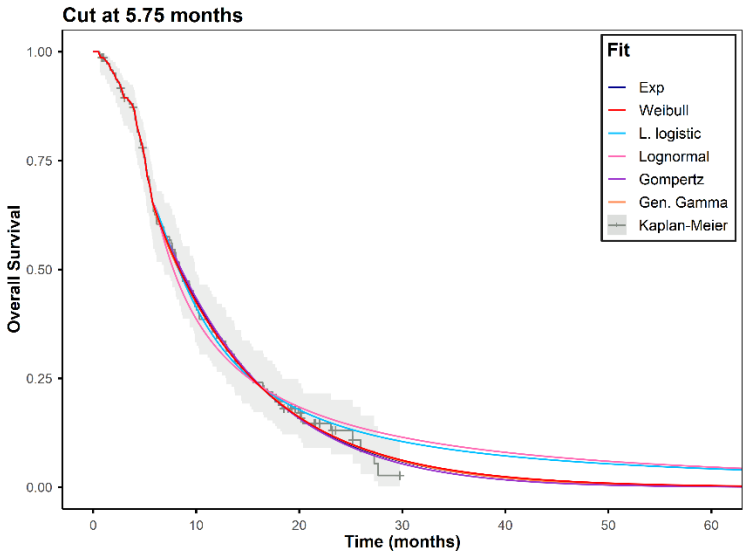
Thus, there are two potential options to addressing this response: provide a cut point before or after the assessment window. In line with the perceived request from the ERG, the decision was taken to ensure at least 5.5 months of Kaplan-Meier data were modelled directly. Hence, a cut-point of 5.75 months was applied, as this is outside the assessment window and directly models at least 5.5 months of Kaplan-Meier data.

As can be seen, the provided semi-parametric models typically have improved fit over the spline models; however, it is arguable whether there is improved fit over the semi-parametric approach where the cut point is provided at 2.99 months. Some models may be considered inappropriate, due to the poor fit to the available data or the implausible predicted mean time to event.

It should be noted that applying a cut point at 5.75 months, as opposed to 2.99 months, has a number of disadvantages. Although it allows for a greater number of people to demonstrate response in the nivolumab arm, it reduces the number of data points available to inform the extrapolation period. This particularly impacts the analysis of PFS, where median is 1.68 months in the nivolumab arm and 3.35 months in the taxane arm, and time on treatment, where the majority of patients have discontinued by 5.75 months.

Cost-effectiveness outcomes are provided in Table 6 and Table 7. As can be seen, the majority of plausible semi-parametric models provided comparable ICERs to the base case analysis, with few exceptions: nivolumab OS generalised gamma model (£60,946) and nivolumab time on treatment lognormal (£63,468). Overall, ICERs ranged from £41,488 to £63,468 for plausible semi-parametric models.

Nivolumab semi-parametric models (5.75 month cut point)



Fit Name	AIC	BIC	Parameters	Median (Months)	Mean (Months)
Exp.	485.11	487.61	lambda: 0.0994	8.47 (7.02, 9.87)	11.71 (10.14, 13.51)
Weibull	487.03	492.03	shape: 0.9725; scale: 10.0407	8.37 (6.95, 10.04)	11.77 (10.13, 13.79)
L. logistic	493.30	498.30	shape: 1.2573; scale: 6.4939	8.31 (7.03, 9.87)	22.85 (13.57, 105.91)
Lognormal	500.47	505.47	mu: 1.7869; sigma: 1.5015	7.79 (6.69, 9.24)	17.19 (12.15, 27.38)
Gompertz	487.04	492.04	shape: 0.0063; rate: 0.0955	8.56 (7.08, 10.20)	11.58 (10.08, 13.31)
Gen. Gamma	488.68	496.18	mu: 2.3639; sigma: 1.0000; Q: 1.1429	8.29 (6.83, 9.69)	11.68 (10.18, 13.71)

Figure 42. Semi-parametric model (5.75 month cut point) of overall survival: ATTRACTION-3 nivolumab arm (Kaplan-Meier plot)

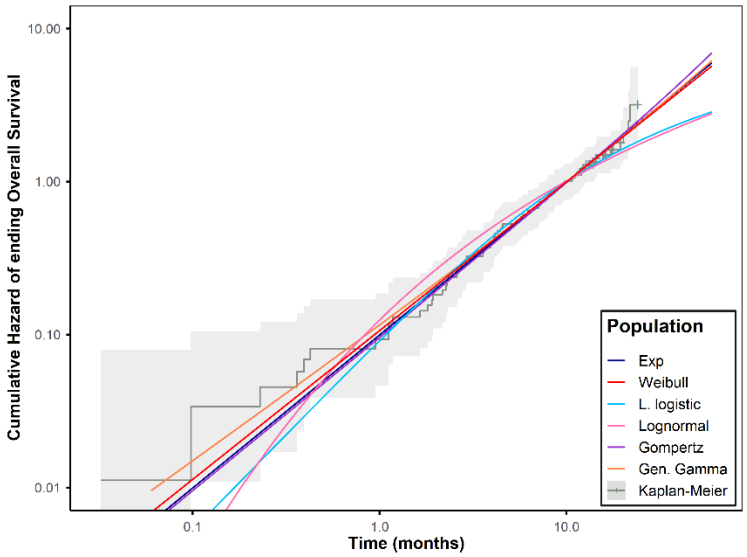
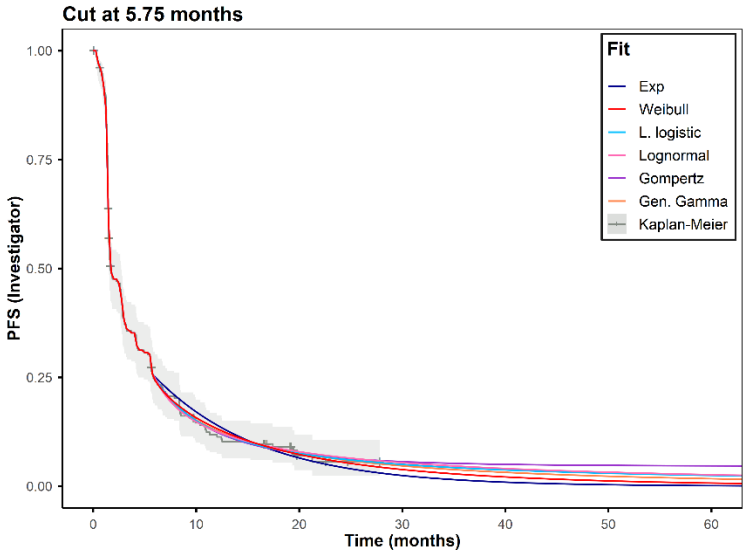


Figure 43. Semi-parametric model (5.75 month cut point) of overall survival: ATTRACTION-3 nivolumab arm (cumulative hazard plot)



Fit Name	AIC	BIC	Parameters	Median (Months)	Mean (Months)
Exp.	235.58	237.51	lambda: 0.0966	1.68 (1.54, 2.79)	5.66 (4.53, 7.11)
Weibull	234.18	238.04	shape: 0.7758; scale: 10.5831	1.68 (1.54, 2.79)	6.15 (4.67, 8.91)
L. logistic	232.86	236.73	shape: 1.0135; scale: 5.9694	1.68 (1.54, 2.79)	116.80 (6.45, 119.18)
Lognormal	233.70	237.57	mu: 1.7759; sigma: 1.7550	1.68 (1.54, 2.79)	10.09 (5.72, 29.29)
Gompertz	232.84	236.70	shape: -0.0885; rate: 0.1519	1.68 (1.54, 2.79)	N/A (4.32, N/A)
Gen. Gamma	235.15	240.95	mu: 2.0383; sigma: 1.5691; Q: 0.4170	1.68 (1.54, 2.79)	7.26 (4.84, 32.87)

Figure 44. Semi-parametric model (5.75 month cut point) of progression-free survival: ATTRACTION-3 nivolumab arm (Kaplan-Meier plot)

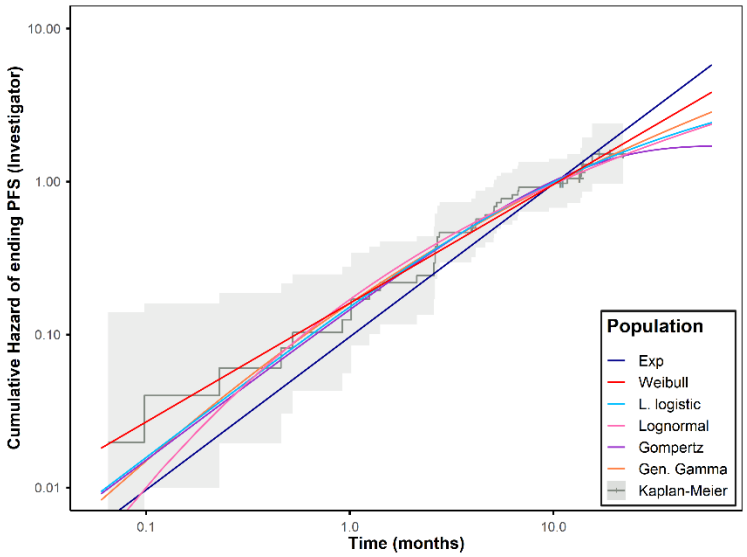
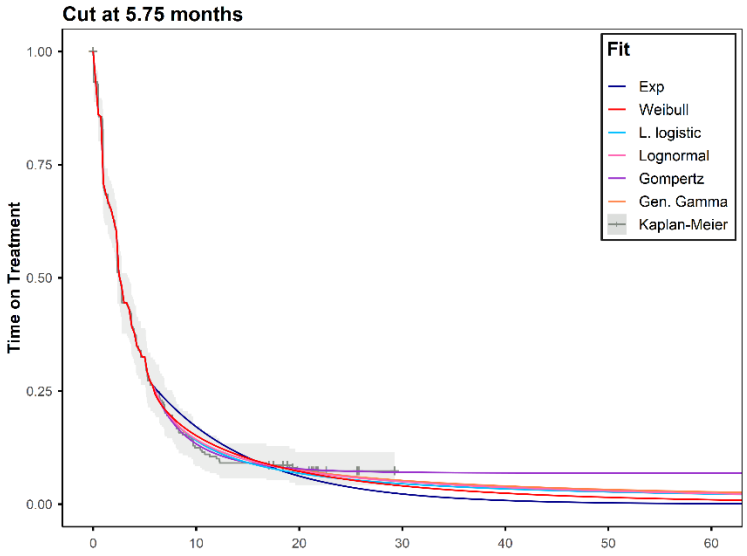


Figure 45. Semi-parametric model (5.75 month cut point) of progression-free survival: ATTRACTION-3 nivolumab arm (cumulative hazard plot)



Fit Name	AIC	BIC	Parameters	Median (Months)	Mean (Months)
Exp.	258.45	260.46	lambda: 0.1015	2.56 (2.33, 3.48)	5.65 (4.51, 7.20)
Weibull	252.52	256.53	shape: 0.7021; scale: 9.9060	2.56 (2.33, 3.48)	6.35 (4.75, 9.56)
L. logistic	246.76	250.78	shape: 0.9841; scale: 4.8897	2.56 (2.33, 3.48)	N/A (6.06, N/A)
Lognormal	246.98	250.99	mu: 1.6369; sigma: 1.7587	2.56 (2.33, 3.48)	9.41 (5.38, 28.56)
Gompertz	243.31	247.32	shape: -0.1669; rate: 0.2237	2.56 (2.33, 3.48)	N/A
Gen. Gamma	248.84	254.87	mu: 1.5159; sigma: 1.8034; Q: -0.1704	2.56 (2.33, 3.48)	11.99 (5.20, 128.88)

Figure 46. Semi-parametric model (5.75 month cut point) of time on treatment: ATTRACTION-3 nivolumab arm (Kaplan-Meier plot)

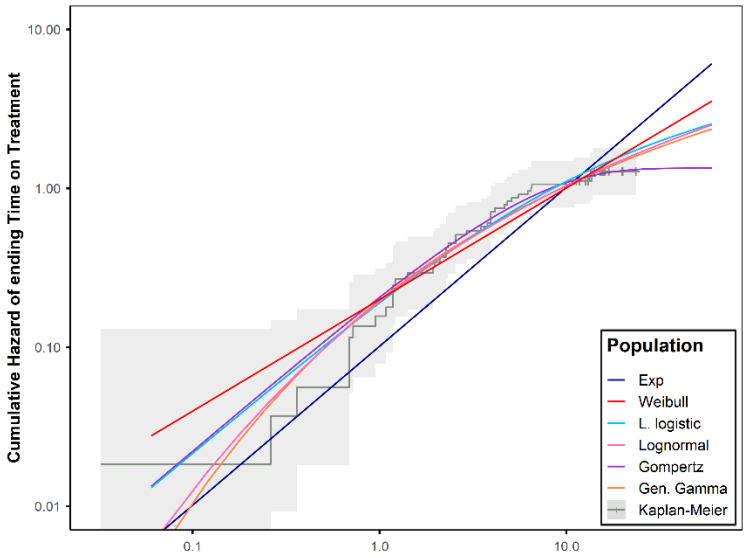
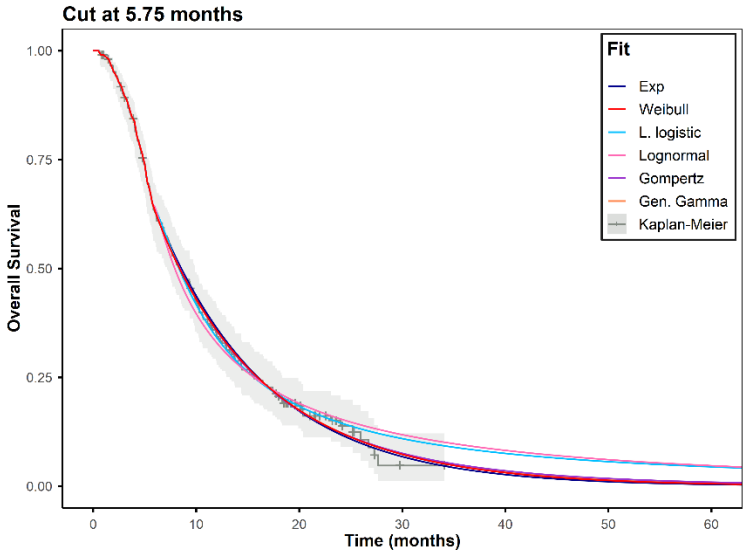


Figure 47. Semi-parametric model (5.75 month cut point) of time on treatment: ATTRACTION-3 nivolumab arm (cumulative hazard plot)

Taxane semi-parametric models (5.75 month cut point)



Fit Name	AIC	BIC	Parameters	Median (Months)	Mean (Months)
Exp.	690.43	693.29	lambda: 0.0930	8.56 (7.28, 9.75)	12.05 (10.56, 13.72)
Weibull	692.14	697.87	shape: 0.9560; scale: 10.7213	8.39 (7.10, 9.70)	12.18 (10.62, 14.07)
L. logistic	696.06	701.80	shape: 1.2528; scale: 6.7769	8.34 (7.19, 9.54)	23.70 (15.39, 79.12)
Lognormal	703.27	709.00	mu: 1.8579; sigma: 1.4681	7.92 (6.90, 9.04)	17.30 (13.06, 24.87)
Gompertz	692.20	697.94	shape: -0.0091; rate: 0.0988	8.43 (7.14, 9.86)	12.42 (10.60, 13.61)
Gen. Gamma	694.12	702.73	mu: 2.3552; sigma: 1.0607; Q: 0.9616	8.38 (7.13, 9.56)	12.22 (10.69, 14.85)

Figure 48. Semi-parametric model (5.75 month cut point) of overall survival: ATTRACTION-3 taxane arm (Kaplan-Meier plot)

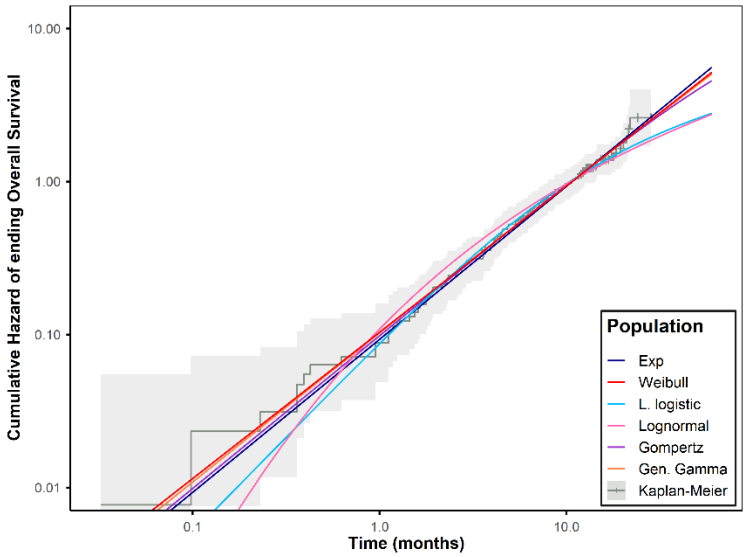
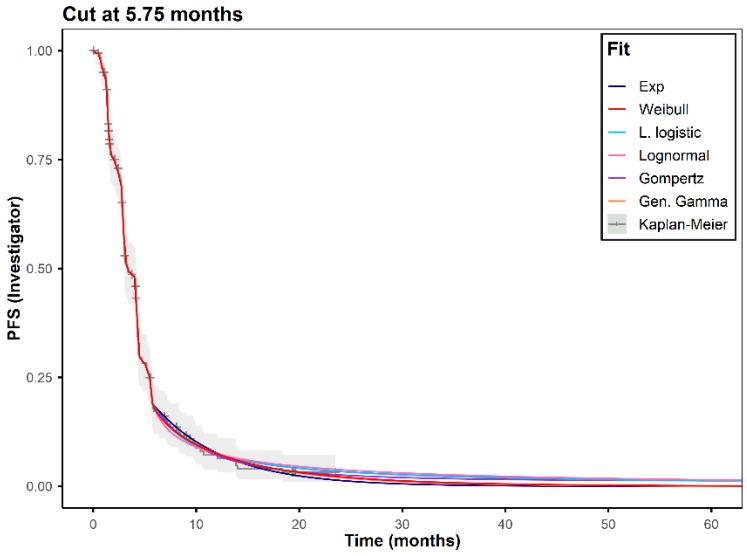


Figure 49. Semi-parametric model (5.75 month cut point) of overall survival: ATTRACTION-3 taxane arm (cumulative hazard plot)



Fit Name	AIC	BIC	Parameters	Median (Months)	Mean (Months)
Exp.	136.61	138.08	lambda: 0.1457	3.35 (2.99, 4.17)	4.87 (4.16, 5.81)
Weibull	136.58	139.51	shape: 0.7896; scale: 6.8302	3.35 (2.99, 4.17)	5.05 (4.20, 6.71)
L. logistic	138.00	140.93	shape: 1.0230; scale: 4.0903	3.35 (2.99, 4.17)	37.43 (4.64, 78.32)
Lognormal	140.44	143.37	mu: 1.3249; sigma: 1.8764	3.35 (2.99, 4.17)	7.72 (4.61, 25.88)
Gompertz	137.04	139.97	shape: -0.0718; rate: 0.1959	3.35 (2.99, 4.17)	N/A (4.05, N/A)
Gen. Gamma	138.52	142.92	mu: 1.9892; sigma: 1.2019; Q: 1.1373	3.35 (2.99, 4.17)	4.99 (4.20, 8.12)

Figure 50. Semi-parametric model (5.75 month cut point) of progression-free survival: ATTRACTION-3 taxane arm (Kaplan-Meier plot)

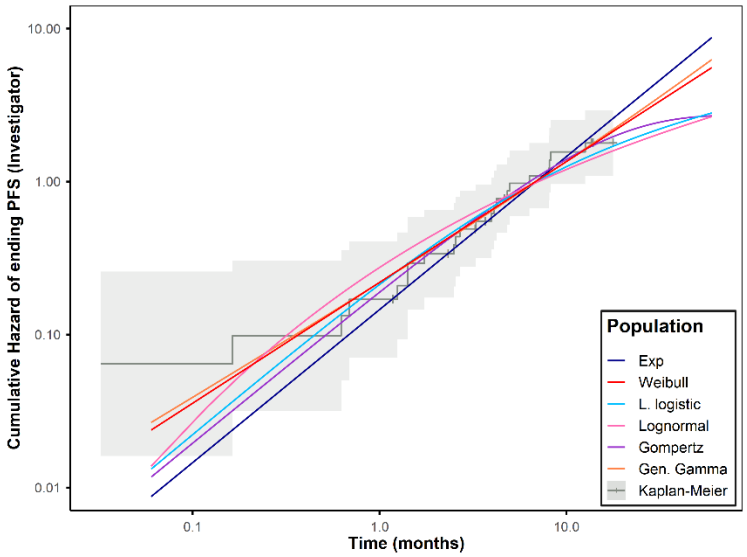
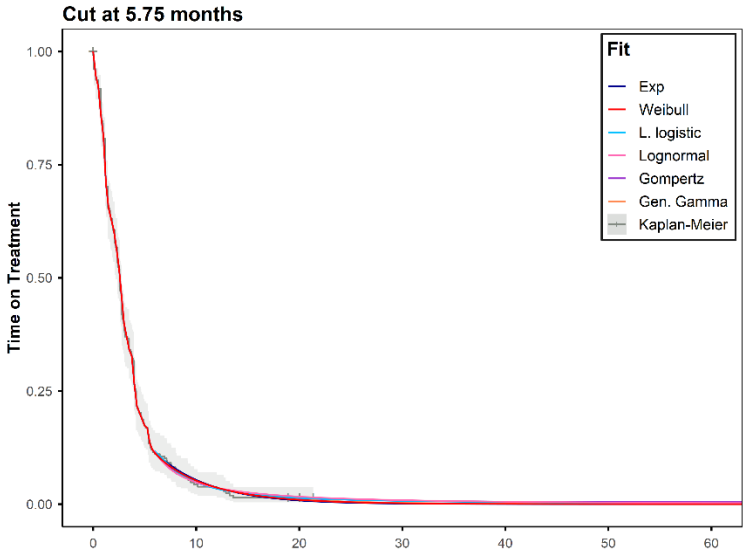


Figure 51. Semi-parametric model (5.75 month cut point) of progression-free survival: ATTRACTION-3 taxane arm (cumulative hazard plot)



Fit Name	AIC	BIC	Parameters	Median (Months)	Mean (Months)
Exp.	119.15	120.37	lambda: 0.1897	2.56 (2.33, 2.79)	3.40 (2.92, 3.96)
Weibull	120.89	123.32	shape: 0.9162; scale: 5.1662	2.56 (2.33, 2.79)	3.42 (2.92, 4.07)
L. logistic	119.89	122.33	shape: 1.3573; scale: 3.1709	2.56 (2.33, 2.79)	3.97 (3.02, 10.95)
Lognormal	123.19	125.63	mu: 1.0844; sigma: 1.4466	2.56 (2.33, 2.79)	3.78 (3.01, 6.05)
Gompertz	119.66	122.10	shape: -0.0772; rate: 0.2535	2.56 (2.33, 2.79)	N/A (2.90, N/A)
Gen. Gamma	122.59	126.25	mu: 1.5079; sigma: 1.1624; Q: 0.7357	2.56 (2.33, 2.79)	3.44 (2.99, 6.23)

Figure 52. Semi-parametric model (5.75 month cut point) of time on treatment: ATTRACTION-3 taxane arm (Kaplan-Meier plot)

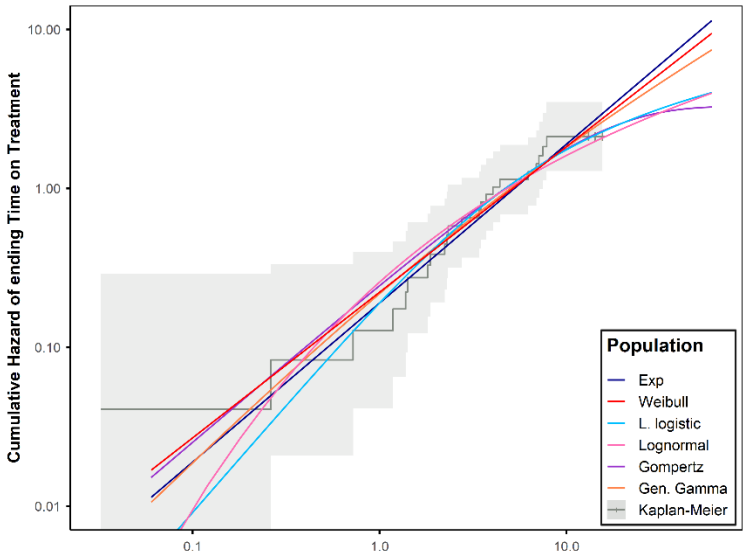


Figure 53. Semi-parametric model (5.75 month cut point) of time on treatment: ATTRACTION-3 taxane arm (cumulative hazard plot)

Time-permitting, the company may also wish to provide additional sensitivity analyses using some/all of the cut-points identified as "optimal" based on the ATTRACTION-3 trial data (i.e. [redacted], [redacted], or [redacted] months).

Please ensure the ability to reproduce the results of these sensitivity analyses is incorporated within the economic model file.

All analyses are provided as separate economic models.

Results: all scenario analyses

Table 6 and Table 7 provide the cost-effectiveness results from all extrapolations of survival estimates; all analyses are provided as separate economic models.

Across all analyses with plausible extrapolations (nivolumab and taxane), ICER estimates ranged from £41,448 (when a semi-parametric log-normal curve was applied after 5.75 months for nivolumab OS) to £69,068 (when a parametric log-logistic curve was applied for nivolumab time on treatment).

Varying taxane extrapolations generally had a lower impact on the ICER, with estimates ranging from £41,644 (when a parametric log-logistic curve was applied for taxane time

on treatment) to £58,782 (when a semi-parametric log-logistic curve was applied after 2.99 months for taxane OS).

Of the 78 scenario analyses conducted for nivolumab, 41 were based on extrapolations considered implausible, of which 22 exceeded the £50,000 willingness to pay threshold. Of the 37 scenario analyses based on extrapolations considered plausible, 29 analyses provided ICERs between £40,000/QALY and £55,000/QALY, indicating that plausible survival extrapolations are relatively consistent in terms of outcome. This is reflected in the ICER scatterplot provided in Figure 48.

While all requested survival analyses are provided, it should be noted that these are not equally relevant to decision-making. Parametric extrapolation of ATTRACTION-3 time to event data was undertaken with reference to the guidance from the NICE Decision Support Unit (DSU)[15] and Bagust and Beale (2014).[16] Assessment of extrapolation was based on several factors. Goodness-of-fit was evaluated using the Akaike and Bayesian Information Criteria (AIC and BIC, respectively). Additionally, the appropriateness of the parametric extrapolation was evaluated by visual inspection of the fit over the observed period and consideration of the log cumulative hazard plots. Further, the plausibility of the extrapolation was assessed through consideration of the long-term hazard profile and the extrapolated mean survival estimates. These factors should be accounted for when assessing different methods of extrapolation.

In addition, the plausibility of combinations of extrapolations should be assessed. As an example, it would be inappropriate to evaluate a longer OS profile in the absence of extending PFS and time on treatment. Similarly, it is inappropriate to evaluate a short OS profile in the absence of shortened PFS and time on treatment, as this could result in predicted time extrapolations overlapping.

With these factors in mind, the time to event extrapolations applied in the base case analysis provided in the company submission can be considered most appropriate for the following reasons:

- Rationale for semi-parametric approach versus spline modelling approach

- A semi-parametric approach avoids the inherent limitations associated with extrapolations using spline models.

- Rationale for timing of cut point:
    - Although a later cut point allows for a greater number of people to demonstrate response in the nivolumab arm, it reduces the number of data points available to inform the extrapolation period. This particularly impacts the analysis of PFS, where median is 1.68 months in the nivolumab arm and 3.35 months in the taxane arm, and time on treatment, where the majority of patients discontinue treatment early in the study.
    - The median time to response in the nivolumab arm was 2.6 months compared with 1.88 months in the taxane arm. A cut point at 2.99 months enables adequate time to allow the majority of patients to respond, and hence sufficiently reflect the change in hazard profile, while maximising the availability of data to inform the extrapolation period

- Rationale for individual extrapolations:
    - Every care was taken to adequately evaluate the extrapolation of ATTRACTION-3 time to event data, following guidance from the NICE Decision Support Unit (DSU)[15] and Bagust and Beale (2014).[16] Goodness of fit statistics (AIC and BIC) were minimised and visual inspection of observed Kaplan-Meier and log cumulative hazard plots was undertaken. Further, the plausibility of the extrapolation was assessed through consideration of the long-term hazard profile and the extrapolated mean survival estimates. These methods were extensively detailed in the company submission in order to justify the provided approach.

- Assessment of extrapolations as a combined model:
    - In addition to assessment of individual extrapolations of time to event data, it is vital to ensure that these are appropriate and plausible when

applied in a combined disease model. The extrapolations applied within the provided base case analysis provide a plausible and appropriate model of OSCC standard of care and the impact of nivolumab treatment.

Table 6. Nivolumab extrapolations

Scenario			Nivolumab		
			Inc. QALY	Inc. Cost (£)	ICER (£/ QALY)
PFS	Parametric	*Exponential*	*0.451*	*£20,842*	*£46,183*
		Generalised Gamma	*0.487*	*£20,842*	*£42,820*
		Gompertz	*0.506*	*£20,842*	*£41,151*
		Log-logistic	*0.449*	*£20,842*	*£46,416*
		Log-normal	*0.450*	*£20,842*	*£46,299*
		Weibull	*0.452*	*£20,842*	*£46,122*
	Semi-parametric with Kaplan-Meier to 2.99 months	Exponential	0.455	£20,842	£45,759
		Generalised Gamma	0.463	£20,842	£44,986
		Gompertz	*0.500*	*£20,842*	*£41,694*
		Log-logistic	*0.479*	*£20,842*	*£43,485*
		Log-normal	*0.480*	*£20,842*	*£43,423*
		Weibull	0.458	£20,842	£45,491
	Semi-parametric with Kaplan-Meier to 4.37 months	*Exponential*	*0.456*	*£20,842*	*£45,692*
		Generalised Gamma	*0.501*	*£20,842*	*£41,636*
		Gompertz	*0.502*	*£20,842*	*£41,499*
		Log-logistic	0.468	£20,842	£44,497
		Log-normal	0.464	£20,842	£44,880
		Weibull	0.457	£20,842	£45,571
	Semi-parametric with Kaplan-Meier to 5.75 months	Exponential	0.457	£20,842	£45,629
		Generalised Gamma	0.473	£20,842	£44,084
		Gompertz	*0.507*	*£20,842*	*£41,110*
		Log-logistic	*0.486*	*£20,842*	*£42,902*
		Log-normal	0.486	£20,842	£42,847
		Weibull	0.463	£20,842	£45,059
	Splines	*1 arbitrary knot*	*0.492*	*£20,842*	*£42,341*
		1 knot at 2.99 months	*0.492*	*£20,842*	*£42,388*
OS	Parametric	*Exponential*	*0.289*	*£18,871*	*£65,236*
		Generalised Gamma	*0.322*	*£19,268*	*£59,873*
		Gompertz	*0.263*	*£18,564*	*£70,649*
		Log-logistic	0.508	£21,426	£42,142
		Log-normal	0.455	£20,798	£45,736
		Weibull	*0.264*	*£18,595*	*£70,384*
	Semi-parametric with Kaplan-Meier to 2.99 months	*Exponential*	*0.286*	*£18,848*	*£65,796*
		Generalised Gamma	0.317	£19,214	£60,571
		Gompertz	*0.267*	*£18,620*	*£69,743*
		Log-logistic	0.458	£20,842	£45,491
		Log-normal	0.435	£20,577	£47,269
		Weibull	*0.260*	*£18,543*	*£71,343*
	Semi-parametric with Kaplan-Meier to 4.37 months	*Exponential*	*0.285*	*£18,836*	*£66,054*
		Generalised Gamma	0.300	£19,011	£63,418
		Gompertz	*0.273*	*£18,697*	*£68,400*
		Log-logistic	0.457	£20,831	£45,558
		Log-normal	*0.461*	*£20,876*	*£45,262*
		Weibull	*0.262*	*£18,572*	*£70,780*
	Semi-parametric with	*Exponential*	*0.276*	*£18,733*	*£67,890*
		Generalised Gamma	0.315	£19,187	£60,946
		Gompertz	*0.349*	*£19,585*	*£56,065*

Scenario			Nivolumab		
			Inc. QALY	Inc. Cost (£)	ICER (£/ QALY)
	Kaplan-Meier to 5.75 months	*Log-logistic*	*0.515*	*£21,495*	*£41,752*
		Log-normal	0.519	£21,545	£41,488
		Weibull	*0.281*	*£18,788*	*£66,944*
	Splines	*1 arbitrary knot*	*0.286*	*£18,850*	*£65,938*
		1 knot at 2.99 months	*0.285*	*£18,844*	*£66,029*
ToT	Parametric	Exponential	0.458	£19,110	£41,709
		Generalised Gamma	0.458	£20,842	£45,491
		Gompertz	*0.458*	*£68,899*	*£150,383*
		Log-logistic	0.458	£31,644	£69,068
		Log-normal	0.458	£30,004	£65,488
		Weibull	0.458	£19,833	£43,289
	Semi-parametric with Kaplan-Meier to 2.99 months	Exponential	0.458	£19,923	£43,484
		Generalised Gamma	*0.458*	*£33,277*	*£72,633*
		Gompertz	*0.458*	*£73,261*	*£159,903*
		Log-logistic	0.458	£23,931	£52,232
		Log-normal	0.458	£22,590	£49,307
		Weibull	0.458	£20,633	£45,035
	Semi-parametric with Kaplan-Meier to 4.37 months	Exponential	0.458	£20,127	£43,930
		Generalised Gamma	0.458	£26,975	£58,878
		Gompertz	*0.458*	*£79,917*	*£174,432*
		Log-logistic	*0.458*	*£28,744*	*£62,737*
		Log-normal	0.458	£28,338	£61,853
		Weibull	0.458	£21,951	£47,912
	Semi-parametric with Kaplan-Meier to 5.75 months	Exponential	0.458	£20,576	£44,910
		Generalised Gamma	*0.458*	*£31,327*	*£68,375*
		Gompertz	-	-	-
		Log-logistic	*0.458*	*£29,433*	*£64,241*
		Log-normal	0.458	£29,079	£63,468
		Weibull	0.458	£22,921	£50,029
	Splines	1 arbitrary knot	0.458	£20,646	£45,064
		1 knot at 2.99 months	0.458	£20,802	£45,404

ICER: incremental cost-effectiveness ratio; Inc: incremental; OS: overall survival; PFS: progression-free survival; QALY: quality-adjusted life year; ToT: time on treatment
Grey italics denotes highly implausible extrapolations, defined as extrapolations that exceed the 95% confidence intervals of the Kaplan-Meier data, provides mean survival that cannot be considered plausible or extrapolations that did not converge.

Table 7. All taxane extrapolations

Scenario			Taxane		
			Inc. QALY	Inc. Cost (£)	ICER (£/ QALY)
PFS	Parametric	*Exponential*	*0.458*	*£20,842*	*£45,544*
		Generalised Gamma	0.458	£20,842	£45,503
		Gompertz	*0.458*	*£20,842*	*£45,476*
		Log-logistic	0.458	£20,842	£45,482
		Log-normal	0.459	£20,842	£45,380
		Weibull	*0.460*	*£20,842*	*£45,354*
	Semi-parametric with Kaplan-Meier to 2.99 months	*Exponential*	*0.460*	*£20,842*	*£45,352*
		Generalised Gamma	0.457	£20,842	£45,606
		Gompertz	*0.450*	*£20,842*	*£46,304*
		Log-logistic	0.447	£20,842	£46,584
		Log-normal	0.446	£20,842	£46,747
		Weibull	0.458	£20,842	£45,491
	Semi-parametric with Kaplan-Meier to 4.37 months	*Exponential*	*0.459*	*£20,842*	*£45,445*
		Generalised Gamma	0.455	£20,842	£45,837
		Gompertz	*0.449*	*£20,842*	*£46,376*
		Log-logistic	*0.448*	*£20,842*	*£46,561*
		Log-normal	*0.447*	*£20,842*	*£46,662*
		Weibull	0.456	£20,842	£45,708
	Semi-parametric with Kaplan-Meier to 5.75 months	*Exponential*	*0.457*	*£20,842*	*£45,604*
		Generalised Gamma	0.455	£20,842	£45,781
		Gompertz	*0.451*	*£20,842*	*£46,227*
		Log-logistic	*0.448*	*£20,842*	*£46,508*
		Log-normal	0.447	£20,842	£46,614
		Weibull	0.455	£20,842	£45,852
	Splines	*1 arbitrary knot*	*0.456*	*£20,842*	*£45,727*
		1 knot at 2.99 months	*0.456*	*£20,842*	*£45,724*
OS	Parametric	*Exponential*	*0.446*	*£20,761*	*£46,509*
		Generalised Gamma	0.345	£19,320	£55,989
		Gompertz	*0.473*	*£21,091*	*£44,595*
		Log-logistic	0.411	£20,210	£49,165
		Log-normal	0.432	£20,492	£47,434
		Weibull	*0.469*	*£21,000*	*£44,756*
	Semi-parametric with Kaplan-Meier to 2.99 months	Exponential	0.458	£20,842	£45,491
		Generalised Gamma	0.434	£20,515	£47,306
		Gompertz	*0.434*	*£20,524*	*£47,271*
		Log-logistic	0.324	£19,019	£58,782
		Log-normal	0.352	£19,415	£55,130
		Weibull	0.457	£20,828	£45,580
	Semi-parametric with Kaplan-Meier to	Exponential	0.459	£20,852	£45,408
		Generalised Gamma	0.432	£20,490	£47,456
		Gompertz	*0.395*	*£19,993*	*£50,556*
		Log-logistic	*0.284*	*£18,475*	*£65,164*

Scenario			Taxane		
			Inc. QALY	Inc. Cost (£)	ICER (£/ QALY)
	4.37 months	*Log-normal*	*0.318*	*£18,956*	*£59,539*
		Weibull	0.449	£20,716	£46,185
	Semi-parametric with Kaplan-Meier to 5.75 months	Exponential	0.455	£20,805	£45,716
		Generalised Gamma	0.449	£20,720	£46,182
		Gompertz	-	-	-
		Log-logistic	*0.307*	*£18,798*	*£66,175*
		Log-normal	*0.311*	*£18,859*	*£60,565*
		Weibull	0.450	£20,745	£46,050
	Splines	1 arbitrary knot	0.438	£20,577	£46,956
		1 knot at 2.99 months	0.445	£20,673	£46,436
ToT	Parametric	Exponential	0.458	£20,842	£45,491
		Generalised Gamma	0.458	£20,781	£45,357
		Gompertz	0.458	£20,824	£45,451
		Log-logistic	0.458	£19,080	£41,644
		Log-normal	0.458	£19,433	£42,415
		Weibull	0.458	£20,818	£45,439
	Semi-parametric with Kaplan-Meier to 2.99 months	Exponential	0.458	£20,663	£45,101
		Generalised Gamma	0.458	£20,603	£44,969
		Gompertz	*0.458*	*£17,964*	*£39,210*
		Log-logistic	0.458	£19,935	£43,510
		Log-normal	0.458	£20,123	£43,922
		Weibull	0.458	£20,651	£45,075
	Semi-parametric with Kaplan-Meier to 4.37 months	Exponential	0.458	£20,631	£45,029
		Generalised Gamma	0.458	£20,494	£44,732
		Gompertz	*0.458*	*£17,657*	*£38,539*
		Log-logistic	0.458	£19,890	£43,412
		Log-normal	0.458	£20,162	£44,006
		Weibull	0.458	£20,598	£44,957
	Semi-parametric with Kaplan-Meier to 5.75 months	Exponential	0.458	£20,604	£44,971
		Generalised Gamma	0.458	£20,548	£44,850
		Gompertz	*0.458*	*£18,863*	*£41,172*
		Log-logistic	0.458	£20,095	£43,859
		Log-normal	0.458	£20,063	£43,790
		Weibull	0.458	£20,583	£44,925
	Splines	1 arbitrary knot	0.458	£20,817	£45,436
		1 knot at 2.99 months	0.458	£20,816	£45,434

ICER: incremental cost-effectiveness ratio; Inc: incremental; OS: overall survival; PFS: progression-free survival; QALY: quality-adjusted life year; ToT: time on treatment
Grey italics denotes highly implausible extrapolations, defined as extrapolations that exceed the 95% confidence intervals of the Kaplan-Meier data, provides mean survival that cannot be considered plausible or extrapolations that did not converge.

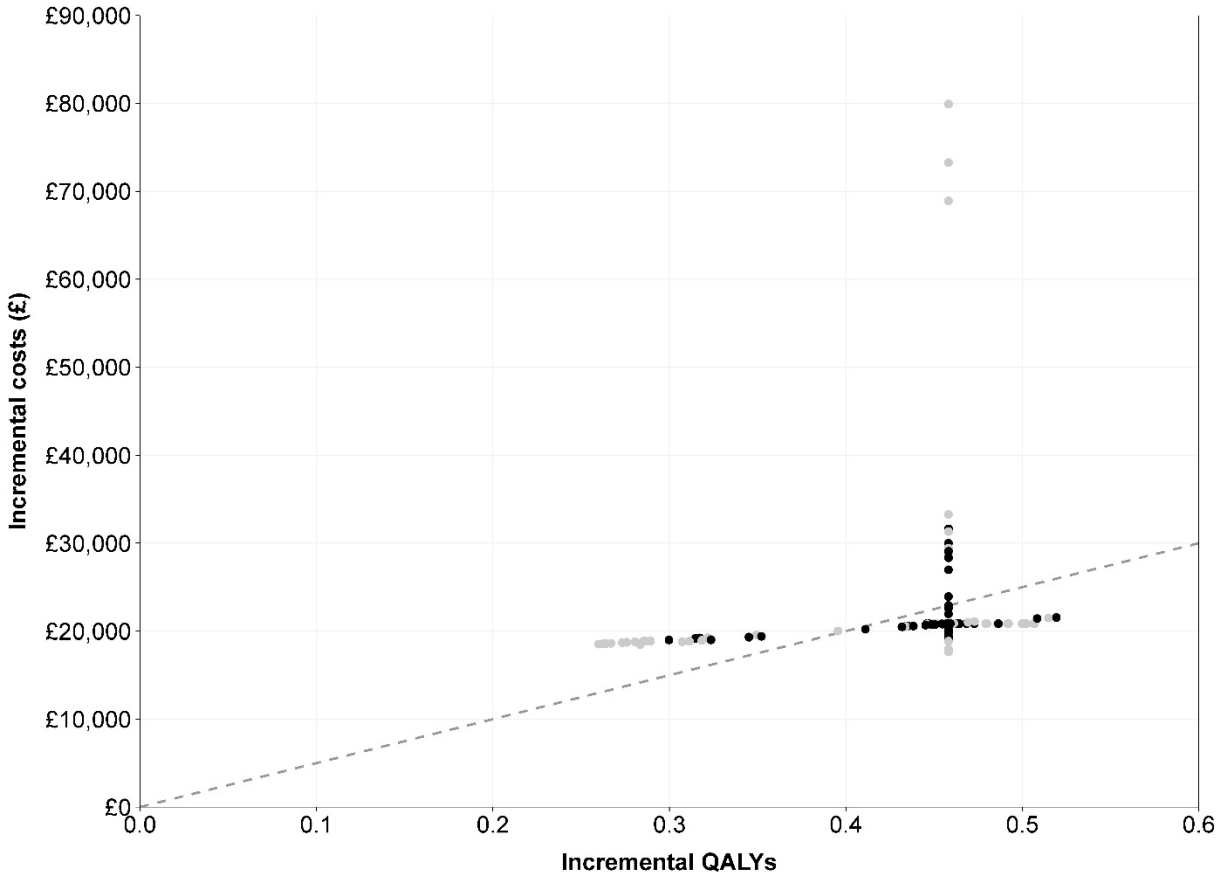


Figure 54. Cost-effectiveness scatter of all nivolumab and taxane extrapolations

Grey dots represent extrapolations deemed implausible.

B11. PRIORITY QUESTION: The company's model provides only one set of utility values to inform results. Please can the company provide sensitivity analysis using utility values from any of the following sources:

- **The systematic literature review conducted regarding health effects**
- **Analyses previously considered based on the ATTRACTION-3 trial data using alternative analytical approaches**
- **Previous NICE technology appraisals conducted in similar populations undergoing similar treatment(s)**

Please ensure the ability to reproduce the results of these sensitivity analyses is incorporated within the economic model file, and a description of any associated limitations is provided

Alternative utility inputs

Only one study from the SLR provided evidence that could be considered appropriate for inclusion in the economic model. Doherty (2018)[10] did not report values by progression status, limiting the extent to which it could be applied in the economic model. However, the value for patients receiving palliative chemotherapy (0.74) was applied across all health states and therapies.

Alternative analytical approaches using data from ATTRACTION-3 is provided in response to Question B12.

There are no NICE HTAs assessing therapies for the treatment of oesophageal cancer. However, a previous NICE HTA (TA378) assessed treatment of previously treated gastric cancer. These utility data (0.737 in pre-progression and 0.587 for post-progression[17]) are applied across treatments.

It should be noted that both these approaches can be considered highly conservative and does not reflect the benefit of nivolumab. Utility values for the taxane arm of ATTRACTION-3 are comparable with those from TA378 ([redacted] versus 0.737[17] in pre-progression and [redacted] versus 0.587[17] for post-progression), which can be considered a validation of the methodology and the output from the study. By contrast, the utility value in the nivolumab arm is higher for both the pre-progression and post-progression state. This improvement can be expected due to the novel mechanism of action that may account for this improvement. In contrast to common oncology therapies, nivolumab enables the patient's own immune system to directly destroy cancer cells (in the same way that it would any other "foreign" cell), resulting in destruction of the tumour through pre-existing, intrinsic processes. Nivolumab is associated with several benefits that impact directly on patient quality of life even when excluding improved PFS and OS, particularly a tolerable AE profile; 65.6% of patients in the nivolumab arm reported a drug-related AE (grade 3-5: 18.2%) versus 95.2% for patients receiving paclitaxel or docetaxel (grade 3-5: 64.0%). Further, it should be noted that quality of life outcomes during ATTRACTION-3 remained relatively stable in the nivolumab arm, as determined by EQ-5D and EQ-VAS; however, patients receiving taxanes frequently

reported worsened quality of life outcomes during the trial period. In addition, the utility values observed during ATTRACTION-3 are broadly equivalent to utility values observed from other nivolumab indications,[18-23] indicating that this utility gain may be due to the novel mechanism of action for nivolumab. Thus, the quality of life data derived from patients during ATTRACTION-3 reflects the expected benefits of nivolumab over taxanes, including the potential for immune system stimulation following progression.

In line with the above rationale, assuming equivalent quality of life outcomes across treatments does not reflect the documented benefits of nivolumab.

Table 8. Alternative source of utility values

State	Nivolumab	Taxane
Systematic literature review (Doherty et al. [10])		
Pre-progression	0.74	0.74
Post-Progression	0.74	0.74
Analyses based on ATTRACTION-3 trial data[1] using alternative analytical approaches		
Provided in response to Question B12		
Previous NICE technology assessment (TA378[24])		
Pre-progression	0.737	0.737
Post-Progression	0.587	0.587

Alternative utility analysis results

Results from the alternative utility analyses are detailed in Table 9. Application of utility values sourced from the systematic literature review and previous NICE technology assessment resulted in ICER estimates of £52,500 per QALY and £63,982 per QALY, respectively, which signals an increase in the estimate from the base case (£45,491 per QALY). However, these outputs should be considered in the context of the highly conservative nature of this analysis.

Table 9. Impact of applying alterative utility sources

Outcome	Systematic literature review		Previous NICE technology assessment	
	Nivolumab	Taxane	Nivolumab	Taxane
Total QALYs	1.114	0.717	0.955	0.629
Total costs	£47,629	£26,786	£47,629	£26,786
Incremental QALYs	-	0.397	-	0.326
Incremental costs	-	£20,842	-	£20,842
ICER (£/QALY)	**-**	**£52,500**	**-**	**£63,982**
ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year				

B12. PRIORITY QUESTION: Please can the company provide sensitivity analysis using a mixed-effects regression model for the estimation of utility values (including fixed covariates for progression status and treatment arm, a variable interacting treatment arm with progression status, and a random effect for subject)? This approach was used in the previous NICE assessment of nivolumab in previously treated advanced renal cell carcinoma (TA417).

In addition, please can the company provide sensitivity analysis removing the use of imputation methods to derive utility values?

Please ensure the ability to reproduce the results of these sensitivity analyses are incorporated within the economic model file

The company fitted a mixed-effects regression on the untransformed utility scores (Dolan TTO tariff) as a complete case analysis. The model was specified as:

dolan.index ~ arm * progressed + (1|usubjid)

providing a fixed intercept, offset for the placebo arm, offset for being confirmed progressed, and offset for being in the placebo arm and confirmed progressed. There was a random offset for subject.

The dataset was as prepared for the submission analysis, i.e. it had been regularised to 12 week intervals to prevent over or under representation of patients in pre/post progression health states due to the differing frequencies of data collection on and off therapy, which did not always coincide with the progression states. Incomplete or

missing observations were removed, and observations where the progression status was indeterminate due to it occurring after the patient's PFS censoring time (due to commencement of following drug or final imaging time) were also removed. The baseline observations were included, as these health state utilities are aimed to be representative of the mean utility in health state; the purpose of the model is not to establish clinical benefit.

Nevertheless, the company does not feel this analysis is appropriate for cost-effectiveness analysis, as mixed models provide a mean per subject effect, and are not representative of the marginal value of utility in a health state over time, where subjects have varying time in state conditional upon their utility. Direct representation of the data collected provides a truer estimate of the mean utility in state over all time as those patients with worse utility spend less time in state.

The complete-case computation of heath state utility was also performed upon this dataset. The prais-winsten standard error correction was used, but this is with the caveat that it is not correct for the small number of intermittent missing data patterns in the dataset.

Results from the analysis is detailed in Table 10, where application of utility values using a mixed effect model and assuming no imputation resulted in ICER estimates of £47,982 per QALY and £44,672 per QALY, respectively, signalling a small impact on ICER estimates in comparison to the base case estimate (£45,491 per QALY).

Table 10. Impact of applying utilities sourced from alternative analytical methods

Outcome	Mixed effect		No imputation	
	Nivolumab	**Taxane**	**Nivolumab**	**Taxane**
Total QALYs	1.059	0.625	1.126	0.660
Total costs	£47,629	£26,786	£47,629	£26,786
Incremental QALYs	-	0.434	-	0.467
Incremental costs	-	£20,842	-	£20,842
ICER (£/QALY)	**-**	**£47,982**	**-**	**£44,672**
ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year				

B13. Throughout the CS, it is stated that paclitaxel treatment involved intravenous administration of 100mg/m^2 weekly for 6 weeks followed by a 2-week

drug holiday. However, the published ATTRACTION-3 manuscript states that paclitaxel was administered weekly for 6 weeks "followed by 1 week off". Please can the company clarify which of these regimens was used in the ATTRACTION-3 trial, which is expected to resemble UK practice, and if different what the potential impacts of this are on the clinical- and cost effectiveness outcomes for the taxanes group?

The protocol refers to paclitaxel being administered once per week for 6 weeks, followed by a 2-week rest (time interval from the last dose of paclitaxel given in the previous cycle to the first dose of paclitaxel given in the next cycle), which defines 1 treatment cycle. This protocol-specified dosing regimen is referred to as paclitaxel administered weekly for 6 weeks followed by 1 week off.

B14. Medical resource use estimates appear to be based on information presented in the previous NICE TA378 (ramucirumab for treating advanced gastric cancer or gastro–oesophageal junction adenocarcinoma previously treated with chemotherapy) which was subsequently validated by a clinical advisory board arranged by the company. For clarity, please can the company confirm which edits were made (e.g. removal/ addition of specific items, or adjustments to assumptions), alongside the reason(s) for these?

Composition of BSC and disease management were derived from an expert survey, not from TA378. However, TA378 was used as a source of validation. The rationale for this approach is provided below.

Within the final scope set out by NICE, BSC is specified as a comparator, with composition as including, but not limited to, anti-emetics, blood transfusion and oesophageal stents.[25]

The composition of BSC and disease management are available from TA378[24], where BSC composition is derived from a company-conducted treatment pattern study in first-line patients and disease management costs are derived from expert elicitation. However, this study did not include anti-emetics and did not report on use of stents.

Due to the uncertainty around use of anti-emetics and stents, the composition of BSC from TA378[24] was presented at a clinical advisory board meeting for the purposes of validating BSC management in the UK. However, clinical experts noted that other forms of BSC were notably omitted from the list, particularly oesophageal stents and ascites drainage.

As clinicians agreed with the NICE scope, a clinician survey was initiated, where the survey was completed by practising oncologists and nurses in the UK based on their experience in treating UK-based gastric and GOJ cancer patients.[25] Hence, composition of BSC and disease management were derived from this survey, not from TA378.

Although the survey was used as the source for resource use composition, TA378 was used as a source for validation. As can be seen in Table 11, composition of BSC and disease management were broadly comparable between the two sources. BSC as detailed in the company submission also includes additional pain control elements, control of gastro-intestinal bleeding, oesophageal stents and ascites drainage. Disease management includes palliative care nursing costs. Further, frequency of resource use and administration assumptions were broadly comparable between TA378 and the company submission.

Table 11. Comparison of BSC and disease management composition in company submission versus TA378

	Best supportive care	Disease management
TA378	• Pain control (morphine only) • Distress management (as cognitive behavioural therapy) • Blood transfusions • Radiation therapy	CT scan Blood count Renal function test Hepatic function test Consultation visit Hospitalisations
Company submission	• Pain control (nerve blocks and several medications, described in Table 66 of company submission) • Distress management (as cognitive behavioural therapy) • Blood transfusions • Radiotherapy • Procedures and drugs to control GI bleeds • Oesophageal stents • Ascites drainage	CT scan Blood count Renal function test Hepatic function test Consultation visit Palliative care nurse Hospitalisations

B15. In Section B.3.8.2, results are presented for alternative comparators, including specific taxanes regimens (i.e. docetaxel and paclitaxel). Throughout the CS it is stated that paclitaxel is expected to be more efficacious than docetaxel, including data provided within Table S1 of the ATTRACTION-3 manuscript which suggests that median OS and PFS were higher for paclitaxel than docetaxel. However, this is not shown in terms of the incremental life-years associated with each taxane compared to nivolumab. Please can the company clarify why the scenario analysis results are misaligned with the clinical data and expectation concerning the differences between treatment with docetaxel and paclitaxel?

As outlined in the survival analysis provided in Appendix M of the company submission, there are several points where the survival curves for docetaxel and paclitaxel crossover, as can be expected given the reduced patient numbers and the degree of similarity between the therapies.

Figure 49 and Figure 50 provide the Kaplan-Meier data for PFS and OS, respectively, from ATTRACTION-3, while Table 12 and Table 13 provide survival estimates at landmark times.

As can be seen, paclitaxel initially provides [redacted] PFS outcomes versus docetaxel (PFS: [redacted]). By nine months, outcomes are [redacted] in the paclitaxel arm, but this is [redacted] by 15 months.

Similarly, there are several crossovers in the OS curve, although this is not reflected in survival at landmark times. Of particular note, one of these crossovers impacts on median OS time, providing the observation that median OS is marginally [redacted] in the paclitaxel arm, although a more robust interpretation is that the two are comparable throughout.

[redacted]*Figure 55: Kaplan-Meier data for investigator-assessed PFS in ATTRACTION-3[1]*

Table 12: Progression-free survival rate at selected landmark times in ATTRACTION-3[1]

PFS at landmark times (%)	Nivolumab (n = 210)	Control group		
		Total (n = 209)	Docetaxel (n = 65)	Paclitaxel (n =144)
3 months	[redacted]	[redacted]	[redacted]	[redacted]
6 months	[redacted]	[redacted]	[redacted]	[redacted]
9 months	[redacted]	[redacted]	[redacted]	[redacted]
12 months	[redacted]	[redacted]	[redacted]	[redacted]
15 months	[redacted]	[redacted]	[redacted]	[redacted]
18 months	[redacted]	[redacted]	[redacted]	[redacted]
21 months	[redacted]	[redacted]	[redacted]	[redacted]

[redacted]

Figure 56: Kaplan-Meier data for OS in ATTRACTION-3[1]

Table 13: Overall survival rate at selected landmark times in ATTRACTION-3[1]

Time point	Nivolumab (n = 210)	Control group		
		Total (n = 209)	Docetaxel (n = 65)	Paclitaxel (n =144)
6 months	[redacted]	[redacted]	[redacted]	[redacted]
9 months	[redacted]	[redacted]	[redacted]	[redacted]
12 months	[redacted]	[redacted]	[redacted]	[redacted]
15 months	[redacted]	[redacted]	[redacted]	[redacted]
18 months	[redacted]	[redacted]	[redacted]	[redacted]
21 months	[redacted]	[redacted]	[redacted]	[redacted]
24 months	[redacted]	[redacted]	[redacted]	[redacted]
27 months	[redacted]	[redacted]	[redacted]	[redacted]
30 months	[redacted]	[redacted]	[redacted]	[redacted]

Section C: Textual clarification and additional points

C1. Please clarify the estimates for PFS in Table 13 (Document B) and estimates by race in Table 16, as there appear to be data entry errors?

- PFS for paclitaxel at 9 months was [redacted] and at 12 months it was [redacted]

- With regards to Table 16, in the subgroup analysis for PFS, [redacted] White patients received nivolumab and [redacted] events occurred among those. In the subgroup analysis for ORR, [redacted] Asian patients received nivolumab and [redacted] Asian patients received chemotherapy.

C2. Please clarify the estimates in Table 5 (Document A), as there appear to be data entry errors?

- The columns for PFS were moved one column to the right and should say median PFS for patients receiving nivolumab was 1.68 months, 3.35 months for patients receiving chemotherapy, 3.02 months for patients receiving docetaxel and 4.11 months for patients receiving paclitaxel.

- Similarly, the columns for the hazard ratio got moved to the left and should say 1.08 for chemotherapy overall, 0.97 for to docetaxel and 1.15 for paclitaxel.

C3. It is stated that "there was a significant PFS benefit for nivolumab-treated patients at all time points from three months through to 21 months". However, the curves appear to cross at about 5 months. Please clarify your interpretation of the PFS benefit.

This is a typographical error and should say "there was a significant PFS benefit for nivolumab-treated patients at all time points from six months through to 21 months.

C4. For the avoidance of doubt, please can the company confirm that the EQ-5D-3L questionnaire was used in ATTRACTION-3 as opposed to the EQ-5D-5L questionnaire)?

ATTRACTION-3 applied the EQ-5D-3L questionnaire, in line with NICE guidance.

C5. In Section B.3.2.2, the CS states that *"ESMO guidelines recommend taxane monotherapy for the second-line treatment (after failure of first-line treatment with taxane combination therapy) of OC"*. Please can the company confirm if the reference to first-line taxane combination therapy is an error?

In the section titled "Management of advanced/metastatic disease" in the ESMO guidelines (Lordick et al., 2016[26]), the following section of text is provided:

Chemotherapy is indicated for palliative treatment in selected patients, particularly for patients with AC who have a good PS [performance score]. Despite scarce evidence, treatment of advanced oesophageal AC [adenocarcinoma] is managed mostly according to the recommendations for gastric cancer. Newer regimens based on oxaliplatin/fluoropyrimidine combinations are an alternative to the 'classical' cisplatin/5-FU schedule. Infusional 5-FU may be replaced by capecitabine if the swallowing of tablets is not compromised. Taxanes are recommended in first-line combinations or as monotherapy in second-line therapy.

In SCC [squamous cell carcinoma], the value of palliative chemotherapy is less proved. Cisplatin-based combinations showed increased response rates but no survival gain compared with monotherapy. Overall, results with palliative chemotherapy are inferior to those in AC. Therefore, best supportive care (BSC) or palliative monotherapy should also be considered.

While the company submission statement is poorly worded and implies that taxane monotherapy requires prior failure of taxane combination therapy, the inference is that taxane monotherapy are the mainstay of palliative chemotherapy, particularly in SCC.

C6. Table 56 of the CS (Document B) states that the dose of nivolumab is expected to be *"3mg/kg by intravenous infusion over 60 mins every 2 weeks"*.

Please can the company confirm that reference to weight-based dose of nivolumab is an error?

The dose of nivolumab is expected to be 240mg by intravenous infusion over 60 minutes every 2 weeks.[1] Therefore, the weight-based dose of nivolumab was given in error. The correct dosing regimen was applied in the economic model.

C7. In Section B.3.7.1, it is stated that nivolumab provides a total of [redacted] discounted QALYs compared with taxane therapy. However, Table 75 states that this value is [redacted]. Please confirm which of these values was provided in error.

This is a typographical error and the correct value is [redacted].

C8. Section 3.5.4 of Appendix D (page 25) states that "The review identified *three* studies, including 275 patients in total, that described the clinical efficacy of nivolumab for the treatment of adult patients with oesophageal cancer30, 31." Please confirm that only two studies were identified.

Only two studies were identified.

References

1. Ono Pharmaceutical Co. Ltd. ONO-4538 Phase III Study. A multicenter, randomized, open-label study in patients with esophageal cancer refractory or intolerant to combination therapy with fluoropyrimidine- and platinum-based drugs - study report. 2019.
2. Ono Pharmaceutical Co. Ltd. ONO-4538 Phase II Study A Multicenter, Open-Label, Uncontrolled Study in Patients with Esophageal Cancer - study report. 2016.
3. West HJ. Immune checkpoint inhibitors. JAMA oncology. 2015;1(1):115-.
4. Chiou VL, Burotto M. Pseudoprogression and immune-related response in solid tumors. Journal of Clinical Oncology. 2015;33(31):3541-3.
5. Borcoman E, Nandikolla A, Long G, Goel S, Le Tourneau C. Patterns of Response and Progression to Immunotherapy. American Society of Clinical Oncology Educational Book. 2018(38):169-78.
6. Jin S, Pazdur R, Sridhara R. Re-Evaluating Eligibility Criteria for Oncology Clinical Trials: Analysis of Investigational New Drug Applications in 2015. J Clin Oncol. 2017;35(33):3745-52.
7. Moriwaki T, Kajiwara T, Matsumoto T, Suzuki H, Hiroshima Y, Matsuda K, et al. Survival analysis of platinum-refractory patients with advanced esophageal cancer treated with docetaxel or best supportive care alone: a retrospective study. Dis Esophagus. 2014;27(8):737-43.
8. Janmaat VT, Bruno MJ, Polinder S, Lorenzen S, Lordick F, Peppelenbosch MP, et al. Cost-Effectiveness of Cetuximab for Advanced Esophageal Squamous Cell Carcinoma. PLoS One. 2016;11(4):e0153943.
9. Bascoul-Mollevi C, Gourgou S, Galais MP, Raoul JL, Bouche O, Douillard JY, et al. Health-related quality of life results from the PRODIGE 5/ACCORD 17 randomised trial of FOLFOX versus fluorouracil-cisplatin regimen in oesophageal cancer. European journal of cancer (Oxford, England : 1990). 2017;84:239-49.
10. Doherty MK, Leung Y, Su J, Naik H, Patel D, Eng L, et al. Health utility scores from EQ-5D and health-related quality of life in patients with esophageal cancer: a real-world cross-sectional study. Dis Esophagus. 2018.
11. Dutton SJ, Ferry DR, Blazeby JM, Abbas H, Dahle-Smith A, Mansoor W, et al. Gefitinib for oesophageal cancer progressing after chemotherapy (COG): a phase 3, multicentre, double-blind, placebo-controlled randomised trial. The Lancet Oncology. 2014;15(8):894-904.
12. Shenfine J, McNamee P, Steen N, Bond J, Griffin SM. A randomized controlled clinical trial of palliative therapies for patients with inoperable esophageal cancer. The American journal of gastroenterology. 2009;104(7):1674-85.
13. Tian D, Wen H, Fu M. Comparative study of self-expanding metal stent and intraluminal radioactive stent for inoperable esophageal squamous cell carcinoma. World Journal of Surgical Oncology. 2016;14(1):18.
14. Xinopoulos D, Dimitroulopoulos D, Tsamakidis K, Korkolis D, Fotopoulou A, Bazinis A, et al. Palliative treatment of advanced esophageal cancer with metal-covered expandable stents. A cost-effectiveness and quality of life study. J buon. 2005;10(4):523-8.
15. National Institute for Health and Care Excellence Decision Support Unit. Technical Support Document 14: Survival analysis for economic evaluations alongside

clinical trials - extrapolation with patient-level data.2013 7 June 2016. Available from: http://www.nicedsu.org.uk/NICE%20DSU%20TSD%20Survival%20analysis.updated%20March%202013.v2.pdf.
16. Bagust A, Beale S. Survival Analysis and Extrapolation Modeling of Time-to-Event Clinical Trial Data for Economic Evaluation An Alternative Approach. Medical Decision Making. 2014;34(3):343-51.
17. Grothey A, Van Cutsem E, Sobrero A, Siena S, Falcone A, Ychou M, et al. Regorafenib monotherapy for previously treated metastatic colorectal cancer (CORRECT): an international, multicentre, randomised, placebo-controlled, phase 3 trial. Lancet. 2013;381(9863):303-12.
18. National Institute for Health and Care Excellence. Nivolumab for previously treated locally advanced or metastatic non-squamous non-small-cell lung cancer [ID900].In development [GID-TAG524].2017 1 September 2017 Available from: https://www.nice.org.uk/guidance/indevelopment/gid-tag524.
19. National Institute for Health and Care Excellence. Nivolumab for previously treated locally advanced or metastatic squamous non-small cell lung cancer [ID811]. Single Technology Appraisal: Evidence Review Group Report. In development [GID-TAG506].2015 14 September 2017. Available from: https://www.nice.org.uk/guidance/gid-tag506/documents/committee-papers
20. National Institute for Health and Care Excellence. Technology appraisal guidance [TA462]. Nivolumab for treating relapsed or refractory classical Hodgkin lymphoma. 2017 14 September 2017. Available from: https://www.nice.org.uk/guidance/ta462.
21. National Institute for Health and Care Excellence. Technology appraisal guidance [TA384]. Nivolumab for treating advanced (unresectable or metastatic) melanoma.2016 14 September 2017. Available from: https://www.nice.org.uk/guidance/ta384.
22. National Institute for Health and Care Excellence. Technology appraisal guidance [TA400]. Nivolumab in combination with ipilimumab for treating advanced melanoma. 2016 14 September 2017. Available from: https://www.nice.org.uk/guidance/ta400.
23. National Institute for Health and Care Excellence. Technology appraisal guidance [TA417]. Nivolumab for previously treated advanced renal cell carcinoma. 2016 14 September 2017. Available from: https://www.nice.org.uk/guidance/ta417.
24. National Institute for Health and Care Excellence. Ramucirumab for treating advanced gastric cancer or gastro–oesophageal junction adenocarcinoma previously treated with chemotherapy (TA378). 2016.
25. National Institute for Health and Care Excellence. Pembrolizumab for previously treated oesophageal or gastro-oesophageal junction cancer ID1357. In development [GID-TA10322].2018 12 September 2018. Available from: https://www.nice.org.uk/guidance/indevelopment/gid-ta10322.
26. Lordick F, Mariette C, Haustermans K, Obermannová R, Arnold D. Oesophageal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of Oncology. 2016;27(suppl 5):v50-v7.

NICE National Institute for Health and Care Excellence

Professional organisation submission

Nivolumab for previously treated unresectable advanced oesophageal cancer [ID1249]

Thank you for agreeing to give us your organisation's views on this technology and its possible use in the NHS.

You can provide a unique perspective on the technology in the context of current clinical practice that is not typically available from the published literature.

To help you give your views, please use this questionnaire. You do not have to answer every question – they are prompts to guide you. The text boxes will expand as you type.

Information on completing this submission

- Please do not embed documents (such as a PDF) in a submission because this may lead to the information being mislaid or make the submission unreadable
- We are committed to meeting the requirements of copyright legislation. If you intend to include **journal articles** in your submission you must have copyright clearance for these articles. We can accept journal articles in NICE Docs.
- Your response should not be longer than 13 pages.

About you	
1. Your name	[redacted]
2. Name of organisation	**The NCRI-ACP-RCP-RCR is grateful for the opportunity to respond to the above consultation. We have liaised with our experts and [redacted], [redacted] and [redacted] have responded as below.**

3. Job title or position	[redacted]
4. Are you (please tick all that apply):	☒ an employee or representative of a healthcare professional organisation that represents clinicians? ☐ a specialist in the treatment of people with this condition? ☐ a specialist in the clinical evidence base for this condition or technology? ☐ other (please specify):
5a. Brief description of the organisation (including who funds it).	**NCRI-ACP-RCP-RCR**
4b. Has the organisation received any funding from the manufacturer(s) of the technology and/or comparator products in the last 12 months? [Relevant manufacturers are listed in the appraisal matrix.]	**No**

If so, please state the name of manufacturer, amount, and purpose of funding.	
5c. Do you have any direct or indirect links with, or funding from, the tobacco industry?	**No**
The aim of treatment for this condition	
6. What is the main aim of treatment? (For example, to stop progression, to improve mobility, to cure the condition, or prevent progression or disability.)	
7. What do you consider a clinically significant treatment response? (For example, a reduction in tumour size by	

x cm, or a reduction in disease activity by a certain amount.)	
8. In your view, is there an unmet need for patients and healthcare professionals in this condition?	
What is the expected place of the technology in current practice?	
9. How is the condition currently treated in the NHS?	Squamous oesophageal cancer has a poor prognosis and has historically been underserved by research and drug development. In fact, most treatments for advanced squamous cell carcinoma are based on trials from gastroesophageal adenocarcinoma patients. However, the recent large scale molecular profiling study from the Cancer Genome Atlas (TCGA) has clearly shown that oesophageal SCC is a different entity from adenocarcinoma and as a result we are now seeing appropriately defined clinical trial populations.[1] ATTRACTION-3 was a practice changing international phase III randomised trial comparing nivolumab to taxane chemotherapy in patients with ESCC which was refractory or intolerant to one prior platinum/fluoropyrimidine chemotherapy regimen [2] As such, the patients in the trial have received similar treatment to patients with advanced oesophageal SCC in the UK. ATTRACTION 3 met its primary endpoint; overall survival was improved for nivolumab treated patients (10.9 months versus 8.4 months [HR 0.77 [95% CI 0·62–0·96]; p=0·02]). This 2.5 month improvement in survival is non-trivial when considered in the context of a disease where median overall survival is less than one year. However, the improvement in median

	overall survival may not be the most important metric when considering the benefit of nivolumab for oesophageal SCC patients, landmark survival i.e. 12 month and 18 month survival are more substantially increased by the use of the PD-1 inhibitor (notably 18 month survival is improved by almost 50% from 21% to 31% for nivolumab treated patients). An important second metric to consider this that nivolumab has a better safety profile compared to chemotherapy (patients treated with nivolumab had more than three times lower Grade 3/4 treatment-related adverse events). This treatment tolerability is critical for patient quality of life. Importantly, there was a high rate of completion of quality of life questionnaires in both arms of the trial and nivolumab showed a statistically significant improvement in quality of life and decreased time to deterioration in quality of life compared to chemotherapy. In ATTRACTION-3, the PD-L1 biomarker was not significantly associated with response to nivolumab, and therefore nivolumab is recommended for all patients independent of biomarker status. We acknowledge that limited non-Asian enrolment was noted in ATTRACTION-3. However, outside of endemic areas the drivers of oesophageal SCC are very similar, regardless of geography (namely alcohol and tobacco use) suggesting that the underlying biology of oesophageal SCC should be similar globally. Indeed, the landmark TCGA analysis which demonstrated the biological differences between oesophageal SCC and adenocarcinoma at the molecular level did not show any significant differences between oesophageal SCC from different countries. For this reason we believe that nivolumab is equally likely to be as effective in UK compared to Asian patients. In summary, we fully support the application for NICE funding for nivolumab in previously treated patients with oesophageal squamous cell carcinoma. This drug has the potential to provide long term benefit with

	low toxicity compared to chemotherapy and also improves quality of life. In a disease where there are very few useful treatment options this can offer new hope to patients after chemotherapy. 1. Cancer Genome Atlas Research Network. Integrated genomic characterization of oesophageal carcinoma. Nature. 2017 Jan;541(7636):169. 2. Kato K, Cho BC, Takahashi M, Okada M, Lin CY, Chin K, Kadowaki S, Ahn MJ, Hamamoto Y, Doki Y, Yen CC. Nivolumab versus chemotherapy in patients with advanced oesophageal squamous cell carcinoma refractory or intolerant to previous chemotherapy (ATTRACTION-3): a multicentre, randomised, open-label, phase 3 trial. The Lancet Oncology. 2019 Nov 1;20(11):1506-17.
• Are any clinical guidelines used in the treatment of the condition, and if so, which?	
• Is the pathway of care well defined? Does it vary or are there differences of opinion between professionals across the NHS? (Please state if your experience is from outside England.)	

• What impact would the technology have on the current pathway of care?	
10. Will the technology be used (or is it already used) in the same way as current care in NHS clinical practice?	
• How does healthcare resource use differ between the technology and current care?	
• In what clinical setting should the technology be used? (For example, primary or secondary care, specialist clinics.)	
• What investment is needed to introduce the technology? (For example, for facilities, equipment, or training.)	
11. Do you expect the technology to provide clinically	

meaningful benefits compared with current care?	
• Do you expect the technology to increase length of life more than current care?	
• Do you expect the technology to increase health-related quality of life more than current care?	
12. Are there any groups of people for whom the technology would be more or less effective (or appropriate) than the general population?	
The use of the technology	
13. Will the technology be easier or more difficult to use for patients or healthcare	

professionals than current care? Are there any practical implications for its use (for example, any concomitant treatments needed, additional clinical requirements, factors affecting patient acceptability or ease of use or additional tests or monitoring needed.)	
14. Will any rules (informal or formal) be used to start or stop treatment with the technology? Do these include any additional testing?	
15. Do you consider that the use of the technology will result in any substantial health-related benefits that are unlikely to be included in the	

quality-adjusted life year (QALY) calculation?	
16. Do you consider the technology to be innovative in its potential to make a significant and substantial impact on health-related benefits and how might it improve the way that current need is met?	
• Is the technology a 'step-change' in the management of the condition?	
• Does the use of the technology address any particular unmet need of the patient population?	
17. How do any side effects or adverse effects of the technology affect the	

management of the condition and the patient's quality of life?	
Sources of evidence	
18. Do the clinical trials on the technology reflect current UK clinical practice?	
• If not, how could the results be extrapolated to the UK setting?	
• What, in your view, are the most important outcomes, and were they measured in the trials?	
• If surrogate outcome measures were used, do they adequately predict long-term clinical outcomes?	
• Are there any adverse effects that were not apparent in clinical trials	

but have come to light subsequently?	
19. Are you aware of any relevant evidence that might not be found by a systematic review of the trial evidence?	
20. Are you aware of any new evidence for the comparator treatment(s)?	
21. How do data on real-world experience compare with the trial data?	
Equality	
22a. Are there any potential equality issues that should be taken into account when considering this treatment?	

22b. Consider whether these issues are different from issues with current care and why.	

Key messages

23. In up to 5 bullet points, please summarise the key messages of your submission.

-
-
-
-
-

Thank you for your time.

Please log in to your NICE Docs account to upload your completed submission.

……

Your privacy

The information that you provide on this form will be used to contact you about the topic above.

☐ **Please tick this box** if you would like to receive information about other NICE topics.

For more information about how we process your personal data please see our privacy notice.

……

Nivolumab for unresectable, advanced oesophageal cancer when standard chemotherapy has failed [ID1249]

A Single Technology Appraisal

Produced by Peninsula Technology Assessment Group (PenTAG)
University of Exeter Medical School
South Cloisters
St Luke's Campus
Heavitree Road
Exeter
EX1 2LU

Authors
Mr Ash Bullement, Associate[1] and Analyst[2]
Dr Linda Long, Research Fellow[1]
Dr Kevin Deighton, Associate[1] and Analyst[2]
Ms Naomi Shaw, Information Specialist[1]
Dr Stephen Falk, Consultant Clinical Oncologist[3]
Dr Nicole Dorey, Consultant Clinical Oncologist[4]
Ms Louise Crathorne, Senior Research Fellow[1]
Prof G.J. Melendez-Torres, Professor[1]
Dr Maxwell S. Barnish, Research Fellow[1]
[1] Peninsula Technology Assessment Group (PenTAG), University of Exeter Medical School, Exeter
[2] Delta Hat Limited, Nottingham
[3] Bristol Haematology and Oncology Centre, Bristol
[4] Royal Devon and Exeter NHS Foundation Trust, Exeter

Correspondence to Dr Maxwell S. Barnish
3.09f South Cloisters, St Luke's Campus, Heavitree Road, Exeter, EX1 2LU; m.s.barnish@exeter.ac.uk

Date completed 29/06/2020

Source of funding	This report was commissioned by the NIHR Systematic Reviews Programme as project number 18/54/01.
Declared competing interests of the authors	Dr Dorey declares educational support from Boehringer-Ingelheim; Roche; Lilly and Astra-Zeneca.
Acknowledgments	The authors acknowledge the administrative support provided by Mrs Sue Whiffin and Ms Jenny Lowe (both PenTAG).
Rider on responsibility for document	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	Bullement A, Long L, Deighton K, Shaw N, Falk S, Dorey N, Crathorne L, Melendez-Torres GJ, Barnish MS. Nivolumab for unresectable, advanced oesophageal cancer when standard chemotherapy has failed [ID1249]: A Single Technology Appraisal. Peninsula Technology Assessment Group (PenTAG), 2020.

Author Contributions:	
Ash Bullement	Acted as health economic project lead, critiqued the company's economic evaluation and contributed towards writing the report. Implemented the ERG's preferred settings and exploratory analyses within the company's economic model.
Linda Long	Critical appraisal of the clinical effectiveness evidence. Led and wrote report sections on the critique of methods, trial design and clinical effectiveness outcomes. Contributed to editing of the report.
Kevin Deighton	Acted as health economist, critiqued the company's economic evaluation and contributed to the writing of the report
Naomi Shaw	Critical appraisal of the literature search strategies. Editorial input.
Stephen Falk	Clinical advisor.
Nicole Dorey	Clinical advisor.
Louise Crathorne	Summarised and critiqued the reviews of cost effectiveness reported in the company's submission. Drafted and commented on the final report.
G.J. Melendez-Torres	Critical appraisal of the company submission, writing and editorial input, and supervised the final report. Guarantor of the report.
Maxwell S. Barnish	Led the delivery of the ERG report. Summarised and critiqued the clinical effectiveness data within the company submission. Drafted and commented on the final report.

Table of Contents

List of tables

List of Figures

Abbreviations

Abbreviation	Description
AE	adverse event
AIC	Akaike's Information Criterion
AR	autoregressive
BIC	Bayesian Information Criterion
BNF	British National Formulary
BSC	best supportive care
CEAC	cost-effectiveness acceptability curve
CI	confidence interval
CMH	Cochran-Mantel-Haenszel
CR	complete response
CS	company submission
CT	computed tomography
CSR	clinical study report
DCRDN	day case and regular day/night
DSA	deterministic sensitivity analysis
DSU	Decision Support Unit
ECOG	Eastern Cooperative Oncology Group
eMIT	electronic Market Information Tool
EQ-5D	EuroQol five dimension
EQ-5D-3L	EuroQol five-dimension three level
EQ-VAS	EuroQol Visual Analogue Scale
ERG	Evidence Review Group
ESMO	European Society for Medical Oncology
EU	Europe
FAD	final appraisal determination
FP	fully parametric
GI	gastrointestinal
GOJ	gastro oesophageal junction
HR	hazard ratio
HRG	Healthcare resource use
HRQoL	health-related quality of life
HTA	health technology assessment
ICER	incremental cost-effectiveness ratio

ITC	indirect treatment comparison
ITT	intention-to-treat
IWRS	interactive web response system
KM	Kaplan-Meier
LYs	life years
MAR	missing at random
MCAR	missing completely at random
MICE	Multiple Imputation by Chained Equations
MIMS	Monthly Index of Medical Specialities
MNAR	missing not at random
MRU	medical resource use
NA	not applicable
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NMA	network meta-analysis
NR	not reported
OC	oesophageal cancer
OR	odds ratio
ORR	objective response rate
OS	overall survival
OSCC	oesophageal squamous cell carcinoma
OWSA	one-way sensitivity analysis
PartSA	partitioned survival analysis
PAS	Patient Access Scheme
PD	progressive disease
PD-L1	programmed death-ligand 1
PF	progression free
PFS	progression free survival
PP	post progression
PR	partial response
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PS	performance status
PSA	probabilistic sensitivity analysis
PSS	Personal Social Services
PSSRU	Personal Social Services Resource Unit
QA	quality assessment

QALY	quality adjusted life year
QoL	quality of life
RCT	randomised controlled trial
RES	response evaluable set
RoB	risk of bias
ROBINS-I	Risk Of Bias In Non-randomized Studies - of Interventions
ROW	rest of world
RR	response rate
SD	standard deviation
SE	standard error
SLR	systematic literature review
SP	semi parametric
TA	Technology Appraisal
ToT	time on treatment
TSD	Technical Support Document
UK	United Kingdom
VAS	visual analogue score
VBA	Visual Basic for Applications
vs	versus
WTP	willingness to pay

1. EXECUTIVE SUMMARY

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

The ERG considered that the company's definition of the decision problem generally matched the decision problem in the NICE scope.[1] The population in the company's decision problem was narrower than the NICE scope in only including squamous patients, although this was in line with the proposed marketing authorisation for nivolumab. The ERG was satisfied that the comparators were similar between the NICE scope and the company's decision problem. One comparator in the NICE scope namely irinotecan was excluded from the company's decision problem (Section 2.3). The ERG did not consider this to be an issue given that clinical advice confirmed that irinotecan was not commonly used in the UK for this indication.

1.2. Summary of the key issues in the clinical effectiveness evidence

The key trial included from the company's SLR, and the only trial to inform the company model in the CS, is a Phase III, open label parallel RCT (ATTRACTION-3).[2] This was an open-label trial, which the ERG considered to be the most substantial limitation for internal validity, especially with regard to safety and HRQoL outcomes. The ERG also had concerns regarding the generalisability of the trial to UK clinical practice. This was due to the limitation of only including patients with ECOG PS scores 0-1, which would represent a fitter and healthier group than seen in routine practice (Section 3.2.4). Moreover, 97% of patients in ATTRACTION-3[2] were Asian, while approximately two-thirds of total patients were Japanese. Different treatment guidelines are used in Asia and these may result in substantial differences in the treatment pathways, especially with regard to Japan. Subgroup analyses on Japanese patients versus Rest of the World (ROW) indicated that geography was an important consideration. Considering the primary OS outcome, the hazard ratios for nivolumab versus taxanes were comparable for Japan ([redacted]) and ROW ([redacted]). However, while the relative effects were similar, the absolute OS values were not. Japanese patients receiving nivolumab had considerably longer OS than ROW patients ([redacted] vs [redacted]) and it is notable that Japanese patients on taxanes had superior OS than ROW patients on nivolumab ([redacted] vs [redacted]) (Section 3.2.5.5).

The adverse event (AE) profile for nivolumab was generally favourable compared to taxanes. However, the ERG was concerned that early deaths were considerably higher on nivolumab

than taxanes ([redacted] vs [redacted]) – and the company commented that this was potentially related to the mechanism of action of immunotherapies versus chemotherapy agents (Section 3.2.5.6).

There was no direct trial evidence available to compare nivolumab with best supportive care (BSC), which was a comparator specified in the NICE final scope for this appraisal.[1] Therefore, the company used an indirect treatment comparison (ITC) to compare nivolumab with BSC. The ERG considered that there were several key issues with the ITC analysis that preclude it from producing an appropriate estimate of the relevant effectiveness of nivolumab versus BSC (Section 3.3). The studies used in the ITC were not randomised, it is not clear how comparability of population was ensured, and it is not clear that the evidence can be generalised to UK clinical practice, in particular since the other studies in the ITC besides ATTRACTION-3[2] draw on Japan-only populations.

The ITC itself suffered from a number of issues (Section 3.4), including a lack of transitivity and a sparse network. While the ERG was able to verify and re-run the WinBUGS code used to undertake the analyses, insufficient details regarding burn-in iterations discarded or checks for convergence were provided to provide confidence in the analysis presented. Finally, the ERG was unable to trace back with consistency the input estimates used in the comparison between docetaxel and paclitaxel to the corresponding estimates in the included studies.

1.3. Summary of the key issues in the cost effectiveness evidence

The ERG did not consider the combination of the company's base-case projections of OS for the nivolumab and taxanes arm to be the most appropriate estimates to inform the model, given the follow-up data available and the generalisability issues with ATTRACTION-3.[2] as discussed in Section 1.2. The ERG noted concerns with the way background mortality was modelled (Section 4.2.6), but considered any double counting of mortality to partially address the generalisability issues with ATTRACTION-3.[2]

The ERG also noted concerns regarding how utility values were estimated (Section 4.2.7). It was the ERG's understanding that imputation of missing EQ-5D values is rarely undertaken. The ERG considered it likely that 'true' utility values at later time points were systematically lower than those seen in earlier time points (based on the general principle that utility declines over time, both related to disease progression and natural health decline as patients age). This is especially important within the context of an economic model which is capable of projecting survival outcomes over a 40-year lifetime horizon.

The economic model assumed a 50:50 market share of taxanes (docetaxel and paclitaxel) which, based on advice provided to the ERG, does not reflect UK clinical practice. Moreover, the costs used in the company model for the comparator taxanes were not reflective of the average price paid by NHS trusts (obtained through drugs and pharmaceutical electronic market information tool [eMIT]), and some other medical resource use costs were not considered reflective of the anticipated resources required or were misaligned with the stated references.

The model structure assumed that patients continued their third-line therapy after discontinuation of nivolumab or a taxane until death, and that survival was explained via the specification of a single over-arching OS curve. Should any active intervention be used in the third-line setting, this assumption may lead to an over-estimation of costs incurred. No adjustment to efficacy was made for any beneficial effects of active third-line therapy.

The ERG highlighted that in its submission the company urged caution when considering the outcome of disease progression owing to the potential role of pseudo-progression in response to checkpoint inhibitor therapy. In spite of this potential issue, the company developed a progression-based model within its submission, though PFS has a limited impact on model results.

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

The ERG's preferred assumptions for the cost-effectiveness analysis of nivolumab compared with taxanes are outlined in Table 1.

Table 1: ERG preferred assumptions

Preferred assumption	Section in ERG report	Cumulative ICER £/QALY
Company base-case	***5.1.1***	***45,491***
SP generalised gamma (5.75) OS models	*6.3.1*	*62,440*
SP Weibull (5.75) ToT models	*6.3.2*	*68,343*
Correction of taxanes costs	*6.3.3*	*80,614*
ERG's preferred administration costs	*6.3.4*	*77,198*
ERG's preferred utility values	*6.3.5*	*106,643*
Update of unit costs for MRU	*6.3.6*	*125,984*

Abbreviations: ERG, Evidence Review Group; ICER, incremental cost-effectiveness ratio; MRU, medical resource use; OS, overall survival; QALY, quality adjusted life year; SP, semiparametric; ToT, time on treatment.

The ERG's preferred base-case settings lead to an ICER of £125,984 per QALY gained.

Table 2: ICER resulting from ERG's preferred assumptions

	Total costs	Total QALYs	Δ costs	Δ QALYs	ICER £/QALY
Taxane	[redacted]	[redacted]			
Nivolumab	[redacted]	[redacted]	27,845	0.221	125,984

Abbreviations: ICER, incremental cost effectiveness ratio; QALYs, quality adjusted life years

1.5. Summary of exploratory and sensitivity analyses undertaken by the ERG

The ERG's exploratory analyses are presented in Table 3.

Table 3: Exploratory analyses undertaken by the ERG

Analysis description	Section in ERG report	ICER £/QALY
Company base-case	***5.1.1***	***45,491***
Remove AE costs	*6.2.1*	*47,671*
ERG background mortality	*6.2.2*	*42,749*
Remove background mortality		*42,299*
Pragmatic ToT estimation (1% at 3 years)	*6.2.3*	*41,501*
Pragmatic ToT estimation (1% at 4 years)		*45,323*
Pragmatic ToT estimation (1% at 5 years)		*49,034*
ATTRACTION-3 taxane split	*6.2.4*	*44,703*
100% docetaxel		*47,578*
Average PP value	*6.2.5*	*55,449*
Minimum PP value		*59,215*
Custom* small benefit (both states)		*58,830*
Custom* moderate benefit (both states)		*56,119*
Custom* large benefit (both states)		*53,646*
Change clinician consultation cost	*6.2.6*	*62,008*
Change hospitalisation cost		*45,575*
Change both		*62,092*

Abbreviations: AE, adverse event; ERG, Evidence Review Group; ICER, incremental cost-effectiveness ratio; PP, post-progression; QALY, quality adjusted life year; ToT, time on treatment.

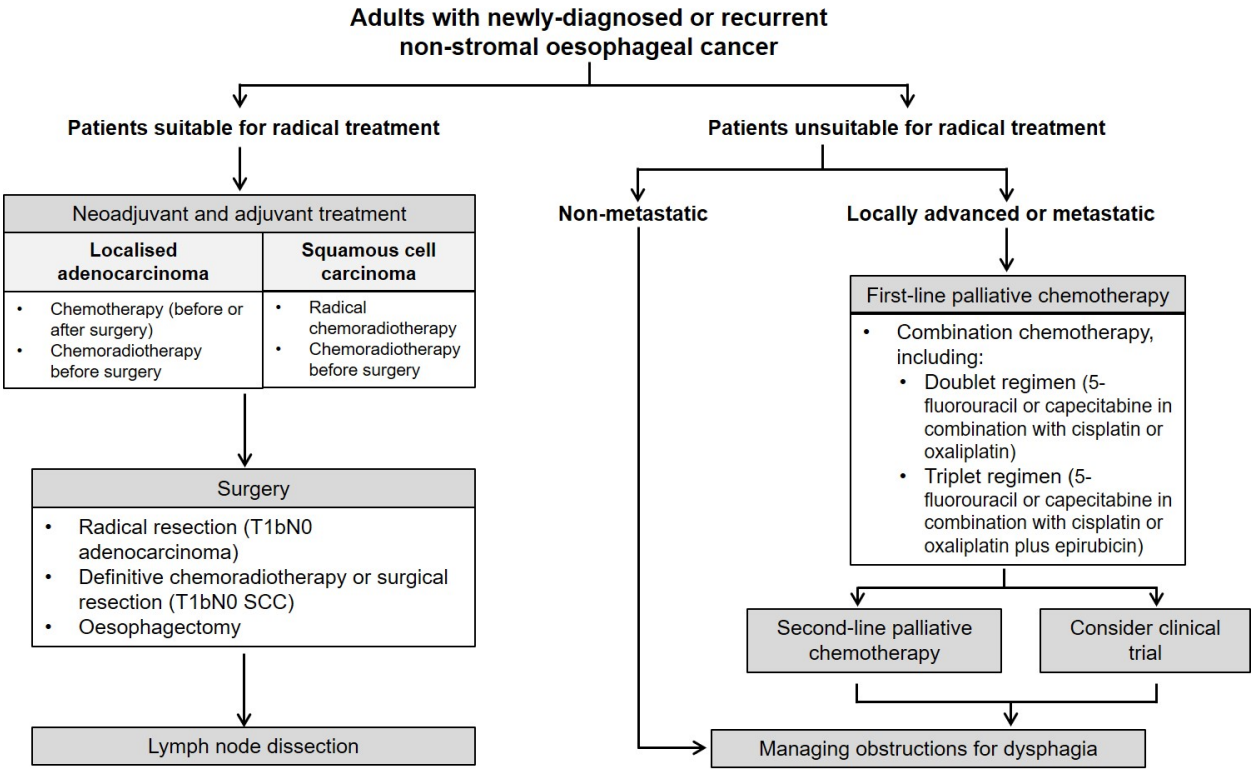
2. INTRODUCTION AND BACKGROUND

2.1. Introduction

Oesophageal cancer (OC) is a malignant tumour developing from the cells lining the oesophagus.[3] Prognosis for unresectable OC is poor with less than half (42%) of patients in England surviving 12 months regardless of stage of diagnosis.[4] While relatively rare in terms of incidence, OC represents the seventh most common cause of cancer death in the United Kingdom (UK), responsible for an estimated 7,295 deaths in the UK in 2017, reflecting extremely poor survival rates, with only around 15% of people diagnosed with OC surviving five years or more.[5,6] Around 64% of OC cases are adenocarcinoma with around 31% being oesophageal squamous cell carcinoma (OSCC).[5] In the UK, OC is often diagnosed at a late stage and adenocarcinoma is much more common in men, while the prevalence of OSCC is similar in men and women.[5] The Evidence Review Group (ERG) considered that the Company Submission (CS) offered an acceptable description of the condition; its pathophysiology, natural course and epidemiology; and the current treatment options available.

Current National Institute for Health and Care Excellence (NICE) guidance on the management of OC (NG83)[7] specifies several stages of treatment of newly diagnosed non-stromal OC. Treatment options differ according to suitability for radical treatment, which depends on clinical characteristics and patient fitness. Therefore, those receiving radical treatment tend to be younger and fitter. Clinical advice to the ERG confirmed that the specific part of the treatment pathway for OC that is of direct relevance to this appraisal is the right hand pathway under 'locally advanced or metastatic' and in particular the box currently occupied by second-line chemotherapy, which in the UK typically comprises taxane monotherapy using docetaxel or paclitaxel, as well as the box labelled 'managing obstructions for dysphagia', which in the UK typically comprises best supportive care (BSC), such as the use of stents, with or without the addition of palliative radiotherapy.

Figure 1. Treatment pathway for oesophageal cancer in the UK (NG83)



Source: CS Document B Figure 4, p.19

2.2. Background

Nivolumab is a fully human monoclonal antibody immunotherapy drug administered intravenously. It belongs to a different class of drug than the taxanes docetaxel and paclitaxel, currently used as second-line treatments for unresectable OSCC. Nivolumab is currently used for a range of other cancer indications in existing practice. The ERG considered that the company's intended positioning, as compared to current standard of care, was appropriate and well-described.

The company's intended positioning for nivolumab is as monotherapy in second-line position for unresectable locally advanced OSCC when standard chemotherapy has failed. This is the position in the treatment pathway currently occupied by the taxane chemotherapy agents, docetaxel and paclitaxel. Clinical advice to the ERG was that both of these agents were used in UK clinical practice, but that there may be a preference for docetaxel for service provision and resource reasons due to lower frequency of administration. The ERG was advised that irinotecan is not commonly used in the UK in this position in the treatment pathway.

2.3. Critique of company's definition of decision problem

The ERG considered that the company's definition of the decision problem generally matched the decision problem in the NICE scope.[1]

Table 4: Summary of decision problem

	Final scope issued by NICE	Decision problem addressed in the CS	Rationale if different from the final NICE scope	ERG comment
Population	Adults with previously treated advanced or recurrent unresectable OC that is refractory or intolerant to chemotherapy.	Nivolumab as monotherapy for the treatment of adult patients with unresectable advanced, recurrent or metastatic OSCC after prior fluoropyrimidine and platinum-based combination therapy.	Not applicable.	The ERG was satisfied that the population matched the proposed marketing authorisation for nivolumab. It was narrower than the NICE scope in only including squamous patients.
Intervention	Nivolumab.	Nivolumab.	As per NICE scope.	The ERG was satisfied that the intervention matched the NICE scope.
Comparator(s)	• Chemotherapy including taxanes (docetaxel/paclitaxel) or irinotecan • Best supportive care (including, but not limited to antiemetics, blood transfusions, oesophageal stents)	• Chemotherapy including taxanes (docetaxel/paclitaxel) • Best supportive care (including, but not limited to antiemetics, blood transfusions, oesophageal stents)	The main treatment options in this setting are primarily palliative. However, the majority of patients in this setting will receive taxane monotherapy, based on market research and clinician opinion. Some patients are unable to receive chemotherapy and these patients will receive BSC. Clinicians felt strongly that irinotecan would not be used in the UK setting for treatment of second-line OSCC. This view is supported by market research, where irinotecan comprises only 6% of current usage.	The ERG was satisfied that the comparators were similar between the NICE scope and the company's decision problem. Clinical advice to the ERG agreed that irinotecan was not commonly used in the UK for this indication.
Outcomes	The outcome measures to be considered include: • OS • PFS • RR	The outcome measures to be considered include: • OS • PFS • RR	As per NICE scope.	The ERG agreed that the outcomes in the company's decision problem matched those in the NICE scope.

	• AEs of treatment • HRQoL	• AEs of treatment • HRQoL		
Economic analysis	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per QALY. The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared. Costs will be considered from an NHS and Personal Social Services perspective.	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per QALY. The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared. Costs will be considered from an NHS and Personal Social Services perspective.	As per NICE scope.	The ERG agreed that the economic analysis presented is aligned with the reference case.
Subgroups	No patient subgroups have been identified.	No patient subgroups have been identified.	As per NICE scope.	N/A.
Special considerations including issues related to equity or equality	No equality issues have been identified.	No equality issues have been identified.	As per NICE scope.	N/A.

Abbreviations AEs, adverse events; BSC, best supportive care; CS, company submission; ERG, Evidence Review Group; N/A, Not applicable; NHS, National Health Service; NICE, National Institute for Health and Care Excellence; OC, oesophageal cancer; OS, overall survival; OSCC, oesophageal squamous cell carcinoma; PFS, progression free survival; QALY, quality adjusted life year; RR, response rate

Source: CS Document B Table 3, p.10

3. CLINICAL EFFECTIVENESS

The sections below discuss the evidence submitted by the company in support of the clinical effectiveness of nivolumab as a second-line therapy for unresectable advanced, recurrent or metastatic OSCC after prior fluoropyrimidine and platinum-based combination therapy. The ERG has critiqued the details provided on:

- Methods implemented to identify, screen and data extract relevant evidence;
- Clinical efficacy of nivolumab;
- Safety profile of nivolumab;
- Assessment of comparative clinical effectiveness of nivolumab against relevant comparators.

A detailed description of an aspect of the CS is provided only when the ERG disagrees with the company's assessment or proposal, or where the ERG has identified a potential area of concern that the ERG considered necessary to highlight for the Committee.

Broadly speaking, the ERG considered that the methodology and outcome data relevant to the decision problem were adequately reported in the CS. Where gaps were identified, the ERG was largely able to identify the information from elsewhere (e.g. from trial publications or the clinical study report), although certain requested analyses were not provided by the company. In particular, the company did not provide four-way stratified analyses by programmed death-ligand 1 (PD-L1) status and Japan vs rest of world (ROW).

3.1. Critique of the methods of review(s)

The company undertook a systematic review to identify relevant publications on the efficacy and safety of nivolumab monotherapy, compared to other potential second-line treatments, in patients with oesophageal cell carcinoma (both squamous and adenosquamous) refractory or intolerant to combination therapy with fluoropyrimidine and platinum-based drugs. The company considered docetaxel and paclitaxel monotherapies, and best supportive care (BSC) to be the most relevant comparators.

In total, 54 unique studies (describing 74 treatment arms) were identified. Most studies identified in the SLR were not randomised and had not attempted to blind patients or assessors to

treatment allocation. One RCT (ATTRACTION-3)[2] was identified that included the target population (patients with squamous oesophageal carcinoma). The identified evidence is critiqued in Section 3.2.

A summary of the ERG's critique of the methods implemented by the company to identify evidence relevant to the decision problem is presented in Table 5.

Table 5: Summary of ERG's critique of the methods implemented by the company to identify evidence relevant to the decision problem

Systematic review step	Section of CS in which methods are reported	ERG assessment of robustness of methods
Searches	Appendix D, Section 2.1	The ERG was broadly satisfied that the searches identified the most relevant evidence for nivolumab in the population of interest. The ERG noted the following limitations: use of a filter other than Cochrane for RCTs; use of limited search terms to identify cohort and observational studies; use of search terms for Outcomes (instead of Intervention and Comparator terms) in combination with Population search terms; no searches of trial registers; no adverse event search.
Inclusion criteria	Appendix D, Section 2.1.1 Table 1	Broadly appropriate. The ERG noted the population was narrower than the NICE scope. Patients with adenocarcinoma only were excluded and studies reporting patients with mixed squamous/adenocarcinoma were excluded if there was no breakdown of results for squamous carcinoma or less than 70% of patients had squamous cell carcinoma. Populations were eligible only if they had received at least one prior treatment i.e. the treatment regimen assessed was a second-line therapy.
Screening	Appendix D, Section 2.3	Appropriate. All abstracts were dual screened versus pre-defined eligibility criteria with discrepancies resolved with a third party. Potential full text articles were retrieved and screened in the same way.
Data extraction	Appendix D, Section 2.4	Appropriate. Data was extracted by a single reviewer using a pre-defined data extraction template, and data was checked by a second reviewer.
Tool for quality assessment of included study or studies	Appendix D, Section D.1.5	Broadly appropriate. Quality assessment was undertaken by two independent reviewers using a checklist modified from the Downs and Black (1998) instrument. Any discrepancies between the two reviewers were resolved by consensus or involvement of a third reviewer.

		The ERG considered that the use of Downs and Black (1998) quality appraisal tool in the SLR is reasonable given the inclusion of both RCTs and non-RCTs in the SLR. The ERG considered that using more recently developed quality appraisal tools such as ROBINS-I (for non-RCTs) and the Cochrane Risk of Bias tool, vs 2 (for RCTs) would have been more robust. However, the ERG did not perform an independent RoB assessment for ATTRACTION-3 using the Cochrane Risk of Bias tool vs 2, considering it would not highlight any relevant additional information already presented in the CS.
Evidence synthesis	Document B, Section B.2.7	Direct comparative evidence for nivolumab vs taxane chemotherapy was available in ATTRACTION-3. The ERG agreed that no standard meta-analysis was required. There was no direct trial evidence comparing nivolumab vs BSC, so a network meta-analysis (NMA) was used for this comparison. The ERG considered that the NMA was not robust, especially due to issues relating transitivity, and possibly unnecessary.

Abbreviations: CS, Company submission; ERG, Evidence Review Group; RCT, randomised controlled trial; RoB, risk of bias; ROBINS-I, Risk Of Bias In Non-randomized Studies - of Interventions; SLR, systematic literature review

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

3.2.1. Study design

The key trial included from the company's SLR, and the only trial to inform the company model in the CS, is a Phase III, open label parallel RCT (ATTRACTION-3)[2] evaluating nivolumab in patients with advanced oesophageal squamous cell carcinoma from Denmark, Germany, Italy, Japan, South Korea, Taiwan, the UK. The data reported in the CS are from planned subgroup comparisons of nivolumab and chemotherapy (paclitaxel or docetaxel) in patients who broadly met the NICE decision problem criteria. The CS additionally presented clinical effectiveness data from one further study (ATTRACTION-1).[8] It was not used to inform the economic model, since it was a single-arm study and used a different dose of nivolumab treatment. The ERG considered this to be appropriate and therefore did not present further critique of this study.

The population, intervention, and outcomes presented in ATTRACTION-3[2] were broadly consistent with the NICE decision problem, with median time to event data available for all the main clinical outcomes in both arms.

The ERG considered that the dose of nivolumab used in ATRACTION-3[2] (240 mg for 30 minutes every two weeks) to be consistent with UK clinical practice and consistent with the NICE decision problem. ERG clinical advisors suggest that the dose of paclitaxel used in ATTRACTION-3[2] (100 mg/m^2 on a six-weekly basis) may be higher than that routinely used in UK clinical practice (60-80mg weekly for three out of four weeks).

The comparator arm in ATTRACTION-3[2] consisted of either docetaxel or paclitaxel. At clarification (question A8), the ERG queried whether the decision to prescribe docetaxel or paclitaxel was undertaken pre-randomisation. This would have allowed comparatively more robust comparisons of nivolumab vs each taxane, considering the effectiveness of nivolumab in 'docetaxel-preferred' or 'paclitaxel-preferred' populations. In response, the company clarified that while the choice of taxanes was recorded pre-randomisation, subgroup analyses by taxane would not be provided as this was not a stratification factor. While the ERG appreciates that these analyses would be unplanned, it notes that a potentially more robust comparison of nivolumab vs each taxane could have been made available to support decision analytic modelling.

3.2.2. Randomisation stages and protocol amendments

The ATTRACTION-3[2] trial involved the randomisation of patients (1:1) to either nivolumab or investigator's choice of chemotherapy (paclitaxel or docetaxel). Randomisation was carried out appropriately, and stratified according to geographical region (Japan vs ROW), number of organs with metastases (≤1 vs ≥2), and expression of PD-L1 (<1% vs ≥1%).

ATTRACTION-3[2] was subject to eight protocol amendments. However, the ERG did not find any protocol amendments that may have introduced a high risk of bias in addition to the high risk of bias inherent in a non-blinded open-label trial design.

3.2.3. Quality assessment of the trials of the technology of interest

The company reported no notable quality issues in relation to ATTRACTION-3[2] as well as for the single-arm nivolumab ATTRACTION-1[8] in the CS (CS B1.5, p.30). The complete quality assessment is available in Appendix D of the CS. The company evaluated both trials using the Downs and Black[9] assessment tool which is an old tool predominantly used for the assessment of non-RCTs. The ERG considered risk of bias using the published literature and the data presented in the CS specifically for the outcomes reported in the CS from ATTRACTION-3[2] that informed the decision problem / economic model (primarily overall survival [OS], response,

progression free survival [PFS], health-related quality of life [HRQoL] and also adverse events [AEs]).

While the ERG noted some strengths of trial quality such as appropriate randomization, intention to treat analysis of primary outcome measures and broadly similar baseline characteristics and withdrawals between the two arms, the ERG noted some quality issues specifically relating to the open-label design of ATTRACTION-3,[2] where patients and investigators are not masked to treatment allocation. The ERG noted potential limitations with the open-label treatment of nivolumab in ATTRACTION-3,[2] particularly given that nivolumab, docetaxel and paclitaxel are all intravenously administered drugs. The ERG noted a substantial limitation with the open-label design in respect of internal validity, especially with regard to safety and HRQoL outcomes. Specifically, while for the objective measurement of the main clinical outcomes (PFS, response and OS), the risk of bias arising from lack of blinding is likely to be low, the ERG noted that in subjective measures of HRQoL and some safety data the risk of bias might be higher.

3.2.4. Baseline characteristics

Baseline characteristics for patients included in the ATTRACTION-3[2] study were reported in Table 12 Section B1.6.1.8 of the CS (p.13), While the ERG notes a tendency for more early stage disease in the comparator arm, the ERG considered there to be no major imbalances in baseline characteristics between the two arms of ATTRACTION-3.[2] Clinical advisors to the ERG considered the baseline characteristics to generally match the patient population that would be treated in the UK. However, it should still be seen as a potential limitation in terms of the generalisability of trial results to UK practice that 96% of patients in ATTRACTION-3[2] were Asian, of which approximately two-thirds were Japanese. While clinical advice to the ERG was that ethnicity was unlikely to have a major effect *per se*, it is important to consider differences in treatment pathways. In particular, it was noted that there is a specific pan-Asian adapted version[10] of the European Society for Medical Oncology (ESMO) practice guidelines[11] which introduces certain differences into the typical pathways seen in the UK. It may thus be that subgroup analyses excluding Japanese sites may provide evidence that is of greater generalisability to the UK context. Furthermore, only patients with Eastern Cooperative Oncology Group (ECOG) performance status (PS) 0-1 are only included in the trial, which suggests that participants in ATTRACTION-3[2] are fitter and otherwise systematically different than those encountered in routine UK practice.

3.2.5. Clinical effectiveness results

Data in the target population were presented for OS, PFS, response, HRQoL and safety (AEs). The ERG consider efficacy outcomes were measured appropriately. Statistical methods were broadly appropriate. For OS and PFS outcomes, analyses were performed in the intention to treat population. The company provided Kaplan–Meier plots for PFS and OS in Figure 10 and Figure 11 respectively (CS, Document B, p.38 and 39), from which median PFS and OS (and corresponding two-sided 95% confidence intervals) were estimated. Hazard ratios and corresponding two-sided 95% confidence intervals for the nivolumab group relative to the control group were estimated using a stratified Cox proportional hazards model. For objective response rate (ORR), analyses were performed in the response evaluable set (RES). ORR was compared between the two treatment groups using the Cochran-Mantel-Haenszel (CMH) test with the randomisation factors as stratification factors.

The ERG requested (clarification question A17) analyses with stratification by PD-L1 expression and geographic location (Japan vs ROW). The company indicated (clarification response A17) that this information was not available. The ERG agreed that four-way stratified analyses were not provided by the company, however it did identify analyses stratified separately by PD-L1 expression and by Japan vs ROW in the clinical study report[12] for OS (Figure 11.4-8, p.169), PFS (Figure 11.4-10, p.171) and response (Figure 11.4-12, p.173 and Figure 11.4-14, p.175), as presented below (Section 3.2.5.5).

3.2.5.1. Overall survival

For the primary endpoint of OS (defined as the time from randomisation until death from any cause), a stratified log-rank test by three stratification factors (Japan vs ROW; the number of organs with metastases (≤1 vs. ≥2); and PD-L1 expression (≥1% vs. <1% or indeterminate)), demonstrated the superiority of nivolumab over the control groups (two-sided [redacted]). The hazard ratio of the nivolumab group relative to the control group was [redacted]. The Kaplan–Meier estimate of median OS was [redacted] in the nivolumab group and [redacted] in the control group. Clinical advice to the ERG suggested this was a clinically meaningful difference. However, hazard ratios may not appropriately summarise relative treatment effects given that the two treatment curves crossed, suggesting a violation of the proportional hazards assumption. In the CS and CSR, the OS rates estimated by the Kaplan–Meier method were reported to be numerically higher in the nivolumab group than in the control group from follow-up at Month 6 through to Month 30. The ERG noted

that while analysis of survival rates from the Kaplan–Meier curve were planned for six, nine, 12, 15 and 18 months, the OS rates at 21, 24 and 30 months were derived from an unplanned post-hoc analysis (CSR,[12] Section 11.4.2.9, p.177 and Table 11.4-2, p. 148).

3.2.5.2. Progression-free survival

The secondary endpoint PFS was calculated from the following equation: "Time from date of randomisation until either the overall response was assessed as progressive disease or the patient died of any cause, whichever was the earlier"+1)/30.4375), The hazard ratio of the nivolumab group relative to the control group was [redacted]. The Kaplan–Meier estimate of median PFS was [redacted] in the nivolumab group and [redacted] in the control group. Although the median PFS was higher in the control group compared to the nivolumab group, the hazard ratio of the nivolumab group to the control group was [redacted] and the 95% CI of the hazard ratio included 1. As for OS, hazard ratios may not appropriately summarise relative treatment effects given that the two treatment curves crossed, suggesting a violation of the proportional hazards assumption. The PFS rate estimated by the Kaplan–Meier method was reported in the CS to be numerically higher in the control group at Month 3, but PFS rate was numerically higher in the nivolumab group at Month 6 onwards. The PFS rates in the nivolumab and control groups were reported in the CS to be [redacted] and [redacted] at Month 6, 11.9% and 7.2% at Month 12, and [redacted] and [redacted] at Month 18, respectively (CS Document B, Table 13, p.37).

3.2.5.3. Response rate

Overall response rate was similar between the nivolumab and chemotherapy groups (objective response rate (ORR): [redacted], complete response (CR: [redacted]), partial response (PR): [redacted] stable disease (SD): [redacted], progressive disease (PD): [redacted] (CS Document B, Table 13, p.37).

Objective response rate (ORR) (defined in the CS as the percentage of patients whose best overall response is assessed as either CR or PR) is specified in the decision problem. ORR was similar between nivolumab and chemotherapy, with an odds ratio (OR) of the nivolumab group relative to the control group of [redacted] ([redacted]) (CSR,[12] Section 11.4.9, p.177).

3.2.5.4. Health-related quality of life

Clinical advisors to the ERG agreed with the company's assertion that in the case of OC, a worsened quality of life (QOL) from the time of diagnosis is often observed, resulting from

swallowing disorder and nutritional disorder caused by oesophageal narrowing, cough caused by aspiration and fistula, and chest pain due to tumour. As a pre-specified exploratory endpoint, health-related quality of life was assessed based on the three-level version of the EuroQol five-dimension three level (EQ-5D-3L) questionnaire, comprising the visual analogue scale (VAS) and descriptive system, which is used to generate the utility index.

A summary of EQ-5D index scores at each time-point in ATTRACTION-3[2] (up to 54 weeks) is provided in Table 14 (p.41) and Figure 12 (p.42) of the CS. The company noted that additional time points are available in the CSR[12] but represent smaller patient numbers.

The general health condition was observed to be better maintained in the nivolumab group compared to the control group. In the nivolumab arm, no meaningful changes in the proportion of patients who reported QoL-related problems were observed during the treatment period in any of the EQ-5D categories. In the control arm, however, the proportion of patients who reported QoL-related problems in the mobility, self-care and usual activities categories after commencing chemotherapy increased by [redacted] compared with the proportion at the screening stage. (CSR,[12] p. 164).

A summary of EuroQol Visual Analogue Scale (EQ-VAS) scores at each time point is presented in Table 15 of the CS (CS Document B, p.42).

Patients treated with nivolumab had a decreased risk of deterioration in QoL compared with patients treated with chemotherapy for the VAS (HR [redacted]; median time to deterioration [redacted]) and the utility index (HR [redacted]).[2] While EQ-VAS scores remained relatively stable among patients treated with nivolumab, a worsening of [redacted] scores was observed among patients treated with chemotherapy from early time points after commencing treatment.

Clinical advisors to the ERG confirmed that improvements in HRQoL with nivolumab vs chemotherapy are clinically meaningful. However, the ERG noted that the lack of blinding inherent in the ATTRACTION-3[2] study design could bias subjective measures of QoL, potentially inflating the effect of nivolumab on HRQoL.

3.2.5.5. Subgroup analyses

Prespecified, exploratory subgroup analyses assessed the association between overall survival and stratification factors or baseline variables: PD-L1 expression (<1%, ≥1%, >5%, ≥5%, <10%,

and ≥10%), age (<65 years vs ≥65 years), sex (male vs female), race (Asian vs white), ECOG PS (0 vs 1), previous surgery (no vs yes), previous radiotherapy (no vs yes), and history of smoking (never, former, or current).

Results of subgroup analyses for the ATTRACTION-3 nivolumab and chemotherapy arms, as at the database lock on 30 November 2018, are shown in Table 13, Figure 13 and Figure 14 of the CS (CS, Document B, p. 37, 46-47). For OS, the superior treatment effect of nivolumab over chemotherapy was consistently observed across the majority of subgroups (Figure 13 and Figure 14) (CS Document B, p. 46-47). Median OS in patients with tumour PD-L1 expression of less than 1% was [redacted] and at least 1% was [redacted] with nivolumab, while with chemotherapy it was [redacted] respectively.[2] The ERG concurs that superior treatment effect of nivolumab over chemotherapy was consistently observed across the majority of subgroups for PFS and ORR.

Considering the primary OS outcome, the hazard ratios for nivolumab versus taxanes were comparable for Japan ([redacted]) and ROW ([redacted]). However, while the relative effects were similar, the absolute OS values were not. Japanese patients receiving nivolumab had considerably longer OS than ROW patients ([redacted] vs [redacted]) and it is notable that Japanese patients on taxanes had superior OS than ROW patients on nivolumab ([redacted] vs [redacted]). There was a statistically non-significant advantage in PFS for taxanes over nivolumab in terms of the hazard ratios for both Japanese ([redacted]) and ROW ([redacted]) patients. For ORR, the pattern of results differed between Japanese and ROW participants. In Japan, ORR was similar between arms ([redacted]), while in ROW ORR was lower in the nivolumab arm ([redacted]) than the taxanes arm ([redacted]), although there was considerable uncertainty and statistical significance was not reached ([redacted]). Considering duration of response, there was evidence of a statistically significant benefit for nivolumab over taxanes in Japan ([redacted]) but not in ROW ([redacted]).This again suggested that analyses excluding Japanese sites may provide more generalisable evidence to the UK context.

3.2.5.6. Adverse effects

Adverse events (AEs) in the ATTRACTION-3[2] study were reported in Section B.2.10 of the CS. Overall, the ERG agreed with the company that nivolumab had an acceptable safety profile.

AEs were very common. However, the overall AEs rate was lower on nivolumab ([redacted]) than control [redacted]. Similar patterns were found for serious adverse events (SAEs) ([redacted] vs [redacted]) and drug-related SAEs ([redacted] vs [redacted]). The ERG noted that early deaths, defined as [redacted] (CS, p.72) were notably higher on nivolumab than control ([redacted] vs [redacted]).The CS largely attributed these early deaths to 'initial disease' (CS, Table 29, p.75). The ERG asked the company for clarification regarding why the early death rate was so much higher on nivolumab than control (clarification question A13). The company responded (clarification response A13) that this may relate to differences in the mechanism of action of immunotherapy treatments such as nivolumab compared to chemotherapy agents. Potentially of relevance, according to the company, are a longer time to response in immunotherapies compared to chemotherapy agents, and the indirect anti-tumour mechanism associated with immunotherapies, which may result in initial growth of existing lesions or formation of new lesions, prior to potential tumour shrinkage or eradication.

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

The company presented an indirect treatment comparison (ITC) to estimate the relative impact of docetaxel, paclitaxel and BSC on OS in the target population. A total of three studies beyond ATTRACTION-3[2] contribute to the ITC (p. 7, CS Appendix L). One study,[13] compared best supportive care to docetaxel, whereas two studies[14,15] in addition to ATTRACTION-3[2] provided estimates comparing paclitaxel to docetaxel.

Collectively, this body of evidence has several major flaws that do not permit appropriate estimation of relative treatment effectiveness for the population in the decision problem. First, none of the evidence draws from randomised comparisons. Relatedly, it is unclear what steps were taken to ensure comparability of populations in the contributing studies. For example, it appears that the docetaxel-paclitaxel comparison drawn from ATTRACTION-3[2] was naïve, without due regard to baseline differences between patients receiving docetaxel and patients receiving paclitaxel; similarly, the additional studies used either multivariable adjustment or naïve comparison to estimate relative effectiveness. Second, included populations were sicker than the population included in ATTRACTION-3,[2] including a wider range of ECOG PS scores, suggesting incommensurability with ATTRACTION-3 estimates. Third, outcomes were not

measured consistently across included studies. Specifically, Moriwaki *et al.* (2014)[13] used post-progression survival instead of OS. The exchangeability of this effect estimate with OS is a question of assumption rather than fact, though the company note in response to clarification question A19 that post-progression survival was 'comparable to the definition of OS applied in other studies'. Finally, all studies except for ATTRACTION-3[2] draw on Japan-only populations. As discussed elsewhere (Section 3.2.4), treatment pathways and disease presentation vary significantly between Japan and ROW, including the UK.

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

The ITC does not provide a reliable basis for a comparison between BSC and treatments included in ATTRACTION-3.[2] Hypothetically, the ITC presented in this submission could have been used to estimate the relative effectiveness of nivolumab, docetaxel, paclitaxel, irinotecan and BSC. However, the company asserted that the inclusion of nivolumab would be inappropriate given the distinct survival curves that immunotherapies produce (and thus the inappropriateness of a hazard ratio to summarise effectiveness), and subsequently excluded irinotecan as well. While the ERG regarded that setting aside irinotecan as a comparator in this appraisal was defensible, this limited the relevant comparators that could enter into an analysis. The result is a sparse network of evidence with no closed loops and thus no 'borrowing strength' from indirect evidence. The company asserted that including wider loops of evidence based on comparators not directly relevant to the analysis would 'decrease the power of the NMA to estimate the links of interest' (p. 8, CS Appendix L). The basis for this assertion was unclear.

The company presented an ITC comparing docetaxel, paclitaxel and BSC. As described above and in Section 3.3, this network was ultimately sparse, and was only estimated for OS as an outcome. Thus, its functional purpose was not to estimate the relative effectiveness of different treatment strategies, but to estimate a hazard ratio that could be used to compare BSC against nivolumab (p. 70, CS Document B). This was done by applying the hazard ratio from the ITC against data from the docetaxel arm from ATTRACTION-3.[2] During clarification (clarification question A22), the company noted that a comparative estimate of BSC against paclitaxel, which would have been based strictly on indirect evidence, was not considered to be 'appropriate' because there was 'no evidence available to validate the assumption of proportional hazards'. This suggests that an approach that weights the two model-generated hazard ratios for docetaxel vs BSC and paclitaxel vs BSC would not be useful, but it also suggests that in the event there was no additional value gained from undertaking an ITC. It also underscores that the only link between BSC and ATTRACTION-3 is a single retrospective study with a

systematically different population in a different practice context, with a different outcome definition than that used in ATTRACTION-3. It is possible that an alternative meta-analysis including BSC against a pooled taxane arm could have provided a more direct approach to constructing a comparison between nivolumab and BSC.

Statistical methods used in the ITC were broadly appropriate as described, including using log hazard ratios and a fixed effects model. Analyses were undertaken in a Bayesian framework, but minimally informative priors were used, despite the challenges these priors can pose in sparse networks. The company did not present sensitivity analyses; while Section 4.2.2 of CS Appendix L made reference to a base case analysis and page 15 of CS Appendix L mentions additional models that were not presented, response to clarification question A21 denies that any additional analyses were undertaken. Thus, it is unclear the degree to which the company checked their analysis for robustness. While the ERG was able to verify and re-run the WinBUGS code used to undertake the analyses, insufficient details regarding burn-in iterations discarded or checks for convergence were provided to provide confidence in the analysis presented.

Finally, the ERG was unable to trace back with consistency the input estimates used in the comparison between docetaxel and paclitaxel to the corresponding estimates in the included studies, in particular for Shirakawa *et al.* (2014)[15] While CS Appendix L makes reference to reconstructed data used to estimate hazard ratios, which is required for studies where no Cox proportional hazards model is presented, the response to clarification question A18 seems to be at variance with this, noting that reconstructed data 'were not used as inputs to the NMA'. This further underlined the unreliability of the models presented.

3.5. Additional work on clinical effectiveness undertaken by the ERG

The ERG scrutinised the code presented for the ITCs, including accuracy of the inputs and of the code provided (refer to critique in Section 3.4).

3.6. Conclusions of the clinical effectiveness section

The ERG considered that the company had identified all relevant clinical evidence for this appraisal. Data were available for all outcomes indicated in the NICE final scope for this appraisal.[1] Information related to the methodology and outcomes for clinical effectiveness was generally available in the CS. Where this was not the case, the ERG was generally able to gain the required information through the CSR, trial publications and requests to the company during

clarification. However, some additional analyses requested by the ERG were not provided by the company. In particular, the company did not provide four-way stratified analyses by PD-L1 status and Japan vs ROW. Only analyses stratified separately by these factors were provided.

There was only one clinical trial for nivolumab versus taxanes that could inform the economic model: the ATTRACTION-3 study.[2,12] The ERG considered that there were several strengths of the ATTRACTION-3[2] study, but considered the open label design to be the most substantial limitation for internal validity, especially with regard to safety and HRQoL outcomes. The ERG was satisfied that nivolumab was superior to taxanes with regard to the primary efficacy outcome of OS as well as the secondary outcome HRQoL, while response rate was similar between the groups and PFS was numerically superior in the nivolumab arm but significance was not reached. The AE profile of nivolumab was generally favourable compared to taxanes, but the ERG was concerned about the fact that early deaths were around twice as likely in the nivolumab arm as the taxane arm. The generalisability of ATTRACTION-3[2] to UK practice is a concern given the fact that 97% of patients were Asian and approximately two thirds were from Japan. Differences in routine treatment pathways as a result of country of treatment were identified. While the relative effect of nivolumab vs taxanes on the primary OS outcome was comparable between Japan and ROW, absolute OS values were considerably higher in Japan, and it was notable that Japanese patients in the taxane control arm had superior OS than ROW patients on nivolumab. Therefore, analyses excluding Japanese sites may provide more generalisable evidence to the UK context.

There was no direct trial evidence comparing nivolumab with BSC, which is a relevant comparator in the NICE final scope for this appraisal.[1] Therefore, an ITC was used to make the comparison between nivolumab and BSC. However, the ERG did not consider the body of evidence used for the ITC or the ITC itself to permit appropriate estimation of the relative treatment effectiveness of nivolumab and BSC in the population for the decision problem. The studies used for the ITC were not randomised, it is not clear how comparability of population was ensured, and it is not clear that the evidence can be generalised to UK clinical practice, in particular since the other studies in the ITC besides ATTRACTION-3[2] draw on Japan-only populations. The ITC itself suffered from a number of issues, including a lack of exchangeability and a sparse network. While the ERG was able to verify and re-run the WinBUGS code used to undertake the analyses, insufficient details regarding burn-in iterations discarded or checks for convergence were provided to provide confidence in the analysis presented. Finally, the ERG

was unable to trace back with consistency the input estimates used in the comparison between docetaxel and paclitaxel to the corresponding estimates in the included studies.

4. COST-EFFECTIVENESS

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

The company carried out an SLR, using three separate search strategies, to identify existing cost-effectiveness evidence, HRQoL evidence, and cost and healthcare resource use evidence associated with previously treated advanced or recurrent unresectable OC that is refractory or intolerant to chemotherapy.

4.1.1. Systematic review of cost-effectiveness models

A summary of the ERG's critique of the methods implemented by the company to identify relevant evidence is presented in Table 6.

Table 6. Summary of ERG's critique of the methods implemented by the company to identify cost-effectiveness models

Systematic review step	Section of CS in which methods are reported	ERG assessment of robustness of methods
Searches	Appendix G (Section G.1.1, Section and G.1.2, and Section G.1.3)	Broadly appropriate. The ERG noted the following limitations: inconsistent use of free-text search terms for oesophageal cancer across bibliographic databases; highly relevant subject heading terms were unexploded in the filter to identify economic evaluations in Embase.
Inclusion criteria	Appendix G (Section G.1.4)	Appropriate. Broad criteria were applied. Full economic evaluations of interventions aimed at managing advanced, unresectable OC whose major lesion was in the oesophagus (mixed populations were considered if the majority of patients [>50%] fulfilled the criteria), published in English language from Year 2000 were in scope.
Screening	Appendix G (Section G.1.5)	Appropriate[a]. Initial reporting discrepancies identified in the PRISMA were resolved during clarification (clarification question B4)
Data extraction	Appendix G (Section G.1.6)	Appropriate[b]
QA of included studies	Appendix G (Section G.3)	Appropriate[c]

Abbreviations: CS, Company Submission; ERG, Evidence Review Group; OC, oesophageal cancer; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; QA, quality assessment

Notes:

[a] Abstracts were dual screened versus pre-defined eligibility criteria. Discrepancies were resolved with a third party. Potential full text articles were retrieved and screened in the same way. A list of excluded studies was provided in Appendix G Section G.4 of the CS together with reasons for exclusion

[b] Data was extracted by a single reviewer, and data was checked by a second reviewer

[c] Critical appraisal was conducted using the Drummond checklist

The ERG was satisfied with the company's review of the cost-effectiveness literature. One economic model[16] was identified that evaluated the cost-effectiveness of cisplatin and fluorouracil alone compared with cisplatin and fluorouracil in combination with cetuximab. A tabulated summary of characteristics is provided in table format in the CS (Document B [Table 37] and Appendix G [Table 7]); however, no discussion as to how the model informed the development of the economic model in the CS was provided. In response to clarification question B1, the company added a brief comment highlighting the linear modelling approach common with many cancers and approaches used within that had been applied in the current model (specifically listing quality of life and treatment discontinuation in respect of cost and resource use). The company also highlighted issues with transferability of the utility values (derived from patients with Barrett's oesophagus which has a different aetiology to oesophageal squamous cell carcinoma). Although the ERG considered the commentary was limited, it was broadly satisfied.

4.1.2. Systematic review of health effects

A summary of the ERG's critique of the methods implemented by the company to identify relevant evidence reporting health effects (health-related quality of life and utilities) is presented in Table 7.

Table 7. Summary of ERG's critique of the methods implemented by the company to identify studies reporting health effects

Systematic review step	Section of CS in which methods are reported	ERG assessment of robustness of methods
Searches	Appendix H (Section H.1.1, Section H.1.2, Section H.1.3)	Broadly appropriate. The ERG noted the following limitation: inconsistent use of free-text search terms for oesophageal cancer across bibliographic databases.
Inclusion criteria	Appendix H (Section H.1.4 and Table 6)	Appropriate. Broad criteria were applied. Studies reporting HRQoL or utility values related to pre-treated advanced, unresectable OC whose major lesion was in the oesophagus (mixed populations were considered if the majority of patients [>50%] fulfilled the criteria), published in English language from Year 2000 were included.
Screening	Appendix H (Section H.1.5 and Figure 1)	Appropriate[a]
Data extraction	Appendix H (Section H.1.6)	Appropriate[b]

QA of included studies	Appendix H (Section H.3)	Appropriate[c]

Abbreviations: CS, Company Submission; ERG, Evidence Review Group; HRQoL, health-related quality of life; OC, oesophageal cancer; QA, quality assessment

Notes:

[a] Abstracts were dual screened versus pre-defined eligibility criteria with discrepancies resolved with a third party. Potential full text articles were retrieved and screened in the same way. A list of excluded studies was provided in Appendix H Section H.4 of the CS together with reasons for exclusion

[b] Data was extracted by a single reviewer using a pre-defined data extraction template, and data was checked by a second reviewer

[c] Quality assessment conducted using the scale in NICE Decision Support Unit Technical Support Document 9 (Table 37)

The ERG was satisfied with the company's review of the literature reporting health effects. Six studies were identified[17-22] which are summarised in tables in Appendix H (Table 7 and Table 8) of the CS. While the company did not provide comment on the applicability of the identified studies to the economic model within the CS, this information was provided during clarification (clarification question B2). The assessment of suitability of the identified evidence to inform the economic modelling provided by the company was considered appropriate.

4.1.3. Systematic review of healthcare resource use and costs

A summary of the ERG's critique of the methods implemented by the company to identify relevant evidence reporting healthcare resource use and costs is presented in Table 8.

Table 8. Summary of ERG's critique of the methods implemented by the company to identify health economic (healthcare resource use and costs) evidence

Systematic review step	Section of CS in which methods are reported	ERG assessment of robustness of methods
Searches	Appendix I (Section I.1.1, Section I.1.2, Section I.1.3)	Broadly appropriate. The ERG noted the following limitation: inconsistent use of free-text search terms for oesophageal cancer across bibliographic databases.
Inclusion criteria	Appendix I (Section I.1.4 and Table 6)	Appropriate. Broad criteria were applied. Studies reporting cost and healthcare resource use related to the treatment of adults with advanced OC whose major lesion was in the oesophagus (mixed populations were considered if the majority of patients [>50%] fulfilled the criteria), conducted in a UK or EU setting were included. Studies published in English language from Year 2000.
Screening	Appendix I (Section I.1.5 and Figure 1)	Appropriate[a]
Data extraction	Appendix I (Section I.1.6)	Appropriate[b]

QA of included studies	Not completed	No critical appraisal was conducted of identified studies.

Abbreviations: CS, Company Submission; ERG, Evidence Review Group; OC, oesophageal cancer; QA, quality assessment

Notes:

[a] Abstracts were dual screened versus pre-defined eligibility criteria with discrepancies resolved with a third party. Potential full text articles were retrieved and screened in the same way. A list of excluded studies was provided in Appendix I, Section I.3 of the CS together with reasons for exclusion

[b] Data was extracted by a single reviewer using a pre-defined data extraction template, and data was checked by a second reviewer

[c] Quality assessment conducted using the scale in NICE Decision Support Unit Technical Support Document 9 (Table 37)

The ERG was broadly satisfied with the company's review of the literature reporting healthcare resource use and costs. Seven studies were identified,[22-28] and were summarised in Appendix I (Table 7) of the CS. However, there was no discussion of the applicability of the identified studies to the economic model within the CS. During clarification (clarification question B3), the company commented that none of the identified studies had reported the composition of standard of care and, as such, none of the studies identified were used in the economic model.

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

4.2.1. NICE reference case checklist

Table 9: NICE reference case checklist

Attribute	Reference case	ERG comment on CS
Perspective on outcomes	All direct health effects, whether for patients or, when relevant, carers	✓
Perspective on costs	NHS and PSS	✓
Type of economic evaluation	Cost–utility analysis with fully incremental analysis	✓
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	✓ Time horizon of 40 years used. A shorter time horizon would likely have been sufficient as the majority of patients with current care in the UK have an estimated survival of <12 months
Synthesis of evidence on health effects	Based on systematic review	✓ Systematic review undertaken to identify relevant evidence. Used to validate utility estimates based on EQ-5D data available from ATTRACTION-3, and to perform sensitivity analysis

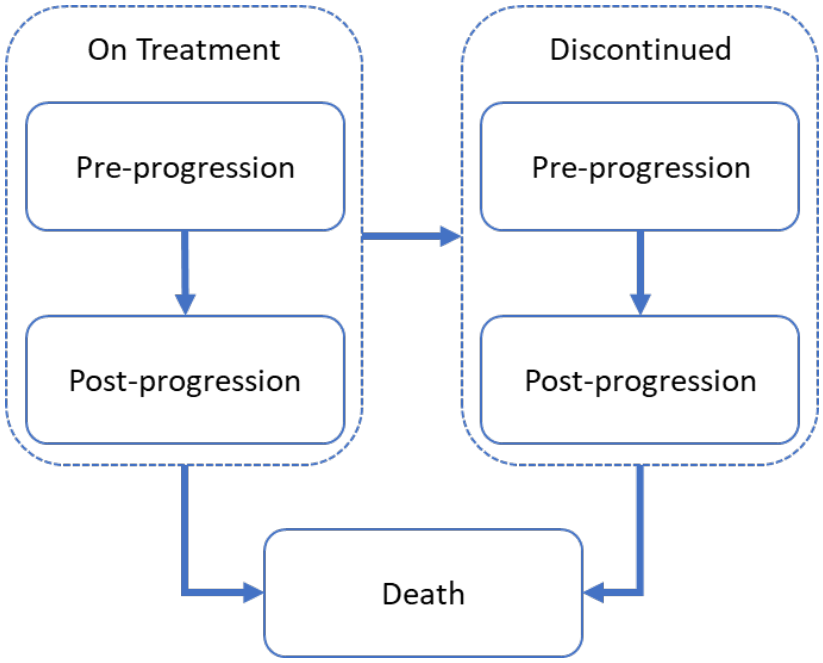
Measuring and valuing health effects	Health effects should be expressed in QALYs. The EQ-5D is the preferred measure of HRQoL in adults.	✓ EQ-5D utility values used to inform the model, though large differences in values between treatment arms and by progression status
Source of data for measurement of HRQoL	Reported directly by patients and/or carers	✓ Reported directly by patients in the ATTRACTION-3 trial
Source of preference data for valuation of changes in HRQoL	Representative sample of the UK population	✓ Based on EQ-5D UK value set
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	✓
Evidence on resource use and costs	Costs should relate to NHS and PSS resources and should be valued using the prices relevant to the NHS and PSS	✗ Costs for docetaxel and paclitaxel are not reflective of the average price paid by NHS trusts (obtained through eMIT)
Discounting	The same annual rate for both costs and health effects (currently 3.5%)	✓

Abbreviations: EQ-5D, EuroQol 5 dimension; CS, company submission; eMIT, electronic Market Information Tool; HRQoL: health-related quality of life; NHS, National Health Service; PSS, Personal Social Services; QALY: quality-adjusted life year; TA: technology appraisal

4.2.2. Model structure

The company developed a partitioned survival analysis (PartSA) economic model to estimate the cost effectiveness of nivolumab versus either docetaxel or paclitaxel ('taxane'), shown in Figure 2.

Figure 2: Company's model schematic



Source(s): CS Figure 21

Based on data from ATTRACTION-3,[2] the model makes use of OS and PFS curves to determine the proportion of patients who are alive with non-progressed disease ('progression-free'), alive with progressed disease ('progressed'), or dead. Progression-based health state occupancy is demonstrated via CS Figure 22. The model adopts a cycle length of seven days (one week).

Separately to the estimation of progression-based health state occupancy, the model also makes use of time on treatment (ToT) data to estimate the proportion of patients expected to remain on treatment with either nivolumab or a taxane at each model cycle. ToT does not affect the estimation of QALYs within the model, but is used to determine drug acquisition, administration, and some medical resource use costs.

The model allows for patients to discontinue before disease progression *and* continue after disease progression (discussed further in Section 4.2.6). However, as progression and treatment status are modelled independently, it is not possible to infer from the company's model the relative proportions of patients that are on-/off-treatment and progression-free/progressed for each model cycle. After discontinuation, the model assumes patients will

remain on their third-line treatment until death (which is set to BSC in the company's base-case analysis).

The PartSA model has been used extensively to inform previous company submissions to NICE of cancer treatments, particularly for those used in an advanced disease setting.[29] The ERG considered the choice of model structure to be suitable to inform decision making in this appraisal, though notes the following limitations which are important to consider when interpreting the results of the model and its functionality:

- Use of PFS means that the proportion of patients that progress each model cycle cannot be explicitly calculated. Should the timing of progression be of importance (e.g. for costs of additional tests and investigations initiated upon documented progression), it would not be possible to assign these accurately within the model. However, as no model inputs are dependent upon a precise estimate of the proportion of progression events per cycle, the ERG does not consider this to be of great concern

- Use of a single overarching OS curve means that the proportion of deaths occurring for patients still on treatment versus patients off treatment cannot be separately calculated. The company's model assumes that the rate of death for third-line patients is identical to that of the whole population (i.e. second- and third-line patients). This limitation only affects the estimation of medical resource use costs incurred in the third-line setting (which have a limited impact on the model results)

- The model assumes that patients continue their third-line therapy after discontinuation of nivolumab or a taxane until death. Should any active intervention be used in the third-line setting, this assumption may lead to an over-estimation of costs incurred. The company's base-case analysis predicts more time spent receiving a third-line treatment for nivolumab versus a taxane ([redacted] versus [redacted] years, respectively), and so the incorporation of any active third-line treatment costs would, *ceteris paribus*, lead to an over-estimate of the ICER for nivolumab versus taxanes (when considering BSC reflects current practice for the UK)

- No adjustment to efficacy has been made for any beneficial effects of active third-line therapy. This is because a single over-arching OS curve has been used to estimate the number of deaths per model cycle. At clarification stage, the ERG asked the company to clarify which proportion of patients received treatment beyond progression in ATTRACTION-3 (clarification question A9). 39.0% of the nivolumab arm received treatment

after progression for a median of 32.5 days*, versus 1.4% of taxanes patients for a median of one day. Consequently, the benefits associated with nivolumab may be over-estimated (even if by a small quantity) owing to the specification of a survival model which masks potential benefits accrued concerning the use of treatment(s) after discontinuation of nivolumab (which are not considered standard UK practice). Subsequent therapy costs are discussed further in Section 4.2.8

In addition to these structural limitations associated with the PartSA structure, the ERG highlights that the CS urges caution when considering the outcome of disease progression owing to the potential role of pseudo-progression in response to checkpoint inhibitor therapy (CS Section B.1.3.5.1). In spite of this potential issue, the company developed a progression-based model within its submission. The ERG notes that this has also been the case for a number of previous checkpoint inhibitor appraisals conducted by NICE (that is, potential issues with progression are highlighted yet a progression-based model was submitted nevertheless).

To understand the potential impact of progression within the company's model, the ERG undertook an exploratory analysis wherein the PFS curve was assumed to be equal to the OS curve (essentially disabling progression events in the model). This analysis found that progression only affects the total estimated QALYs in the company's base-case analysis, and that the total costs incurred and life-years accrued are unaffected by the choice of PFS curve. As such, while the model is (by definition of the health states) 'progression-based', disease progression only affects the application of utilities and does not have an explicit impact on modelled costs or life-years (in the company's base-case analysis).

4.2.3. Population

The company stated that its model considers people with unresectable advanced, recurrent or metastatic OSCC that is refractory or intolerant to fluoropyrimidine and platinum-based combination therapy (CS Section B.3.2.2). This population is aligned with the company's anticipated licensed indication. As the model is predominantly constructed around the population of ATTRACTION-3,[2] it is important to acknowledge several important features of the

* The ERG notes that it is unclear how much of the subsequent treatment comprises of continued treatment with nivolumab beyond progression versus initiation of another line of therapy (e.g. subsequent use of taxanes or other anti-cancer therapies).

ATTRACTION-3 study that affect the generalisability of the modelled patient population to those anticipated to be treated in NHS practice.

The inclusion criteria for ATTRACTION-3[2] stipulated that patients must have an ECOG PS of 0 or 1, and a life expectancy of at least three months. As a direct consequence of these criteria, it is expected (as with many other clinical trials conducted in advanced cancer populations) that the trial population is generally fitter than the UK OC population, which would include those with an ECOG PS >1 and those with a life expectancy shorter than three months. This is acknowledged by the company within its submission, wherein it is stated *"Individuals randomised into clinical trials are likely to be slightly younger and healthier than the overall oesophageal cancer patient population in the UK."* (Section B.3.3.2.2).

Clinical advice provided to the ERG suggested that patients would only be considered as candidates for systemic anticancer therapy (either with a taxane or nivolumab) if they had an ECOG PS of 0 or 1. In addition, clinical advice suggested that some patients may opt for treatment with nivolumab but would otherwise decline to receive a further line of chemotherapy (i.e. a taxane). Based on this, the ERG considered the modelled population to be reflective of a taxane-eligible, ECOG PS 0-1, second-line subgroup of the broader OSCC population.

ATTRACTION-3[2] was conducted in a predominantly Asian population (less than 5% of the ITT population were non-Asian), of which the majority of patients resided in Japan (approximately 65% of the ITT population). For some gastro-oesophageal cancers, geographical region can have an important impact on histology and prognosis (for example, related to regional diets); though clinical advice provided to the ERG confirmed that this issue is not expected to have a large impact on outcomes for patients with advanced OSCC. However, the ERG notes that the pan-Asian ESMO guidelines highlight that in Japan, oesophageal cancer is diagnosed histologically according to the Japanese Classification of Oesophageal Cancer, and that differentiating between poorly and undifferentiated histological subtypes has both prognostic and clinical significance.[10]

OSCC patients are potentially managed with an escalated version of standard treatment available to patients in Europe (and more specifically, the UK). For example, the pan-Asian ESMO guidelines[10] indicate that there are five different systemic treatment options that are currently approved or reimbursed in Japan, including nedaplatin (which, to the ERG's knowledge, is not available in Europe). In addition, it is well recognised that the Japanese general population have a longer life expectancy than Western populations. A clinical adviser to

the ERG commented that differences in population treatment mean that data from ATTRACTION-3[2] are unlikely to fully reflect outcomes that would be seen in an equivalent Western population, but outcomes are expected to be similar in terms of the way the disease behaves and how treatment may impact outcomes.

A review by Watanabe (2018)[30] reported that the overall five-year survival rate for OC ranges from 15% to 25% worldwide, yet in Japan, the five-year survival rates of male and female OC patients were estimated to be 36% and 44%, respectively. While these estimates are subject to several limitations (including, but not limited to, the inclusion of adenocarcinoma patients, multiple treatment lines, and naïve comparisons of different populations), these findings indicate that in general, Japanese patients may be expected to have improved survival compared with ROW.

The following (limited) information was found in the Clinical Study Report for ATTRACTION-3, and is discussed as well in Section 3.2.5.5:

Table 10: Comparison of overall survival in ATTRACTION-3 between Japan and ROW

[redacted]

Further commentary concerning the population from ATTRACTION-3[2] is provided in Section 3.2.4. In light of the limitations of the ATTRACTION-3 highlighted above (and elsewhere throughout this report), the ERG considered the trial-estimated effect of nivolumab versus taxanes (e.g. in terms of the HR) may be generalisable to the anticipated population of patients to be treated in NHS practice. However, the ERG considered the ATTRACTION-3[2] population to be relatively fitter at baseline (mostly due to a combination of their ECOG PS, life expectancy, and differences in region). Therefore, any benefits associated with nivolumab versus taxanes from the trial estimated on an absolute scale (e.g. incremental life-years) are likely greater than the equivalent outcomes which would be seen in clinical practice.

The ERG also notes that allocation of specific taxanes (the comparator arm in ATTRACTION-3,[2] described further in Section 4.2.4) was determined prior to allocation of nivolumab or taxanes. This means that it would be theoretically possible to consider a comparison of patients considered suitable for treatment with docetaxel or paclitaxel, which would enable an assessment of how similar outcomes were for nivolumab-treated patients that were deemed suitable for each taxane. These analyses were not provided (Section 3.2).

4.2.4. Interventions and comparators

The company's model considered nivolumab monotherapy as the intervention under consideration, administered intravenously at a fixed dose of 240 mg by intravenous infusion over 60 minutes every two weeks until treatment discontinuation. Treatment with nivolumab is expected to be continued until disease progression (based on CS Table 2); however, in ATTRACTION-3,[2] treatment with nivolumab (or its comparator) could be continued after documented disease progression provided the following criteria were met (CS Section B.2.6.1.3.1):

- No rapid disease progression and the continuation of study treatment is expected to lead to clinical benefits
- Treatment was tolerated

- A stable ECOG PS Score
- Continuation of study treatment will not cause a delay of any prophylactic intervention for serious complications associated with disease progression (such as brain metastasis)

The comparator to nivolumab included within the model is 'taxanes': a blended population comprising patients treated with either docetaxel or paclitaxel. This corresponds to the comparator arm in ATTRACTION-3,[2] where the choice of either docetaxel or paclitaxel was determined by the treating clinician (as confirmed in response to clarification question A8).

Clinical advice provided to the ERG suggested that both paclitaxel and docetaxel may be used in UK practice, though each regimen is considered to have similar efficacy. In general, it was felt that docetaxel would be the preferred choice for most patients, owing to the fact that it is administered less frequently (i.e. once every three weeks instead of once per week). However, due to potential issues with tolerability, some patients may instead be treated with weekly paclitaxel which is considered to have a more favourable safety profile.

In addition, the company presented an exploratory comparison to BSC. The studies and methods used to elicit an indirect comparison to BSC are described further in Section 3.3 and Section 3.4. By virtue of the ATTRACTION-3[2] study comparing nivolumab to taxanes, all patients included within the ITT population may be considered eligible to receive chemotherapy. The company highlights within its submission (Section B.3.2.4) that BSC is *"only a valid comparator in patients unable to receive alternative therapies"*. As such, the population enrolled within ATTRACTION-3[2] may be considered inappropriate for comparison to BSC.

The final scope issued by NICE also included irinotecan as a comparator. In the CS, it is stated that clinicians consulted by the company felt strongly that irinotecan would not be used in the UK setting for treatment of second-line OSCC (CS, Document B, Table 1). The company cited market research conducted by the company which revealed that irinotecan comprises a 6% market share of current second-line treatments.

The ERG's clinical advisers noted that while irinotecan could theoretically be used in this setting (and indeed is used in other European countries), there is relatively little evidence available concerning its use and so taxanes would remain the primary choice of active second-line therapy. Importantly however, one advisor to the ERG noted that irinotecan use could be reserved for the third-line setting for when taxanes have failed yet patients are still considered fit enough to tolerate another line of therapy.

The ERG agreed with the company's decision to focus predominantly on a comparison to the taxanes arm of ATTRACTION-3,[2] given that this is generally aligned with the treatment pathway in the UK, and makes use of the full ITT population of the study. However, the ERG also considered it important to acknowledge that it is expected that the majority of patients in the UK would opt for treatment with docetaxel instead of paclitaxel, and so any differences in the modelled costs for each of these taxanes may be important to consider.

The value of 6% is indicative of very limited use of irinotecan in UK practice, especially in light of an estimated population size in the region of 266 second-line eligible patients per year (CS, Section A.18). However, the ERG acknowledged that irinotecan was included as a comparator in the final scope, and that by definition of this value being greater than zero the market research confirms some (though few) patients may be treated with irinotecan. Nevertheless, the ERG considered the company's rationale for excluding irinotecan as a comparator to be justifiable based on the lack of evidence to robustly inform cost-effectiveness estimates, and the small number of patients that would be treated with irinotecan in practice.

The ERG also agreed with the company's position that the comparison made to BSC should be considered as a scenario analysis only, owing to the study design of ATTRACTION-3[2] and associated logic in terms of eligibility for treatment (as described in Section 4.2.3). Any inferences of the cost-effectiveness results for nivolumab versus BSC are inherently flawed based on the differences in the population of patients eligible for nivolumab (as part of ATTRACTION-3) and the population that would receive BSC in NHS practice.

4.2.5. Perspective, time horizon and discounting

The company's economic model adopted an NHS and PSS perspective on costs and outcomes (QALYs and LYs), which were discounted at 3.5% per annum. The ERG was satisfied that the perspective and discounting adopted by the company's model are both aligned with the NICE reference case.

A time horizon of 40 years was used to inform the company's base-case analysis, which the ERG notes that while certainly sufficient to capture all relevant costs and effects, is arguably excessive for patients with advanced, unresectable OC. The choice of survival curve extrapolation is therefore especially important, as any residual, prolonged plateau in survival may have an important effect on the estimation of modelled life-years.

In the company's base-case analysis, a log-logistic model is used to inform the nivolumab arm, which is associated with a decreasing hazard of death in the longer term (and thus, an emergent plateau in the curve). The choice of extrapolation has a highly-influential effect on cost-effectiveness results, and is discussed further in Section 4.2.6 of this report. However, in relation to the time horizon, the ERG highlights that the combination of a decreasing long-term hazard of death and an arguably excessive time horizon has the potential to yield unrealistic mean life-year estimates.

As a part of the development of the ERG's report, two practising oncologists independently confirmed that the majority of patients undergoing second-line taxane therapy in current UK practice would have an estimated survival of less than 12 months. As such, it may be considered reasonable that by 10 years (that is, ten-times the maximum life expectancy for most patients with current care), the majority of relevant costs and effects would be captured by the model.

4.2.6. Treatment effectiveness and extrapolation

4.2.6.1. Modelling approach

Data from the ATTRACTION-3 study constitute the primary evidence base from which estimates of treatment effectiveness are made to inform the economic model. In ATTRACTION-3,[2] patients were randomised to receive treatment with nivolumab or a taxane (docetaxel or paclitaxel). Further information concerning ATTRACTION-3[2] (including the ERG's critique) is provided in Section 3.2 of this report. However, for the purpose of time-to-event outcome estimation, no adjustments are made to the ATTRACTION-3[2] data to account for potential differences between the trial population and the NHS patient population.

Three time-to-event outcomes are used to inform the economic model:

- **Overall survival (OS):** the proportion of patients who are alive at each model cycle, regardless of disease progression status
- **Progression-free survival (PFS):** the proportion of patients who are alive with non-progressed disease at each model cycle
- **Time on treatment (ToT):** the proportion of patients who are still receiving their second-line treatment (i.e. nivolumab or taxane), regardless of disease progression status

In addition to these time-to-event outcomes, the model also includes HRQoL and AE data from ATTRACTION-3.[2] These data do not affect health state transitions, and so are discussed separately in Section 4.2.7 (HRQoL) and Section 4.2.8 (resources and costs) of this report, respectively.

For each of the time-to-event outcomes, survival modelling methods were used to extrapolate over the lifetime horizon of the model. As highlighted in NICE DSU TSD 19, when extrapolation of the trial evidence is required to appropriately inform cost-effectiveness, PartSA models should easily facilitate the investigation of alternative assumptions in accordance with current NICE methods guidance.[29] The company's economic model adopted a time horizon equivalent to 40 years, and at the end of follow-up in ATTRACTION-3[2] (at approximately 34 months), there were approximately 17% (nivolumab) versus 5% (taxane) of patients still alive. The PFS and ToT curves were also incomplete, and so extrapolation of time-to-event data was necessary to inform the economic model.

In this section, a summary of the broad analytical approaches taken is described. The models related to specific endpoints are discussed in turn in the relevant sections that follow.

Two separate approaches were taken to inform the model in the original CS:

- **Fully-parametric (FP), standard parametric models:** A parametric model was fitted to the time-to-event data from ATTRACTION-3[2] (separately for each treatment arm)
  - Six parametric models were considered: exponential, Weibull, Gompertz, lognormal, log-logistic, and generalised gamma
- **Semi-parametric (SP) models:** The Kaplan–Meier curve was used to inform the time-to-event outcome up until a given cut point, after which the remainder of the time-to-event outcome was informed by a parametric model fitted to data from ATTRACTION-3. [2] This approach is often also termed a piecewise model
  - Like the standard parametric models, these models were fitted independently by treatment arm. However, events and censored observations before the cut point were discounted from the parametric component (i.e. the parametric models were fitted to 're-based' data, where $time_{re-based} = time_{original} - cut\ point$)

The CS also acknowledged that a third approach (spline models) was also considered, but these models were not presented. In addition, SP models with only one cut point were originally

provided. At clarification stage, the ERG requested SP models with alternative cut points and spline models be provided by the company for consideration. The company provided a range of alternative models based on this request, allowing for a more in-depth assessment of the impact of different models on cost-effectiveness results.

Model selection was based on the following considerations highlighted by the company:

- **Statistical goodness-of-fit:** Akaike's and Bayesian Information Criteria (AIC and BIC, respectively) were used to indicate goodness-of-fit to the ATTRACTION-3[2] data while incorporating a penalty related to model complexity
  - The ERG understands that statistical goodness-of-fit scores cannot be compared across the FP and SP approaches, as the parametric components of these models were (by definition) not fitted to the same data. This means that the relative additional complexity of specifying an SP model versus an FP model cannot be assessed using these traditional measures
- **Visual assessment of fit:** The Kaplan–Meier plot was compared to the fitted models to (informally) establish how well they fitted the data from ATTRACTION-3[2]
- **Inspection of log-cumulative hazard plot:** The log-cumulative hazard of a given event was plotted over time to guide the model selection and viability of models

The company explained within its submission that the Kaplan–Meier plots for all outcomes (most notably, PFS and OS) demonstrated a high initial hazard of experiencing a given event (e.g. death) during the initial study period (i.e. shortly after initiating treatment with either nivolumab or a taxane). The company speculated that this may be due to *"the high mortality impacting patients with oesophageal cancer"* (CS, Section B.3.3.2.1.1). After the initial study period, the company explains that lower hazards of experiencing a given event may be inferred from data for both treatment arms. Consequently, the company explained that FP models were not capable of reflecting such a change in hazards, yet an SP approach may better reflect the high initial hazard (while also making use of the 'maximum amount of data' to inform the parametric component to extrapolate over the model time horizon).

In order to specify an SP modelling approach, it is necessary to identify a relevant 'cut point' – i.e. the time at which the model switches from using the Kaplan–Meier curve directly to a parametric model fitted to the remainder of the Kaplan–Meier curve. Ultimately, the company's

SP models each considered a cut-point of 2.99 months, based on the following explanation provided in Appendix M:

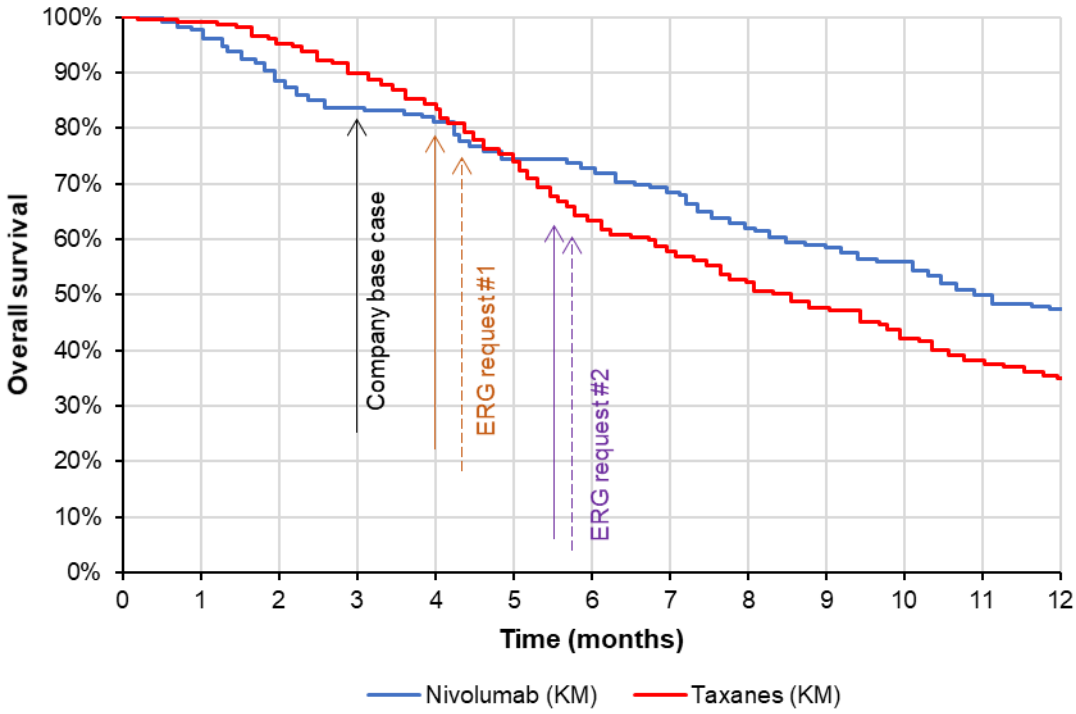
> *To enable consistency and transparency, it was determined that a cut point at 2.99 months should be used based on the following rationales:*
>
> - *This time point exceeded the mean and median time to response in all treatment arms, so that the majority of responses will already have occurred and any separation in long-term outcomes would be reflected in the data.*
> - *The point of maximal rate of hazard change would have occurred in all treatment arms, so that parameterisations would be able to reflect the ongoing change in hazard.*
> - *The time point is not impacted by assessment periods.*

As described previously, in the original submission SP models with only one cut point (2.99 months) were provided. It is generally recognised that the choice of any specific cut point has the potential to lead to markedly different extrapolations of survival, and the use of an SP approach requiring selection of a cut point has been the subject of debate in a number of previous assessments conducted by NICE, including those of other checkpoint inhibitor treatments.[31,32]

Davies *et al.* (2012)[33] highlighted using a case study in advanced melanoma that *"selection of a suitable point on the* [Kaplan–Meier] *function from which to extrapolate becomes increasingly arbitrary as the effective sample size decreases*".[33] While certainly a limitation of the SP approach in general, the company's selection of a cut point where there are still a substantial proportion of patients still at risk means that the impact of this is limited.

In response to clarification question B10, the company provided SP models with two alternative cut points. Originally, the ERG requested cut points of approximately 4 and 5.5 months – these two time points correspond to a time point avoiding the short period for the outcome of OS in the nivolumab arm where no events are experienced (after 4 months) and a time point after the period over which the curves for each arm cross (after 5.5 months). The company explained within its response that as some patients may be assessed for progression ±7 days either side of a planned assessment point, it is ill-advised to select a cut point that falls within this period, and so instead provided analyses where the cut points were set to 4.37 and 5.75 months. The ERG agreed with this modification to the initially-requested analyses, and Figure 3 shows where these cut points correspond to the OS Kaplan–Meier curves.

Figure 3: Alternative cut points specified in SP models provided to the ERG versus OS



Abbreviations: KM, Kaplan–Meier.

Note(s): Figure produced by the ERG. The ERG requested two alternative cut points of 4 and 5.5 months (#1 and #2, respectively, shown in the diagram as solid arrows). The company explained within its response that as some patients may be assessed for progression ±7 days either side of a planned assessment point, it is ill-advised to select a cut point that falls within this period, and so instead provided analyses where the cut points were set to 4.37 and 5.75 months (shown in the diagram as dashed arrows)

Source(s): Kaplan–Meier curves digitised based on reporting in the pivotal trial publication by Kato *et al.*, (2019).[2]

In consideration of an SP approach, the specification of this type of model assumes that there are some time periods wherein the implied hazard of death is zero (where the Kaplan–Meier curve is horizontal), and other times where the implied hazard of death is infinite (where the Kaplan–Meier curve is vertical). This may have unusual impacts on the estimation of longer-term survival, as depending on where the parametric component is joined to the Kaplan–Meier curve, there could have been a period of no events and/or recent event(s) meaning the curve is initiated at a higher or lower starting proportion than would be realistically expected (e.g. based on an FP approach). The ERG highlighted that this potential issue is partially noted within the company's response to clarification question B10, but remains a limitation nonetheless.

As described previously, the company provided additional spline-based models in response to clarification question B10. The ERG requested a spline-based analysis using one internal knot,

to allow for an alternative means of addressing the potential change in hazards based on the initial portion of the Kaplan–Meier curve versus the longer-term estimation of survival. The company provided two alternative spline-based models – one where the internal knot was manually placed at 2.99 months (per the cut point in the SP approach used in the company's base-case analysis), and another where knot location was set to the default positioning (based on the median uncensored survival time for each arm).

While the company provided spline-based models as a sensitivity analysis, the company cautioned their application to data from ATTRACTION-3.[2] More specifically, in its response to clarification question A10, the company cautions against spline-based models to *"model data with clear structural discontinuities, as this imposes conditions of smoothness on the hazard that are observed to be inconsistent with the data and can result in oscillation of the predictions and poor long-term performance due to the rapidly changing derivative of the hazard function."*

The ERG disagrees that there is a *"clear structural"* discontinuity present in the data from ATTRACTION-3,[2] and that while inferences concerning the pattern of hazards may be made, hazards are not observed and thus spline-based models cannot be shown to be definitively consistent or inconsistent with data from the trial. While the explanation provided above relates to the application of a spline-based model, the ERG notes that the same argument would technically also apply to the estimation of an FP, non-spline-based model – that is, that the *"rapid change in hazard"* in the initial period may not be appropriately estimated using a parametric function to smooth the estimated hazards in this period.

In addition, the ERG did not consider the fitting of spline-based models to data from ATTRACTION-3[2] to be subject to substantial risk of *"oscillation of the predictions"*, given that the company has elected to fit a hazards-based spline with only one internal knot (which, as noted in the company's response to clarification question A10, may be considered an extension to the standard Weibull model). However, the ERG agreed that (as with all FP or SP models), longer-term predictions may be poor, and so the plausibility of longer-term extrapolations should be carefully considered.

The ERG accepted that there was a potentially-important 'elbow' in the OS Kaplan–Meier curve for nivolumab just before three months (which may be related to the mechanistic properties of nivolumab, and is shown in Figure 3). However, it did not agree that this was sufficient evidence in itself to overtly reject an FP model (either based on standard parameterisations or splines). Equivalently, the ERG does not consider the taxanes OS curve to be affected by the same

'elbow', and so the specification of an FP approach for the taxanes arm in particular would not seem unreasonable.

In choosing between an FP (spline-based or otherwise) versus an SP approach, NICE TA525[32] guidance (atezolizumab for treating locally advanced or metastatic urothelial carcinoma after platinum-containing chemotherapy) provided information in a similar context where this decision was also required (for a checkpoint inhibitor versus taxanes). In TA525, the company used an FP generalised gamma model to estimate OS for patients on both treatment arms of the pivotal IMvigor 211 study. However, the ERG and committee ultimately preferred an SP approach using the Kaplan–Meier curve with the tail extrapolated from the point when 20% of patients were still alive.

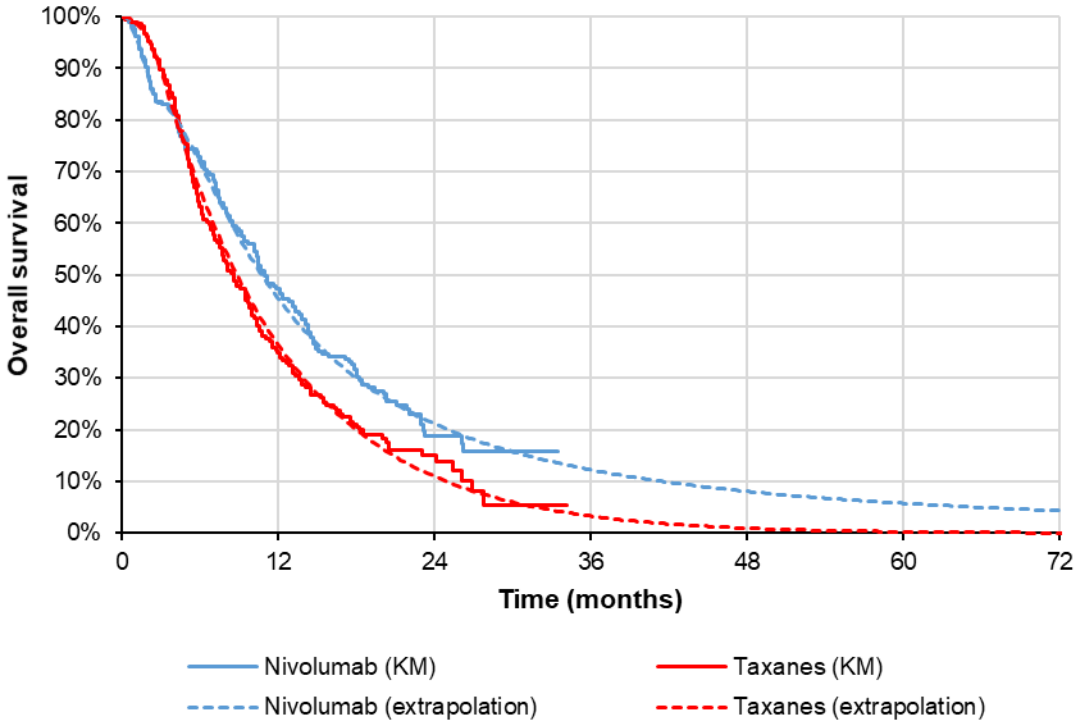
The decision to move away from an FP approach to an SP approach in TA525[32] was based on the estimated survival proportions at later time points (more specifically, five-year OS for the taxanes are in particular), as opposed to relative improvements in fit within the observed period. Based on this, the ERG considered it important to acknowledge that both approaches may provide reasonable fits to the Kaplan–Meier curve (especially the SP approach, given part of the model is directly based on the Kaplan–Meier curve itself), but it is the plausibility of the extrapolated portion which is expected to be the key determinant of optimal model selection.

4.2.6.2. Overall survival

In the company's base-case analysis, an SP approach was adopted to inform the estimation of OS for both treatment arms. In this approach, the Kaplan–Meier curve was used until 2.99 months followed by a parametric model for the remainder of the model time horizon. For the taxanes arm, an exponential model was applied, whereas for the nivolumab arm a log-logistic model was applied (both after 2.99 months).

The company's base-case projections of OS are presented in Figure 4, which the ERG produced by overlaying the projections from the economic model on top of a digitised copy of the Kaplan–Meier curves reported in the pivotal trial publication.

Figure 4: Superimposition of company's base-case OS projections and Kaplan–Meier curves from ATTRACTION-3



Abbreviations: KM, Kaplan–Meier.

Note(s): Figure produced by the ERG. These projections include the company's adjustment to account for background mortality.

Source(s): Kaplan–Meier curves digitised based on reporting in the pivotal trial publication by Kato *et al.*, (2019).[2]

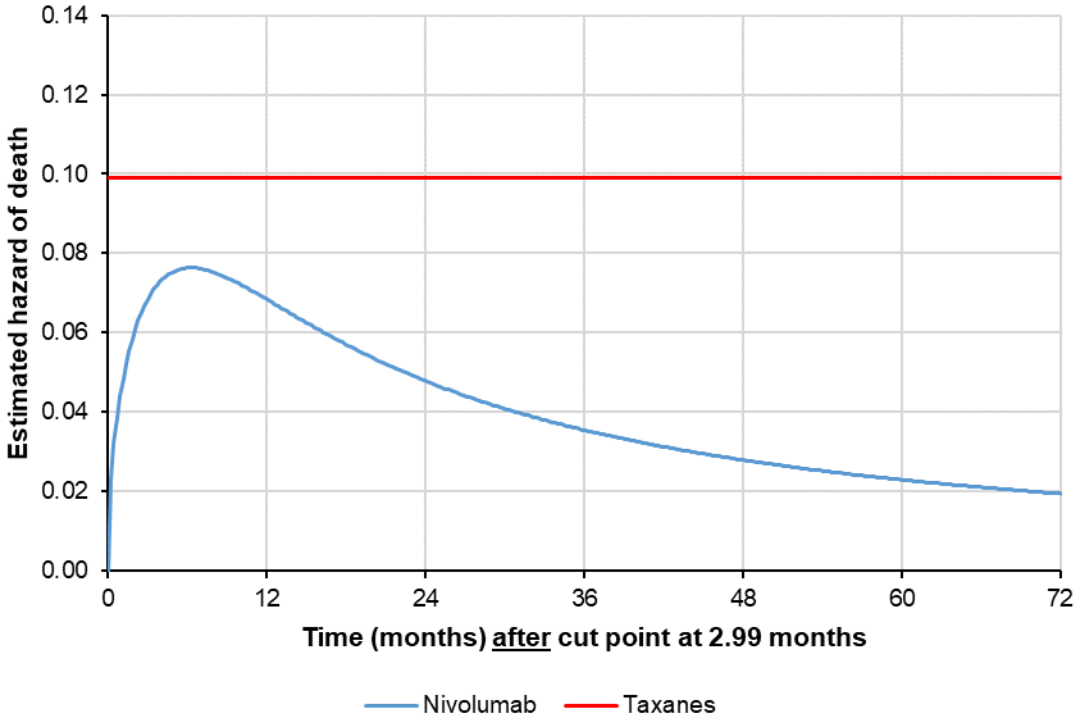
The base-case extrapolation of OS yields a reasonably good visual fit to the Kaplan–Meier curve after the cut point (at 2.99 months) for the nivolumab arm. However, the specification of an SP log-logistic model is not without limitations. The log-logistic model is associated with a non-monotonic hazard function, wherein there is an initial "peak" in the hazard of death, after which the hazard consistently decreases indefinitely. Using information provided in Appendix M of the CS (more specifically, Figure 38 of Appendix M), the ERG was able to re-create the assumed pattern of hazards applied after the cut point, using the following formula for the hazard function of the log-logistic distribution:

$$h(t) = \frac{\left(\frac{shape}{scale}\right) \times \left(\frac{time}{scale}\right)^{shape-1}}{1 + \left(\frac{time}{scale}\right)^{shape}}$$

The base-case projection of OS for the taxanes arm was based on an exponential model, which assumed a time-invariant hazard of death. For the taxanes arm, the fitted exponential model did not appear to provide a good visual fit to the Kaplan–Meier curve, which can be seen by the extrapolation sitting slightly above the Kaplan–Meier curve between approximately six and 12 months, before crossing the Kaplan–Meier curve at around 15 months, and then under-estimating the Kaplan–Meier curve until approximately 27 months. This pattern of over-estimation followed by under-estimation points to the fact that a constant hazard rate is likely too simplistic in order to fully reflect the pattern of survival seen for taxane-treated patients. Again, using information reported in Appendix M, the ERG identified the equivalent pattern of hazards (i.e. a constant hazard rate) for this model.

The resultant plots of hazards over time (relative to timing of the cut point) are presented in Figure 5 for both the nivolumab and taxanes arms. Figure 5 shows that the hazard of death for the nivolumab arm is assumed to increase up until approximately six months after the cut point (equivalent to 8.99 months from baseline), after which the hazard of death decreases indefinitely. For the taxanes arm, the hazard of death is consistently greater than the nivolumab arm, and does not vary over time (resulting in a horizontal line).

Figure 5: Pattern of hazards exhibited by company base-case projections of overall survival



Note(s): Figure produced by the ERG. These projections do not include adjustment to account for background mortality, and are based entirely from reported curve fit parameters in Appendix M of the CS.

The ERG considered the modelled pattern of hazards to be an important consideration, particularly when determining the most appropriate parameterisation of OS, as the specification of a log-logistic component in the nivolumab model may artificially impose a peak in hazards after the initial period of higher hazards already captured by the Kaplan–Meier component. The CS does not provide any explicit justification for why a later peak in hazards would be expected, and instead focuses on the general fit of the different models both to the Kaplan–Meier curves and based on a log-cumulative hazard plot.

The log-cumulative hazard plot presented in the CS for the taxanes arm illustrates that the exponential model provides one of the poorer representations of hazards (along with both the Weibull and Gompertz models) compared to the generalised gamma, log-logistic, and lognormal models (Appendix M, Figure 41). Also, as cited previously, other appraisals in similar contexts assumed a non-constant hazard rate for patients receiving taxanes (e.g. TA525, wherein a generalised gamma model was selected).

In addition to the specification of alternative survival models, the company adjusted estimates of OS to account for background mortality. The CS explained the rationale for this approach based on the following:

> *"Individuals randomised into clinical trials are likely to be slightly younger and healthier than the overall oesophageal cancer patient population in the UK. The mean age of patients in ATTRACTION-3 is 63.8 years, increasing the likelihood that most deaths observed over the trial period were cancer-related. Therefore, the model includes age and gender-adjusted mortality based on information from UK life tables…"* (CS Section B.3.3.2.2).

While the ERG agreed that the majority of deaths that occurred within the ATTRACTION-3[2] study are likely cancer-related, it was not possible to identify if the proportion of non-cancer related deaths was broadly generalisable to the UK patient population that would be considered eligible to receive nivolumab in NHS practice.

To account for background mortality, the company's model estimated the all-cause mortality risk per cycle in order to produce a background mortality OS curve. Background mortality rates were sourced from published Office for National Statistics Life Tables.[3,4] After this, the product of the OS curve estimated from the trial and the background mortality OS curve evaluated at each model cycle is taken to produce an adjusted OS curve, shown in the equation below:

$$OS_{Total} = OS_{Trial} \times OS_{Background}$$

The company acknowledged within its submission that while some double counting occurs through the use of this approach, the effect applies to both arms and is therefore *"likely to have a minimal impact on predicted survival (and hence cost-effectiveness)"* (CS Section B.3.3.2.2).

Through this specific application of background mortality, the ERG noted (all other things equal) an over-estimate of the ICER would be produced, due to nivolumab being associated with greater OS (after the initial four months) and thus the multiplicative effect on background mortality will have a relatively greater impact versus the taxanes arm. This was especially noteworthy in the company's base-case analysis, as the log-logistic component of the SP model applied for nivolumab leads to a greater proportion of patients that are alive in the longer-term (and are thus subject to higher background mortality risk owing to their relative age). Equivalently, no adjustment for background mortality would lead to an under-estimate of the ICER (as the log-logistic component of the SP model for nivolumab is associated with decreasing hazards in the longer-term, that do not take into account increasing risk of death from other causes as patients age).

The company's model included the ability to disable background mortality adjustment. Consequently, the impact of enabling or disabling background mortality adjustment is explored within the ERG's additional analyses (Section 6.2). However, for the purpose of informing the ERG's preferred base-case, the application of background mortality is unchanged from the company's base-case analysis. This is because as described previously, the ERG expected outcomes in the trial to be better than those achieved in NHS practice, and therefore a small over-estimation of background mortality is likely to lead to survival models that are slightly closer to the outcomes that would be seen in a UK population. In addition, the ERG explored an alternative application of background mortality within its exploratory analyses (also reported in Section 6.2).

Based on the reasons provided above, the ERG did not consider the combination of the company's base-case projections of OS for the nivolumab and taxanes arm to be the most appropriate estimates to inform the model. However, the ERG noted that the provision of a range of survival models allowed for alternative approaches to be considered, and commended the company for providing a broad range of FP and SP approaches to allow exploration of alternative settings and assumptions.

As is often the case with PartSA models of late-stage cancer treatments, the choice of model for the outcome of OS greatly influenced the modelled total costs and QALYs, and thus the ICER. The ERG's preferred base-case analysis includes the specification of an SP generalised gamma model, using the cut point at 5.75 months for both treatment arms. The impact of the ERG's preferred models for the outcome of OS (as well as PFS and ToT) on the modelled ICER is described in Section 6.3 of this report, alongside the reasons behind why this model was considered most appropriate.

4.2.6.3. Progression-free survival

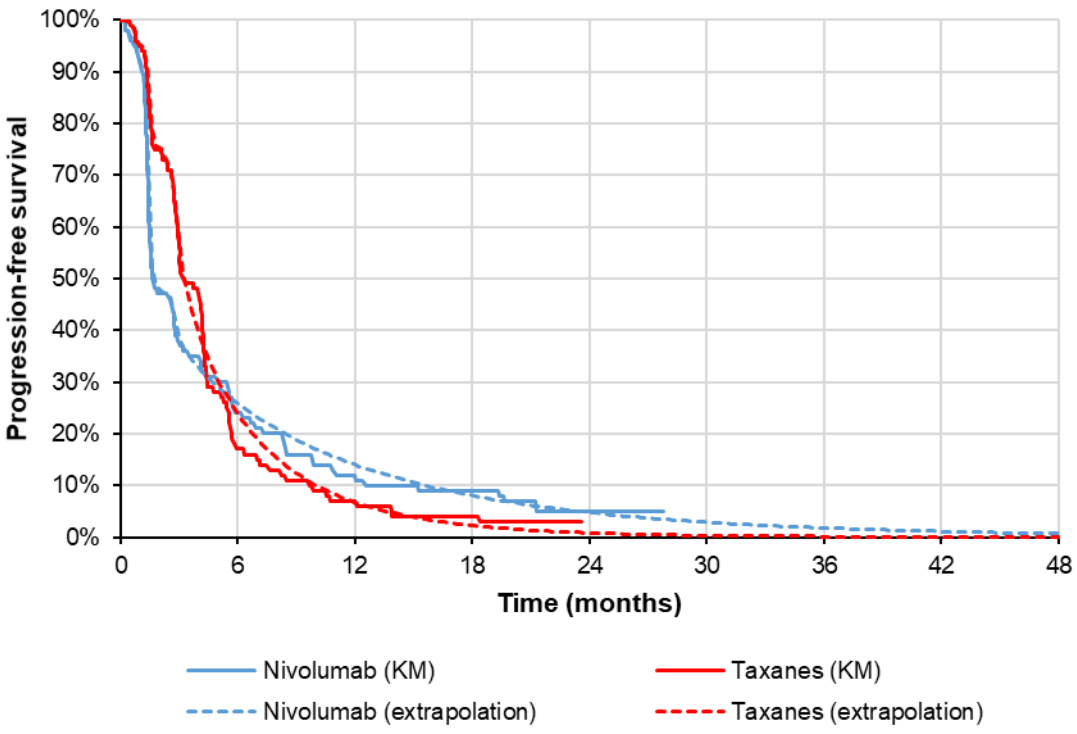
As described in Section 4.2.2, while the company's economic model structure is progression-based, the estimation of a PFS curve for each treatment arm does not have a large overall impact on the estimated ICER. This is because the PFS curve only affects the allocation of utility (i.e. the proportion of 'alive' patients assigned a 'progression-free' or 'progressed' value).

In the company's base-case analysis, an SP approach is specified for both treatment arms, where the Kaplan–Meier curve is used up until the cut point of 2.99 months, after which a Weibull model is fitted to estimate PFS for the remainder of the time horizon.

It is important to acknowledge that the assessment of progression is affected by the timing of progression assessments in ATTRACTION-3.[2] As such, it may be expected that the PFS curve from the trial represents an over-estimate of the 'true' PFS curve (as patients will have progressed at a point in time less than or equal to the time at which progression is recorded). In spite of this, as PFS is a composite outcome of both progression and death events, and the limited impact PFS has on the modelled ICER, the ERG did not consider it necessary to adjust the PFS curve to account for potential discordance between the timing of progression events versus when progression events were recorded.

The company's base-case projections of PFS are presented in Figure 6, which the ERG produced by overlaying the projections from the economic model on top of a digitised copy of the Kaplan–Meier curves reported in the pivotal trial publication (as the CS did not contain a figure which only presented the base-case projections).

Figure 6: Superimposition of company's base-case PFS projections and Kaplan–Meier curves from ATTRACTION-3



Abbreviations: KM, Kaplan–Meier.

Note(s): Figure produced by the ERG.

Source(s): Kaplan–Meier curves digitised based on reporting in the pivotal trial publication by Kato *et al.*, (2019).[2]

While the choice of PFS model is expected to have only a limited impact on the ICER, the ERG considered the base-case models to provide a relatively-poor fit to the Kaplan–Meier curves. The ERG acknowledges that any parametric (either FP or SP) approach will not be able to fully reflect the protocol-driven 'bumps' in the Kaplan–Meier curve; however, the base-case models fitted to both arms may be considered to provide an over-estimate of time spent in progression-free survival (most notably, between six and 12 months). As described above, the ERG considered the PFS curve from ATTRACTION-3[2] to represent an over-estimate of the 'true' PFS curve, and therefore a more conservative estimate of PFS may be more suitable to inform the model (to partially address the 'bumps' in the curve).

In spite of the above, the ERG's preferred base-case analysis includes the specification of the same models for the outcome of PFS as per the company's base-case analysis. This is because the choice of PFS model has relatively little impact on the estimation of the ICER, and that any approach is unlikely to provide a good representation of the Kaplan–Meier curve owing to the protocol-driven 'bumps' noted previously.

4.2.6.4. Time on treatment

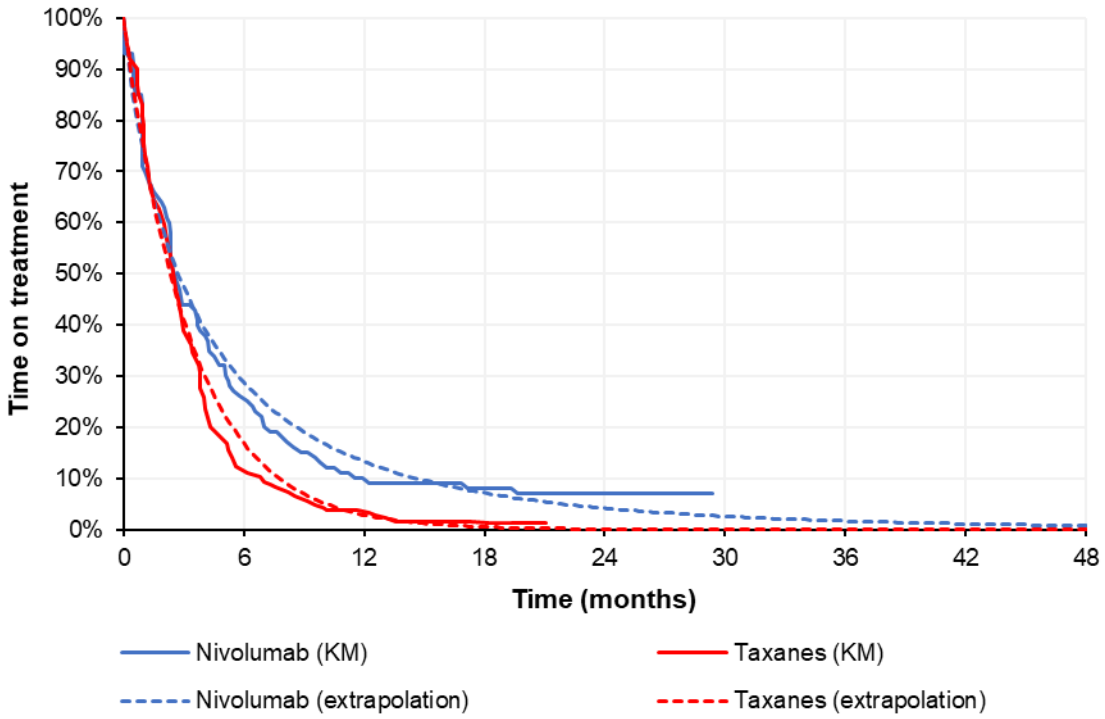
The company's economic 'progression-based' model structure is heavily reliant upon the determination of an appropriate extrapolation of ToT. This is because ToT is used to determine treatment acquisition and administration costs, which combined are responsible for the majority of the incremental costs associated with nivolumab versus taxanes. For this reason, appropriate estimation of ToT is of greater importance versus estimation of PFS (in terms of impact on the model results).

The overall modelling approach for ToT is very similar to that of OS and PFS, in that a range of parametric models were fitted, and subsequently used to inform the proportion of patients that remained on second-line treatment over the model time horizon. However, unlike OS and PFS, the company did not provide SP models for the outcome of ToT in its original submission. Very limited explanation was provided within the CS concerning why SP models were not considered for ToT, which is presented below:

"The log-cumulative hazards for ToT show no reason to disregard a parametric model. Additionally, the data for ToT is relatively complete and so there is less emphasis on the importance of extrapolating outcomes. ToT extrapolation does not require as complex methods of fitting as the PFS and OS curves. As the criteria for parametric models are satisfied and the models fit the data well, there was no need to explore other model types." (CS Appendix M Section 4.3.4).

In the company's base-case analysis, an FP generalised gamma model was used for the nivolumab arm, whereas an FP exponential model was used for the taxanes arm. The company's base-case projections of ToT are presented in Figure 7.

Figure 7: Superimposition of company's base-case ToT projections and Kaplan–Meier curves from ATTRACTION-3



Abbreviations: KM, Kaplan–Meier.

Note(s): Figure produced by the ERG.

Source(s): Kaplan–Meier curves digitised based on reporting in the CS.

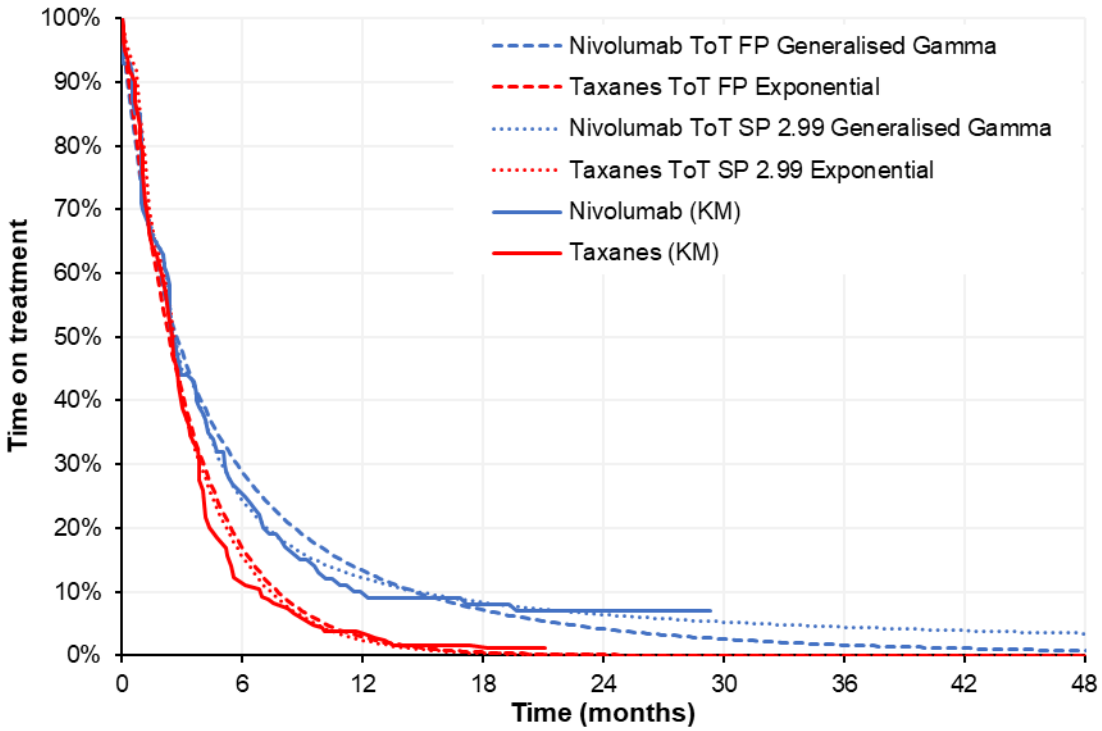
Unlike the choice of PFS model which is expected to have a minimal impact on the ICER, the choice of ToT model has a greater influence on the cost-effectiveness results. From Figure 7, it can be seen that neither of the base-case extrapolations provide a particularly good visual fit to the Kaplan–Meier curves. The ERG disagrees with the company's suggestion that the log-cumulative hazard plots show *"no reason to disregard a parametric model"*, as it is evident from

visual inspection of the company's preferred models (when compared to the Kaplan–Meier curves) that an alternative approach may be warranted given the poor fit of the FP models. However, as with the choice of PFS model, any parametric (FP or SP) approach will not be able to fully reflect all the protocol-driven 'bumps' in the Kaplan–Meier curve.

At clarification stage, the ERG requested additional models, including SP approaches and spline-based models for the outcomes of ToT. The company provided the SP models using the original cut point (2.99 months) and the additional cut points described in Section 4.2.6.1 (4.37 and 5.75 months). The spline-based models produced near-identical fits to the base-case FP generalised gamma models, thus are not discussed further.

To illustrate the differences between FP and SP models for ToT, the ERG has plotted variations of the company's base-case models in Figure 8 (i.e. same functional form as per the company base-case analysis for both arms, using the cut point of 2.99 months).

Figure 8: SP versus FP ToT projections (generalised gamma and exponential)



Abbreviations: FP, fully-parametric; KM, Kaplan–Meier; SP, semi-parametric; ToT, time on treatment.

Note(s): Figure produced by the ERG.

Source(s): Kaplan–Meier curves digitised based on reporting in the CS.

The SP exponential model provides a very similar fit to the FP approach, with similar longer-term extrapolations. For the nivolumab arm, the SP approach yields a better fit to the Kaplan–Meier curve, yet projects a substantial proportion of patients to continue treatment for two years or more.

In the CS, a scenario analysis was presented concerning a potential stopping rule at two years, though limited explanation was provided as to the expectation that this rule would be applied in practice. The ERG was aware that treatment stopping rules have been considered in a range of previous NICE assessments of checkpoint inhibitors, especially concerning treatment beyond two or three years. However, no explanation was provided in the CS concerning whether or not such a stopping rule is expected to be applied in practice, and the stopping rule was not included in the protocol of the ATTRACTION-3 study[35] (as may be inferred through the lack of events in the ToT curve around this time). Therefore, the ERG did not consider the stopping rule further to inform its base-case analysis.

The ERG's preferred analysis includes the specification of a SP Weibull model for both arms, using the cut point of 5.75 years. The impact of the ERG's preferred models for the outcome of ToT (as well as OS) on the modelled ICER is described in Section 6.3 of this report, alongside the reasons behind why this model was considered most appropriate.

Separately to the selection of the ERG's preferred base-case model for ToT, the ERG highlights that it may be helpful to consider the likely pattern of discontinuation based on how reliable estimates from the Kaplan–Meier curve may be until a given time point, and ultimately by which time point nearly all patients are expected to have discontinued treatment. The ERG's additional sensitivity analysis based on an alternative (pragmatic) SP approach is provided in Section 6.2.

4.2.6.5. Subgroup analysis

As an alternative comparison to the broad 'taxanes' group, the company's model included an option to compare nivolumab to each taxane individually. To do this, the company conducted subgroup analyses and separately fitted models for each time-to-event outcome (i.e. OS, PFS, and ToT). At clarification stage, the ERG requested additional analyses be conducted where taxane use was considered a covariate (clarification question B6). However, in response to clarification question A8, the company stated that: *"Investigator choice between paclitaxel and docetaxel was declared and documented in the randomization system (IWRS) prior to randomization".*

Based on this explanation (provided at clarification stage), the ERG did not consider comparisons between individual taxane arms and the full nivolumab arm to be appropriate. This is also discussed in Section 3.2.1.

Based on the above, the subgroup analyses by individual taxane assignment were not considered further, and so discussion of the difference between the extrapolation approaches is not presented here. However, for completeness, the ERG noted that the choice of approach to consider these groups (i.e. subgroup versus covariate based) yielded broadly consistent results, and results of the scenario analyses are presented in Section 5.2.3.

4.2.7. Health-related quality of life

4.2.7.1. Methodological approach (company base-case analysis)

Health-related quality of life (HRQoL) data were obtained via the EQ-5D-3L, collected in ATTRACTION-3. The periodic completion of EQ-5D questionnaires throughout the trial enabled the calculation of utilities for different progression states for each group. Utilities were derived from EQ-5D data using the standard methods of Dolan *et al.* (1997),[36] per the NICE reference case.

As is often the case in clinical trials, EQ-5D data from ATTRACTION-3[2] were collected with varying frequency over the duration of follow-up. The CS explained that completion of the EQ-5D was dependent upon treatment status, which was also closely associated with progression status. Treatment status could relate to whether or not a patient was still being treated, or which arm the patient was allocated regardless of whether or not treatment is being continued.

The CS stated that in order to *"allow fitting of a model assuming an AR(1) autocorrelation structure between observations upon a single patient"*, it is necessary to *"regularise"* the time period between observations (12-week periods). The ERG considered an important omission from the CS is a clear description of why this specific analytical approach to estimate utility values was taken, and therefore considered it important to highlight the following key aspects of the approach taken:

- In this context, the ERG understood the phrase "model" to apply to the general approach taken to impute missing data and thus ultimately generate utility values to populate the cost-effectiveness model

  - This type of “model” is not a regression model, which may be considered more conventional in the context of estimating utility values for use in economic evaluation. Ultimately, the utility values estimated for use in the economic model are simply based on mean values from an imputed data set

  - By considering a simple mean (by progression status) of the imputed data set, this approach is entirely reliant upon the accurate estimation of a large quantity of missing data, both within the observed period of follow-up and in the extrapolated period (which requires the estimation of future progression and death events)

- The ERG understood the phrase “regularise” to be analogous to enforcing the assumption that utility values must fall into specific bounds in order to be considered within the correlation structure being imposed

  - The need to adjust the utility data in this manner was, based on the ERG’s understanding, a direct consequence of (a) deciding that it is necessary to impute missing data, and (b) specifying an AR(1) correlation structure

- The 12-week period between observations was determined based on the lowest frequency of collection on trial

  - In ATTRACTION-3, EQ-5D data were collected every six weeks during the treatment phase, and every 12 weeks in the follow-up phase

  - Consequently, by only taking the observation closest to every 12-week period during the treatment phase, it may be inferred that as much as 50% of the utility data collected for patients on treatment are discarded from the company’s analysis (i.e. scheduled visits falling in the middle of 12-week periods)

Based on the description provided above, the ERG highlighted that this approach to estimating utility values for inclusion within an economic model is unconventional, and subject to several important limitations (as highlighted above). It is the ERG’s understanding that imputation of missing EQ-5D data is seldom undertaken, and as a result of this published guidance is limited.

Through a pragmatic literature search, the ERG identified a simulation study by Simons *et al.* (2015).[37] This study was performed to evaluate the impact of imputing individual domains versus imputing index values to deal with missing EQ-5D-3L data. The authors noted that in

practice (i.e. outside the controlled settings of a simulation study), missing data will be present for unobserved variables in the imputation model, and that this can pose a number of problems. In the case of ATTRACTION-3, the most notable examples of this are missing death and progression times. Furthermore, Simons *et al.* (2015)[37] found that in general, index imputation was more accurate that domain imputation for studies with smaller sample sizes (Figure 2 of the study).

As highlighted by Simons *et al.* (2015),[37] missing data for unobserved variables may be an issue for imputation. Determining the pattern of missingness seen in the EQ-5D data from ATTRACTION-3[2] is therefore important in order to inform the selection of an appropriate imputation method. In Appendix N of the CS, it is explained that it was important to ascertain whether missingness was *"monotonic (missing constantly from one assessment until end of follow-up) or non-monotonic (sporadic)"* and whether missingness was *"temporally correlated with death"*.

The ERG agreed with the company that it is highly unlikely that utility data are missing completely at random (MCAR), and so this type of missingness was not considered further. The company considered the data to be missing at random (MAR) and therefore proceeded with imputation using Multiple Imputation by Chained Equations (MICE). However, the ERG considered there to be a lack of robust evidence to support the expectation that utility data are truly MAR. If the data are MAR, it is required that all unobserved values may be accurately predicted from a combination of other utility values and related patient characteristics/ outcomes (i.e. time to death and progression status).

The ERG noted that based on CS Appendix N Figure 1, it may be inferred that a relatively large number of observations are missing due to observations falling outside of follow-up. For example, at the maximum time period shown within the assessment period (144 weeks from baseline), there were [redacted] (CS Appendix N Table 3). This is because the remaining n=49 patients fell outside of the follow-up period, though it remained unclear to the ERG how the follow-up period for individual patients with respect to EQ-5D data collection was determined (and how this differed from follow-up time for time-to-event endpoints). Regardless of how follow-up was determined, this meant that a substantial quantity of utility data in the longer-term are nevertheless missing.

The ERG considered it likely that 'true' utility values at later time points are systematically lower than those seen in earlier time points (based on the general principle that utility declines over time, both related to disease progression and natural health decline as patients age). This is especially important within the context of an economic model which is capable of projecting survival outcomes over a 40-year lifetime horizon. In addition, as highlighted by Simons *et al.* (2015),[37] patients who are very sick are more likely not to complete the EQ-5D questionnaire, meaning that results will be skewed towards more healthy individuals leading to missing values that are missing not at random (MNAR). Consequently, it remained unclear if the imputation approach would lead to the accurate estimation of missing utility values, particularly those that were not collected based on follow-up in ATTRACTION-3.[2]

While the ERG agreed with the company that there were no established methods for addressing utility data that are MNAR, the ERG considered it likely that this type of missingness is applicable in the case of EQ-5D data from ATTRACTION-3,[2] at least in the longer-term where very little information is known on utility.

Based on the assumption of MAR, imputation was performed in the CS for OS and progression status, as well as the utility values themselves. OS times were imputed based on a covariate-adjusted FP log-logistic model for both treatment arms, as this was model was deemed to be *"consistently capable of fitting the data adequately and preserved the characteristics of the tail of the data sufficiently well for the purposes of this analysis where survival times of more than 18 months beyond maximum follow-up would not impact the utility imputation model."* (Appendix N, Section 4.4). The ERG noted for clarity that the OS model specified for imputation is different to the OS model specified for use in the company's economic model base-case analysis (an SP log-logistic model).

Here, it should also be noted that it was assumed that time-to-death would not be considered to have an impact on utility for patients still alive after 18 months. However, survival times were nevertheless required for the purpose of imputing missing observations for patients with administratively censored survival times. It was unclear to the ERG precisely how OS times were imputed beyond the impact on utility, and if it was appropriate to assume time-to-death would have no impact on utility after 18 months from treatment initiation.

Unobserved progression events and utility values were imputed by predictive mean matching. The CS states that this method of imputation determines which complete observations in the dataset are "closest" to the missing point in the space defined by the independent covariates,

and of the five closest observations, one is chosen at random, and the outcome observation of that record is copied to the missing observation (CS Appendix N, Section 4.5). The independent covariates used were identified via a pragmatic literature search, and are described in further detail within the CS (Section B 3.4.2.1.2 and Appendix N).

As highlighted previously, the appropriateness of these methods of imputation are contingent upon the data being truly MAR, which the ERG did not consider to be the case. For progression times specifically, it was unclear how the approach to imputation dealt with the fact that progression is not a pre-requisite event that should occur prior to death. It was therefore the ERG's expectation that utility values estimated using the imputation methods described by the company will lead to an over-estimate of mean utility for each health state across both treatment arms. However, the extent to which utility values are over-estimated remains unclear.

In summary, the ERG did not consider it possible to appropriately impute the missing utility data from ATTRACTION-3[2] with currently-available methods, which rely upon the accurate estimation of missing utility values, progression events, and survival times. Therefore, the utility values produced to inform the economic model may be considered at best, highly uncertain, and at worst, inaccurate and consequently misleading.

4.2.7.2. Alternative approaches

In the original CS, no alternative utility values were considered in sensitivity analyses, and so the ERG requested alternative values be provided at clarification stage. More specifically, the ERG requested two additional analyses of the ATTRACTION-3 trial data (a mixed-effects regression model and an analysis of non-imputed data), as well as the identification and subsequent application of any relevant utility values identified in the published literature or used in similar previous NICE appraisals.

For the mixed-effects regression (performed on the complete-case data set), the ERG requested a model be fitted including variables for treatment arm and disease progression, as well as a random intercept at the patient level. While limited information was provided in the clarification response concerning model fitting, significance of included variables etc., this model was provided by the company for comparison purposes in the economic model. However, the company did not consider this analysis to be appropriate for cost-effectiveness analysis, stating:

"… mixed models provide a mean per subject effect, and are not representative of the marginal value of utility in a health state over time, where subjects have varying time in state conditional upon their utility. Direct representation of the data collected provides a truer estimate of the mean utility in state over all time as those patients with worse utility spend less time in state." (clarification question B12 response).

From the SLR, the company highlighted that only one study[18] may be considered appropriate for inclusion in the economic model. However, this study did not report utility values by progression status, and so to include this within the model the same value (0.74, for patients receiving palliative chemotherapy) must be applied across both treatment arms and both health states. The ERG did not consider it appropriate to apply a single utility value across both treatment arms and both health states, and so this source was not considered further to inform the model *per se*, yet this value provided a means of validating the results from analysis of ATTRACTION-3[2] data.

In addition, the company provided a scenario analysis using utility values applied in the previous NICE TA378[38] assessment (ramucirumab for treating advanced gastric cancer or gastro–oesophageal junction adenocarcinoma previously treated with chemotherapy). The corresponding utility values of 0.737 (progression-free) and 0.587 (post-progression) were applied across both treatment arms.

4.2.7.3. Values used in the economic model

In the company's base-case analysis, separate utility values were derived for each treatment arm through taking the mean values of the imputed data set across each health state. A summary of the utility values used within the cost-effectiveness analysis is provided in Table 11.

Table 11: Summary of utility values for cost-effectiveness analysis

Health state	Nivolumab utility: mean (SE)	Control utility: mean (SE)
Pre-progression	[redacted]	[redacted]
Post-progression	[redacted]	[redacted]

Abbreviations: SE, standard error.

Source(s): CS Table 53.

The use of higher utility values for nivolumab than taxanes in the pre-progression state is justified within the CS based on a reduced number of sSAEs during nivolumab treatment versus taxanes during ATTRACTION-3.[2] In addition, it was noted within the CS that differences in utility between nivolumab and taxanes across both health states may be expected owing to nivolumab's novel mechanism of action (CS, Section B.3.4.3.1). However, the ERG noted that

the control arm mean baseline utility (taken at screening) was significantly lower than of the nivolumab arm ([redacted] versus [redacted], p = 0.034) (CS Appendix N, Section 4.1). It is therefore unclear how much of the estimated difference in utility between arms (across both health states) could be explained by differences at baseline, especially acknowledging that ATTRACTION-3[2] has an open-label study design (though it is unclear to the ERG whether patients were aware of their treatment allocation at screening visit).

The mean utility for the progression-free disease state for the nivolumab treatment arm is estimated to be [redacted], yet the mean baseline utility for this arm was [redacted] (CS Appendix N, Table 5). For the taxanes arm, the equivalent baseline utility was [redacted], whereas the mean utility for progression-free disease was [redacted]. The ERG questioned the face validity of the mean utility value for nivolumab-treated patients with progression-free disease being essentially identical to their baseline utility, both of which are significantly higher than the mean utility of the UK general population aged 65-70 years, based on Ara and Brazier (2011)[39] of 0.8041 (95% CI: 0.790, 0.817).

It was also notable that the median utility value for nivolumab treated patients was [redacted] based on CS Figure 12. The ERG noted that [redacted], [redacted], appeared unrealistic within the context of a patient population generally aged >65 years with an advanced cancer that has not responded to a previous line of chemotherapy.

After progression, patients randomised to receive nivolumab (some of whom may continue treatment beyond progression) are assigned a utility value of [redacted] versus [redacted] for taxane patients. This difference in utility for patients that have progressed appeared very large, and therefore the ERG considered these values to have questionable face validity. The ERG understood that some benefit accrued while patients are progression-free may be carried over into the progressed disease health state, yet the extent of this is unknown, and estimates of benefit carried through into the progressed state based on imputed data from ATTRACTION-3[2] are expected to be subject to a number of limitations (as discussed in Section 4.2.7.1).

Use of higher utility values for nivolumab versus a comparator in both pre-progression and post-progression health states aligns with previous nivolumab submissions in other indications, as highlighted within the CS for this appraisal. The ERG also acknowledged that the same approach had also been adopted in previous assessments of other checkpoint inhibitors, such as TA519[40] (pembrolizumab for treating locally advanced or metastatic urothelial carcinoma

after platinum-containing chemotherapy). However, the ERG highlighted the following excerpts from the FAD documentation for some of these appraisals:

- **TA519**[41]**:** Utility estimates should be pooled across treatment arms. *"The ERG highlighted that KEYNOTE-045 was open-label, which results in a risk of bias to the utilities because they are a patient-reported outcome… The committee considered that, given the uncertainties raised about treatment-specific utilities, the utilities should be pooled across treatment arms*."

- **TA490**[42]**:** The most appropriate utility values lie between the treatment-dependent and the treatment-independent estimates *"Although the committee preferred the ERG's conservative approach of using treatment-independent utilities, it acknowledged that this scenario was pessimistic and some potential quality-of-life benefits of nivolumab had not been captured."*

- **TA483**[43]**:** The committee concluded that the most appropriate values were likely to be between those presented by the company and those by the ERG (both independent of treatment arm). *"It acknowledged that the company's values of 0.750 and 0.592 (progression-free and progressed-disease health states respectively) were taken from EQ-5D data in the CheckMate 017 trial, but considered that they were likely to have been overestimated; on the other hand, the ERG's values (0.65 and 0.43) were lower, but there were limitations in how they were derived. The committee concluded that the most appropriate values were likely to be between those presented by the company and those by the ERG."*

From the above, it may be inferred that in past appraisals where similar issues have arisen, the general consensus appears to be that a difference in utility beyond progression may be expected, yet the magnitude of this benefit (and how long it should apply for) is unclear. Consequently, a more conservative approach may be considered to be a treatment independent approach to informing utility values, especially in consideration of an open-label trial (such as ATTRACTION-3[2]) which means patient-reported outcomes are subject to a risk of bias.

The corresponding utility values without imputation (i.e. a complete-case analysis of data from ATTRACTION-3[2]) are presented in Table 12. As may be expected, the complete-case analysis yielded larger utility values for each health state across both treatment arms (given that the majority of missing data were for relatively poorer health states). The ERG did not consider it

appropriate to use the complete-case analysis to inform the economic model (as is also stated within the CS), and so these values are not considered further.

Table 12: Comparison of imputed and complete-case utility values

Health state	Nivolumab utility: mean (SE)		Control utility: mean (SE)	
	Imputed	Complete-case	Imputed	Complete-case
Pre-progression	[redacted]	[redacted]	[redacted]	[redacted]
Post-progression	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: SE, standard error.

Source(s): CS Table 53 and clarification question B12 response (corresponding model file).

The results of the mixed-effects regression analysis are presented in Table 13 (versus the imputed values per the company's base-case analysis).

Table 13: Comparison of imputed and mixed-effects regression utility values

Health state	Nivolumab utility: mean (SE)		Control utility: mean (SE)	
	Imputed	Mixed-effects	Imputed	Mixed-effects
Pre-progression	[redacted]	[redacted]	[redacted]	[redacted]
Post-progression	[redacted]	[redacted]	[redacted]	[redacted]

Abbreviations: SE, standard error.

Note: The ERG has assumed the values for the complete-case analysis were missing 0's at the end owing to Excel rounding, and so for consistency with the other values has added a 0.

Source(s): CS Table 53 and clarification question B12 response (corresponding model file).

The ERG did not disagree with the company's criticism of a mixed-effects regression analysis, in that such an analysis (by definition) produces cohort-level values for the effect of treatment arm and progression status, and a mean individual-level effect which does not vary over time. However, such an analysis avoided the need to rely on data imputation, which (as described in Section 4.2.7.1) the ERG did not consider to have been conducted appropriately, nor did it consider it possible to impute these data appropriately with current-available methods.

As shown in Table 13, the utility values for each arm and health state are similar. However, both values for nivolumab decreased slightly, whereas both values for taxanes increased slightly. The ERG was unclear why the specification of a mixed-effects regression has led to these changes in the estimated values, as no explanation was provided by the company within its response. However, given that this approach does not rely upon imputation of missing data, the

ERG considered such an analysis to be more suitable for informing the economic model (versus the company's base-case analysis).

In the analysis considered based on data from NICE TA378,[38] utility values of 0.737 and 0.587 were applied for the progression-free and post-progression health states, respectively (for both treatment arms). These values are reported in the CS for TA378 (Table 73). Elsewhere in the CS for this appraisal, utilities estimated from the pivotal CORRECT trial are referenced from a published gastric cancer study by Grothey *et al.* (2013).[44] It was the ERG's understanding that both TA378 and the study by Grothey *et al.* (2013)[44] are based on the same study and patient population (CORRECT). In relation to this study specifically, the ERG highlighted the following:

- Mean baseline utilities in CORRECT were 0.73 (regorafenib) and 0.74 (placebo) group, which are notably better balanced versus the ATTRACTION-3[2] study. Mean end-of-treatment utility in CORRECT was 0.59 for both treatment arms
- The ERG noted that in TA378,[38] baseline and end-of-treatment values were assumed proxies for the pre- and post-progression health states (based on a mean value taken from the complete-case analysis from CORRECT). Consequently, the ERG noted that comparison of these values was inherently flawed, as the more appropriate comparison to make would be between values at baseline, which would be [redacted] versus 0.73 or 0.74

[redacted]

Based on the information presented above, the ERG also raised concerns regarding the use of utility values frrom TA378[38] for this appraisal, yet external values permitted the ERG to understand how influential alternative estimates of utility are on the cost-effectiveness results for this appraisal.

4.2.8. Resources and costs

The company's model included costs relating to nivolumab and taxane treatments, second- and third-line BSC, medical resource use, and the resolution of AEs; discussed in turn below.

4.2.8.1. Intervention and comparators' costs

Nivolumab has a list price of £2,633.00 per 240 mg vial. The equivalent cost including the commercially-sensitive, [redacted] PAS discount is [redacted]. Each treatment cycle is two weeks, and treatment with nivolumab is administered intravenously over 30 minutes on Day 1 of each cycle at a fixed dose of 240 mg. Treatment will be continued until disease progression, though a ToT curve was used to inform treatment discontinuation within the model (Section 4.2.6.4 for more information).

While each dose is fixed at 240 mg, the CS noted that not all planned doses were administered. The model accounted for this adjustment based on the proportion of actual versus expected doses received, which for nivolumab was equivalent to 95.2% of doses (CS Table 58).

Two taxanes were included as comparators within the model: docetaxel and paclitaxel.

- Docetaxel is administered at 75 mg/m^2 once every three weeks, for each three-week treatment cycle, via intravenous infusion over at least 60 minutes
- Paclitaxel is administered at 100 mg/m^2 once every week for six weeks, followed by one week of rest, for each seven-week treatment cycle, also via intravenous infusion over at least 60 minutes
    - At clarification stage, the ERG noted that some materials related to the ATTRACTION-3[2] study comment on a two-week rest period (clarification question B13). The company confirmed that this relates to the time period from last dose of treatment (on Week 6, Day 1 of Cycle x) to the next dose (on Week 1, Day 1 of Cycle x+1)

The CS stated that *"the lowest possible acquisition costs were applied"* for the taxanes arm, with costs equivalent to £720.10 per 134.25 mg cycle dose and £367.37 per 179 mg cycle dose for docetaxel and paclitaxel, respectively. These values align with the Monthly Index of Medical Specialities (MIMS) price listings, cited in the CS (Tables 60 and 61).

However, an alternative to costs from MIMS would be to use published costs via the British National Formulary (BNF). The BNF[45,46] includes costs for each product from a number of different manufacturers that provide branded equivalents for both treatments at varying prices. The ERG was able to identify equivalent costs via BNF as used in the CS (taken from MIMS).

As both taxanes are generic medicines used in a range of therapeutic areas, prices are also published via the electronic Market Information Tool (eMIT)[47] which captures volume-based discounts provided to the NHS. Published eMIT prices demonstrate equivalent mean costs of £20.96 per 134.25 mg cycle dose and £39.32 per 179 mg cycle dose for docetaxel and paclitaxel, respectively. These represent the prices paid for a total quantity of 13,825 units of docetaxel and 20,976 units of paclitaxel within the NHS.

It was unclear to the ERG why costs from MIMS were used in preference to costs from either eMIT or the BNF (if eMIT unavailable), both of which the ERG would consider standard cost sources to inform submissions to NICE. The use of MIMS costs has led to a substantial over-estimate of costs for the taxanes, and therefore the ERG's base-case considered the use of eMIT costs as a correction to the company's base-case analysis.

In the company's base-case analysis, the costs for the taxanes arm were based on an assumed 50:50 split (i.e. *"a simple average"*) of the costs applied for docetaxel and paclitaxel (CS Section B3.5.1.3). In ATTRACTION-3,[2] there were n=65 docetaxel and n=144 paclitaxel patients that comprised the full n=209 taxanes arm; hence the trial-based ratio of taxane use would be approximately 31:69. Clinical advice provided to the ERG suggested that docetaxel is usually the first choice of taxane for most OSCC patients. The relevance of a 50:50 ratio to UK practice is unclear, and so the ERG considered two alternative scenarios: (1) assuming the split per ATTRACTION-3,[2] and (2) assuming 100% use of docetaxel (purely for costing purposes).

BSC is included within the CS as an adjunct to second-line nivolumab or taxane therapy, as well as a "third-line treatment" after progression (as well as a comparator against nivolumab in a scenario analysis). However, for the purpose of this report, BSC costs are considered disease management costs, and are therefore discussed further in Section 4.2.8.3.

At clarification stage, the ERG asked the company clarify what proportion of patients received post-progression treatment in ATTRACTION-3[2] by treatment arm (clarification question A9), acknowledging that BSC is expected to comprise the mainstay of current NHS practice in this setting. In response, the company stated that 82 of the 210 nivolumab patients ([redacted]) received treatment post-progression, with a median of [redacted] treatments (range: [redacted]) and a median post-discontinuation time on treatment of [redacted]. For the taxanes arm, three of the [redacted] received treatment post-progression; all patients had one subsequent treatment and a median post-discontinuation time on treatment of one day.

However, in the pivotal trial publication of ATTRACTION-3,[2] it is stated that 119 (57%) of 210 patients in the nivolumab group and 115 (55%) of 209 patients in the taxanes group received subsequent therapy for advanced oesophageal cancer (though is not described in relation to progression status). Furthermore, it is noted that the most common subsequent treatments were taxanes (for 100 [48%] of the 210 patients in the nivolumab group and 43 [21%] of 209 patients in the chemotherapy group), fluoropyrimidine-based chemotherapies (24 [11%] of 210 and 39 [19%] of 209), and platinum-based chemotherapies (20 [10%] of 210 and 22 [11%] of 209).

The ERG could not establish why these figures differ to such an extent, and is especially concerned that nearly half of the nivolumab group received subsequent taxane therapy (48%), based on reporting in the ATTRACTION-3[2] pivotal trial publication.

4.2.8.2. Treatment administration

Both nivolumab and taxanes were administered via intravenous infusion. To apply the cost of treatment administration within the model, the company calculated a weighted average of the National Cost Collection cost for "Deliver Simple Parenteral Chemotherapy at First Attendance", based on the SB12Z code (based on a combination of day case, regular day/night, outpatient, or "other" settings).[48] This yielded an estimated cost of £241.06 per administration (CS, Document B, Table 57).

The ERG acknowledged that each administration of nivolumab is expected to be associated with notably less chair time versus the taxanes (i.e. 30 minutes versus at least one hour).[†] As such, on average, the ERG expects the cost per administration of nivolumab would be lower versus taxanes (based on reduced chair time). In addition, clinical advice provided to the ERG was that treatment administration is most likely to take place in a day case setting.

Consequently, the ERG's preferred assumptions for administration costs included a higher cost for the taxanes arm (reflecting increased chair time) and costing based on a day case setting. A comparison of the company- and ERG-preferred administration costs is provided in Table 14. The ERG's preferred administration costs are factored into the ERG's preferred base-case analysis, which is discussed further in Section 6.3.

[†] The ERG notes that in the CS (Table 56) it is stated that nivolumab is expected to administered by intravenous infusion over 60 minutes every 2 weeks. However, based on reporting in the pivotal ATTRACTION-3 publication, the ERG understands that administration would take place over 30 minutes.

Table 14: Comparison of company- and ERG-preferred administration costs

Treatment	Company	ERG
Nivolumab	£241.06 (SB12Z, total HRG's)	£254.14 (SB12Z, DCRDN)
Taxanes	£241.06 (SB12Z, total HRG's)	£385.28 (SB14Z, DCRDN)

Abbreviations: DCRDN, Day case and Regular Day/Night; HRG, Healthcare Resource Group.

Source(s): National Cost Collection.[48]

4.2.8.3. Medical resource use

The costs associated with BSC, routine monitoring and follow-up (henceforth termed 'medical resource use [MRU]') were applied to both the nivolumab and taxane arms in the company's model. Estimates of resource use were determined from clinician survey and unit costs were taken from NHS reference costs. The ERG considered that in general, the methods used for obtaining MRU and unit cost assumptions to inform the company's model are appropriate. However, some discrepancies occur in the costs used within the CS model compared with costs calculated by the ERG. A brief summary of the key components of MRU are described below along with commentary from the ERG.

The composition of BSC was derived from the clinician survey involving practicing oncologists and nurses in the UK, based on their experience in treating UK-based gastric and GOJ cancer patients. Costs for BSC components were derived from the National Cost Collection for the NHS 2018/19, with the proportion of patients requiring each aspect of BSC treatment determined from the UK clinician survey.

The costs of BSC in the CS varied between settings, due to the proportion of patients requiring each aspect of care differing when used as a second-line adjunct to nivolumab or taxanes, compared with BSC treatment as a second- or third-line therapy in isolation. In response to clarification question B14, the company provided a comparison of resource use items captured by the model (based on the clinician survey) versus information from NICE TA378[38] (which served as a validation source). Compared to the population under consideration for TA378 (advanced gastric cancer or gastro–oesophageal junction adenocarcinoma), OSCC patients are expected to also potentially undergo procedures and drugs to control gastrointestinal bleeds, and may also require oesophageal stents and/or ascites drainage.

Some discrepancies occurred in the costs comprising BSC presented in the CS (Document B, Table 64). Radiotherapy is costed as £184.25; however, the ERG calculated this as £92.13

based on the information provided in CS Table 64 (£487.45 [cost of treatment] x 0.189 [proportion of patients requiring treatment]). Correcting these values within the company's base-case analysis increased the ICER to £45,502 (compared with £45,491 in the CS). A similar error is noted when BSC is used as a separate comparator in the scenario analyses (CS Table 64). In this case, radiotherapy is costed as £207.94, whereas the ERG calculated this as £103.80 (£487.45 [cost of treatment] x 0.213 [proportion of patients requiring treatment]).

Additionally, nerve blocks pain relief was costed as £26.62 for the BSC scenario analysis, whereas the ERG calculated this as £2.66 (£532.96 [cost of treatment] x 0.005 [proportion of patients requiring treatment]) based on the information in CS Table 64. The ERG suspected this may be due to a typographical error concerning the number of zeros in the proportion of patients requiring nerve blocks (e.g. 0.005 should perhaps be 0.05), but this is purely speculation.

Based on the limitations of the BSC comparator outlined in Section 4.2.4, further adjustments to this comparator were not made. The third-line BSC treatment costs for the BSC + nivolumab/taxane arms are provided as a single value in the CS model and it is not clear to the ERG how these values were calculated from the data provided in CS Table 68.

The breakdown of costs for pain relief components is provided in CS Table 66. Similar to the taxane costs detailed in Section 4.2.8.1, the costs of pain relief medications are derived from MIMS rather than eMIT or the BNF, both of which the ERG would consider standard cost sources to inform submissions to NICE. However, the low overall costs of pain relief medication at £2.17 for week 1 and ongoing daily costs of £0.31 mean that these costs are unlikely to have a large impact on the economic model results. The source of costs for nerve blocks pain relief is not specified in Table CS 66, despite the reader being directed to this table to obtain further details of the medication(s) used and cost breakdown. Consequently, the ERG was not able to check that the costing used for nerve blocks pain relief was correct, despite this making a moderate contribution to the BSC costs at ~£26/week.

Medication costs to control GI bleeds were sourced from Campbell *et al.* (2015)[49] and provided in CS Table 65 as a component of BSC. The costs from Campbell *et al.* (2015)[49] were derived from data collected in 2012-13, which were stated as £23.76. The ERG assumed that an inflation factor was applied to obtain the cost of £25.71 presented in the CS (Document B, Table 65) but this is not clearly stated within the CS. The cost for ascites drainage is stated as £3,404.20 in the CS (Document B, Table 65), based on a value of £3,146 obtained from White & Carolan-Rees (2012)[50] inflated to 2015/2016 costs. After reviewing the cited manuscript,[50] it was

not clear to the ERG how the value of £3,146 was obtained for ascites drainage. The CS did not provide any breakdown of how this cost was calculated, nor the method used for applying an inflation factor to this cost.

The MRU items captured within the model include clinical consultations, blood tests and investigations, hospitalisation and palliative care. The ERG was satisfied that the included costs cover the key MRU items expected to be required by patients.

The cost of clinical consultation was cited as a weighted average of consultant-led and non-consultant led consultations from the National Cost Collection for the NHS 2018/19. The ERG calculated this as £196.33, in contrast to the value of £187.36 provided in CS Table 70. The ERG suspected the value of £187.36 was provided based on outpatient code 370 (Medical Oncology), and not the average across all HRG codes as stated in the CS. However, it remained unclear to the ERG which cost the company considered most appropriate to inform the model.

The costs for hospitalisation included in the CS (Document B, Table 70) are based on a weighted average of elective and non-elective long-stay hospitalisation (Malignant Gastrointestinal Tract Disorders, weighted average of elective and non-elective long-stay FD11A-FD11K). This was costed as £534.07 per hospitalisation; however, the unit costs of hospitalisation included in the weighted average ranged from £1,907 (FD11K, elective) to £9,650 (FD11A, elective). It was therefore unclear to the ERG how a weighted average of £534.07 was estimated. The ERG was able to estimate a different value of £577.11 using the same codes in a non-elective short stay setting, but was unable to calculate a value of £534.07. The ERG calculated the weighted average for hospitalisation costs as £3,379.73, based on the description provided in the CS.

The difference in costs is displayed in Table 15 when amending the cost for clinician consultations and hospitalisations, as described above.

Table 15. Cyclic (weekly) health state resource use and costs based on CS Table 70 versus ERG calculations

Resource	CS			ERG calculations		
	Unit cost (£)	Weekly cost		Unit cost (£)	Weekly cost	
		Use (%)	Cost (£)		Use (%)	Cost (£)
Clinician consultation*	187.36	0.153	28.67	196.33	0.153	30.04
CT scan	97.15	0.092	8.94	97.15	0.092	8.94
Full blood count	2.79	0.221	0.62	2.79	0.221	0.62
Renal function test	1.10	0.162	0.18	1.10	0.162	0.18
Hepatic function test	1.10	0.170	0.19	1.10	0.170	0.19
Hospitalisation†	534.07	0.095	50.74	3,379.73	0.095	321.07
Palliative care specialist nurse	76.74	0.359	27.55	76.74	0.359	27.55

Abbreviations: CS, company submission; CT, computed tomography.

Notes: Costs of particular interested shaded in grey *ERG unit cost estimated based on outpatient code 370 (medical oncology); †ERG unit cost estimated using description provided in CS (weighted average of elective and non-elective long stay codes FD11A-K).

Source(s): National Cost Collection.[48]

In the ERG's preferred base-case analysis, the re-calculated costs were applied (as shown in the ERG's calculations in Table 15). This was based on the ERG's understanding that the unit costs were miscalculated, and that the references stated were correct. Alternative costs are explored in Section 6.1.

One-time, end-of-life costs prior to death are included in the model, with the costs estimated from prior literature[51] and inflated to reflect estimated current costs. Details of the breakdown of costs is not provided within the CS. A cost of £7,987.00 was derived from the identified literature before being inflated to reflect current costs, which is presumably obtained from Table 9 in the cited study.[51] The ERG summed the items included within Table 9 of the cited literature but obtained a smaller value of £7,287.00. The method used for applying an inflation factor to this cost was not clearly detailed within the CS.

The ERG noted that the one-time, end-of-life costs represent estimated costs within the last three months of life and that several components are included within the MRU items. This included: (some) hospitalisation costs, social care costs (i.e., respite care) and nursing costs. Furthermore, a substantial proportion of patients are estimated to die within the first three months after treatment initiation, and so may not accrue all of the terminal care costs within the time horizon of the model.

4.2.8.4. Resolution of adverse events

The unit costs associated with the resolution of AEs were taken from recent literature[52] and inflated to 2019 costs, with the exception of leukopenia and lymphocyte count decreased, which were assumed to be equivalent to neutropenia.

The ERG noted that AE costs appeared to be higher than costs used to inform past appraisals of late-stage cancer treatments. The Copley-Merriman *et al.* (2018)[52] study provided AE resolution costs for patients with metastatic melanoma, whereas substantially lower costs for AEs have been previously identified for treatment of advanced non-small-cell lung cancer in the United Kingdom,[53,54] which may represent a better proxy for advanced OSCC.

The ERG identified three previous appraisals that reported AE resolution costs to understand the range of values used to inform submissions made to NICE, presented in Table 16. Firstly, values from TA378[38] were identified (based on similarity between the populations, as identified in the CS). Secondly, TA525[32] was searched as an example of a recent appraisal of a checkpoint inhibitor versus taxanes (though in a different patient population). Lastly, TA628[55] was identified as the most recently-published non-OSCC cancer appraisal at the time of writing.

Table 16: Costs applied in the model for resolution of AEs

AE	ID1249 CS	TA378*	TA525†	TA628‡
Anaemia	£1,592.39	£1,211	£329.92	£631.88
Febrile neutropenia	£4,755.76	£3,019	£362.66	£495.48
Leukopenia	£1,308.26	*£364*	£362.22	*£495.48*
Lymphocyte count decreased	£1,308.26	*£364*	*£362.22*	*£495.48*
Neutropenia	£1,308.26	£364	*£362.22*	£495.48
Alanine aminotransferase increased	£268.61	NR	£163.00	NR
Aspartate aminotransferase increased	£268.61	NR	£163.00	NR
Diarrhoea	£2,426.57	NR	£114.00	NR
Rash	£1,039.65	NR	NR	NR

Abbreviations: DCRDN, Daycase and Regular Day/Night; HRG, Healthcare Resource Group; NR, not reported.

Note(s): *Cost year: 2012-13, see Table 115 of TA378 CS. †Cost year: 2015-16, see Table 70 of TA525 CS; ‡Cost year: 2017-18, see Table 58 of TA628 CS. Values shown in italics are based on assumption applied in company base-case (that is, leukopenia, neutropenia, and lymphocyte count decreased can be considered equivalent).

Source(s): National Cost Collection.[48]

Based on Table 16, it can be seen that the cost of resolving AEs is consistently greater for each AE compared to the other previously conducted appraisals (where reported). The ERG highlighted that only three example appraisals have been highlighted, and the approach to identification should not be considered systematic. It was not clear how the data from Copley-Merriman *et al.* were used to inform the costs of resolving AEs; however, this seemed to be derived from data previously reported by Wehler *et al.* (2017),[56] which was initially published online in 2015 (and cited by Copley-Merriman *et al.* (2018)[52]).

The unit costs reported in Table 16 have been *"inflated to 2019 costs"* using *"PSSRU inflation factors"* (CS Section B.3.5.3). The uninflated costs used as a basis were not reported in the CS, nor were the precise methods to inflate the costs. The ERG therefore could not validate the inflated costs used to inform the model.

Due to the occurrence of more AEs for patients treated with taxanes versus nivolumab, the total costs associated with AE resolution are substantially higher for taxanes ([redacted] versus [redacted] per average patient for the base-case analysis disaggregated results, Appendix J). To further understand the impact of AEs (in terms of costs) on model results, the ERG performed an exploratory analysis to disable AE costs from the model (which was made possible through a switch included in the company's model). Results of this analysis are provided in Section 6.2.

In spite of the limitations highlighted above, based on clinical advice provided to the ERG, it was recognised that the resolution of AEs was expected to be associated with substantial costs. Therefore, the ERG's base-case analysis did not adjust the costs included in the company's base-case analysis. However, the ERG still emphasised caution when considering the potential impact of AEs on the economic model results.

5. COST-EFFECTIVENESS RESULTS

5.1. Company's cost-effectiveness results

5.1.1. Base case results

Results of the company's base-case analysis are presented as the ICER for nivolumab versus taxane (pooled comparator), per ATTRACTION-3. The total costs, QALYs and LYs are presented in CS Table 75, replicated in Table 17 below. Of note, the company's base-case analysis incorporated a PAS discount of [redacted] applied to the list price of nivolumab.

Table 17: Company base case results

Arm	Total			Incremental			ICER (£/QALY)
	Costs (£)	LYs	QALYs	Costs (£)	LYs	QALYs	
Company base-case (deterministic)							
Taxane	[redacted]	[redacted]	[redacted]				
Nivolumab	[redacted]	[redacted]	[redacted]	20,842	0.536	0.458	45,491

Abbreviations: ICER, incremental cost-effectiveness ratio; LY, life year; QALY, quality adjusted life year.

Source(s): CS Table 75

The company's base-case ICER was £45,491 for nivolumab versus taxane, based on incremental costs of £20,842 and a QALY gain of 0.458. Of the total 0.458 incremental QALYs, 0.094 were gained in the 'pre-progression' health state, leaving 0.364 which were gained in the 'post-progression' health state. This finding illustrates that the majority of QALYs gained by patients treated with nivolumab were accrued within the 'post-progression' health state, yet some incremental benefit associated with nivolumab is also accrued in the 'pre-progression' health state.

5.2. Company's sensitivity analyses

The CS reported a number of sensitivity analyses to explore the impact of alternative settings and assumptions, as well as the role of parameter uncertainty within the model results. These analyses are discussed in turn below.

5.2.1. Deterministic sensitivity analysis

The company conducted a deterministic sensitivity analysis (DSA) by modifying the parameters presented in CS Section B.3.8.2. The CS stated that upper and lower bounds for the majority of

parameters were varied by ± 20% of the (mean) base-case value. Exceptions to this were discounting parameters (upper and lower bounds of 6% and 0%, respectively), sex (upper and lower bounds of 100% and 0% male, respectively) and time horizon (adjusted to 10 years).

A tornado plot was used to present the DSA results in CS Figure 34, with the outcome of interest being the ICER. The plot showed that the results were most sensitive to second line treatment costs and health state utility post-progression. Discounting of costs and benefits, baseline characteristics, and the proportion of patients receiving a dose also had substantial impacts of the ICER.

The ERG noted that varying each parameter by its reported measure of uncertainty (where available) would have been a more appropriate and superior method for understanding the impact that changes to individual parameters could have had on the model results. More specifically, varying parameters by the bounds of the confidence interval allows inferences to be drawn in relation to the quantifiable uncertainty of each parameter, which cannot be achieved using a fixed percentage of the (mean) base-case value. Table 73 of the CS suggests that measures of uncertainty were available for several parameters which could have been used within the DSA.

Understanding the effects of observed variability for each parameter was also precluded by the presentation of ‘treatment costs’, ‘health state costs’ and ‘AE costs’ as grouped variables, rather than presenting the individual components of each of these groups separately in the form of a one-way sensitivity analysis. The ERG noted that such an approach may mask the impact of component model inputs that may act in opposite directions or apply only for one treatment arm. Conversely, by grouping all parameters together and assuming a large SE, the uncertainty present in a given category may be substantially overestimated.

Owing to the methodological limitations of the DSA presented in the CS, and acknowledging that it is beyond the remit of the ERG to reconstruct a true one-way sensitivity analysis from first principles (given the large quantity of custom VBA code present in the company’s model), the ERG has instead opted to focus predominantly on the range of alternative scenario analyses in order to determine the key drivers of cost-effectiveness results.

5.2.2. Probabilistic sensitivity analysis

The company conducted a probabilistic sensitivity analysis (PSA) to explore the impact of parameter uncertainty, based on each model parameter’s respective distribution (listed in CS

Table 73). PSA results are provided in CS Table 76, replicated in Table 18. The cost-effectiveness plane and cost-effectiveness acceptability curve (CEAC) are also provided in CS (Document B, Figures 32 and 33), respectively.

Table 18: Company base case results (probabilistic)

Arm	Total			Incremental			ICER (£/QALY)
	Costs (£)	LYs	QALYs	Costs (£)	LYs	QALYs	
Taxane	[redacted]	[redacted]	[redacted]				
Nivolumab	[redacted]	[redacted]	[redacted]	21,210	0.547	0.468	45,278

Abbreviations: ICER, incremental cost-effectiveness ratio; LY, life year; QALY, quality adjusted life year.

The PSA results from the CS are similar to the deterministic base-case results. The company stated that at a willingness to pay threshold of £50,000 per QALY gained, the probability of nivolumab treatment being cost-effective versus taxane was [redacted]. The ERG replicated the PSA using the company base case using two different random seeds and achieved results within 1% of those reported.

The ERG noted that the run time for the PSA was unusually long, requiring nearly two hours for each replication of the PSA. The CS stated that it was not possible to ascertain standard errors for all parameters in the PSA, and where necessary the PSA was informed by assuming that the standard error was 20% of the mean value. However, it was not explicitly stated within the CS or the corresponding economic model which parameters had standard errors ascertained and which parameters had standard errors estimated. This limits the inference that can be drawn from the analysis.

The ERG highlighted the following errors and caveats within the company's PSA:

- Several parameters that apply to both treatment arms awere sampled independently within the PSA. These are health state costs related to disease progression status (£116.87 per model cycle, for both health states across both treatment arms) and terminal care costs applied upon death (£8,973.60).
- Some parameters were assumed to be fixed (i.e. not varied) within the PSA, which should be associated with parameter uncertainty. These included the proportion of doses received per cycle, administration costs and second-line BSC costs. Each of these parameters

should be reflected within the PSA, and it was unclear to the ERG why uncertainty around these parameters specifically was not applied within the model.

- Costs associated with parameter uncertainty have been varied using a gamma distribution. Briggs (2005)[57] highlighted that a gamma distribution can be assigned to reflect the uncertainty associated with *single* costs. However, within the context of a cohort-level model, the normal distribution may be a more appropriate reflection of the uncertainty in a given cost, owing to the role of the Central Limit Theorem. While, as highlighted above, it was unclear which parameters have been assumed to have an SE of 20% of the mean, it appeared as though many of the costs have been informed in this manner. Through specifying a large SE, this has the potential to lead to an unrealistic skew in the costs produced to inform the PSA.

- The approach taken to capturing the uncertainty in the SP models appears incorrect, as the curves are simply varied using a normal distribution around the point estimate of survival at each point in time (as opposed to sampling the uncertainty in the curve fit parameters themselves). As the values were pasted into the model directly, the ERG was unable to establish how these bounds were derived; however, it can be seen from the approach taken that there is zero uncertainty assumed in the *shape* of the curve, rather, all uncertainty is assumed to apply to the overall *scale* of the curve.

- Uncertainty around the individual utility values is very small. The ERG expected that this was primarily due to the imputation of missing data which artificially increases the certainty around point estimates of specific utility values (potentially through inflation of the effective sample size). Therefore, the true uncertainty in the estimation of utility values is not considered to be appropriately reflected by the PSA

As with the DSA, the ERG opted to focus predominantly on the range of alternative scenario analyses in order to determine the key drivers of cost-effectiveness results, as opposed to the results of the PSA.

5.2.3. Scenario analyses

The company conducted a number of scenario analyses to assess the impact of structural uncertainties and alternative settings and assumptions on the base-case results. Results are provided in Table 77 and 78 of the CS. Following the requests at the clarification stage, additional scenarios were explored in relation to clarification questions B6 and B10 to B12.

5.2.3.1. Survival extrapolations

The CS identified that the application of alternative clinically-plausible extrapolations (determined by the company) of OS for both treatment arms had a large impact on survival benefit, and consequently the ICER. All alternative clinically-plausible extrapolations of OS for the nivolumab arm increased the company's base-case ICER, with five out of six alternative extrapolations presented in CS Figure 35 increasing the ICER beyond £50,000.

The ICER was largely affected by different extrapolations as the company's base-case analysis (SP log-logistic, 2.99 months) yields a high estimated mean survival (i.e. modelled life-years, equivalent to 24.33 months), compared with the observed restricted mean survival ([redacted] months) and alternative SP extrapolations (15.07 to 20.00 months, CS Appendix M, Figure 38). In other words, approximately 42% of the life-years predicted in the company's base-case analysis are based on the extrapolated portion of the curve.

The application of alternative, clinically-plausible extrapolations of OS for the combined taxanes arm had a smaller effect on the ICER. In CS Figure 36, six out of seven alternative extrapolations increased the ICER of nivolumab versus taxanes, with three of these extrapolations increasing the ICER above a £50,000 per QALY threshold. Alternative extrapolations for PFS for the nivolumab and taxane arms had minimal impact on the ICER of nivolumab versus taxanes.

In consultation with the ERG, two practising oncologists suggested that the majority of patients undergoing second line taxane therapy would have an estimated survival of less than 12 months. However, as seen in ATTRACTION-3,[2] 12-month OS for the taxanes arm was approximately 34%, indicating that patients in the study were likely fitter than those that would be eligible for taxanes in NHS practice. This means that inferences concerning the most plausible extrapolations for each treatment arm should be considered with this potential discrepancy in patient population in mind.

It is important to consider that log-logistic extrapolations (as used for nivolumab OS) and exponential extrapolations (as used for taxanes OS) exhibit unique properties, with log-logistic curves producing a prolonged plateau beyond the observed data and exponential curves producing a sharper decrease beyond observed data (through the specification of a constant hazard of death). It should be noted that both curve fits were deemed clinically plausible in the CS for either treatment arm (as denoted by marking in CS Tables 77 and 78), while the scenario

analysis demonstrated that using the same curve for both treatment arms (either log-logistic or exponential) increased the ICER to £58,782 and £65,796, respectively.

Scenario analyses were not performed whereby alternative clinically plausible extrapolations of OS were applied for both nivolumab and taxane arms simultaneously. In consideration that 12 out of 13 alternative clinically plausible extrapolations for OS in the nivolumab or taxane arm increased the ICER independently of each other, adjusting these simultaneously would cause further increases in the ICER of nivolumab versus taxanes.

In response to clarification question B10, the company provided spline-based models as an additional sensitivity analysis for extrapolation of the survival curves. Spline-based models for OS and PFS in the nivolumab arm and PFS in the taxanes arm were deemed 'clinically implausible' based on the fitted data exceeding the 95% confidence intervals of the observed data. In spite of this, the spline-based extrapolation of OS in the taxanes arm with one knot at 2.99 months produced an ICER of £46,436 (i.e. very similar to the company's base-case ICER).

In further response to clarification question B10, SP models were provided using alternative cut points of 4.37 and 5.75 months. The application of clinically-plausible extrapolations of OS for the nivolumab arm using cut points of 4.37 and 5.75 months produced estimated ICER values ranging from £41,488 to £63,418. The application of clinically-plausible extrapolations of OS for the taxanes arm using cut points of 4.37 and 5.75 months produced estimated ICER values ranging from £45,408 to £47,456. Alternative extrapolations for PFS for the nivolumab and taxane arms had minimal impact on the ICER.

5.2.3.2. Choice of comparator

Scenario analyses were performed to assess the cost-effectiveness of nivolumab versus each individual taxane (docetaxel and paclitaxel) and BSC, as presented in CS (Document B, Table 79). The predicted discounted incremental QALYs ranged from 0.401 (versus docetaxel) to 0.414 (versus paclitaxel) to 0.630 (versus BSC), with variation in discounted incremental costs from £20,971 to £19,371 to £30,434, versus paclitaxel, docetaxel and BSC, respectively. The resultant ICERs for nivolumab versus docetaxel, paclitaxel, and BSC were £52,340, £46,764 and £48,298, respectively.

In response to clarification question B6, sensitivity analysis was included whereby the individual taxanes were included as a covariate in survival models, rather than separating the comparator arm into two subgroups. This provided comparable results to the subgroup analysis in the CS,

with an ICER estimate of £50,176 per QALY for nivolumab versus docetaxel and £47,037 per QALY for nivolumab versus paclitaxel.

The ERG again emphasized the limitations of the indirect comparison between nivolumab and BSC (Section 3.4), and so did not consider this result to be suitable for decision making. The ERG also noted that the nivolumab group used to inform these analyses remains unchanged, even though the clinician's preferred choice of taxane would have been made for the nivolumab patients prior to randomisation. Therefore, the ERG did not consider these scenarios suitable to inform decision making.

5.2.3.3. Treatment discontinuation

Scenario analyses were performed to assess the effects of alternative treatment stopping rules, as presented in CS Table 80 and 81. The implementation of a two-year stopping rule for nivolumab resulted in an ICER of £40,909, which signals a reduction in the estimate from the base case (£45,491). Implementing a scenario with treatment cessation upon progression in patients receiving nivolumab and a transition to third-line BSC resulted in an ICER of £45,455, which is comparable to the base case ICER. Importantly however, neither or these scenarios were associated with any changes to the modelled QALYs or LYs (i.e. no difference in clinical outcomes), and so should therefore be interpreted with caution.

5.2.3.4. Utility values

In response to clarification questions B11, the company provided additional scenario analyses whereby alternative utility values were incorporated into the economic model. The incorporation of utility values of 0.74 for nivolumab and taxanes in both the pre-progression and post-progression states (based on findings from the systematic literature review) resulted in an estimated ICER of £52,500. Alternatively, the incorporation of values of 0.737 for the pre-progression health state and 0.587 for the post-progression health state for both treatment arms (based on TA378[38]) resulted in an ICER of £63,982.

In response to clarification question B12, the company provided sensitivity analyses for the utility values derived from ATTRACTION-3[2] and used in the company base case. The use of a mixed-effects regression model for the estimation of utility values increased the ICER to £47,982 (compared with the base case of £45,491). Alternatively, the estimation of utility values from ATTRACTION-3[2] without the use of any data imputation (i.e. a complete-case analysis) decreased the ICER to £44,672.

5.3. Model validation and face validity check

The ERG performed a range of validation checks on the economic model. The ERG was able to replicate the deterministic base-case results, DSA, and PSA using the model originally submitted by the company. However, the submitted model relied heavily on non-standard custom Visual Basic for Applications (VBA) functions, which the ERG assumed were applied for the purpose of improving the user interface of the model. This approach had the undesired effect of greatly reducing the transparency of the calculations, particularly noting the limited timeframe over which the ERG performed its critique.

Subsequently, the model was replicated by the ERG in a separate file using standard Excel formulae and pasted values where necessary (e.g. survival curves), which replicated the results of the base-case analysis. However, it was beyond the scope of the ERG's review to re-construct the full model, including all relevant sensitivity analyses.

The originally-submitted model did not allow for replication of any of the scenario analyses included within the CS. At clarification stage, the company provided additional information to replicate these scenario analyses. However, the provision of 16 separate Excel files led to difficulty in the ERG making comparisons when adjusting different parameters simultaneously. For example, alternative survival curve extrapolations (e.g. FP or SP for taxanes or nivolumab) had to be selected in separate documents, with manual calculations performed to generate ICER values, and could not be combined with other scenarios (e.g. different utility values). Nevertheless, the ERG was able to replicate the reported outcomes from the CS and responses to clarification questions using the models provided by the company.

An important issue present in the company's economic model was that the SP curves provided in all model documents were hard-coded values, without the presentation of any formulae from which to validate the construction of the curves. The ERG digitised the Kaplan-Meier curves (using WebPlotDigitizer v4.2[58]) in order to ensure the SP models were aligned with the observed data from ATTRACTION-3.[2] The ERG was satisfied that the SP models were aligned with the Kaplan-Meier curve, but was unable to validate the formulae used to produce the parametric components of each SP model.

The ERG noted NICE DSU TSD 14 guidance: *"Where parametric models are fitted separately to individual treatment arms it is sensible to use the same 'type' of model, that is if a Weibull model is fitted to one treatment arm a Weibull should also be fitted to the other treatment arm"*.[59] While

this principle is subject to debate within the context of two treatments with very different mechanistic properties, the ERG calculated ICER values for nivolumab versus taxanes using the same method of OS extrapolation (where deemed clinically plausible for both treatment arms in the CS, based on CS Figures 35 and 36). When matching the methods of extrapolation, the ICER for nivolumab versus taxanes ranged from £58,148 (SP log-normal) to £85,022 (FP generalised gamma). This finding illustrates that the cost-effectiveness results are largely driven by an assumed differential survival profile for nivolumab versus taxanes, and that the specification of the same functional form (even when deemed plausible) can lead to very large ICERs.

6. EVIDENCE REVIEW GROUP'S ADDITIONAL ANALYSES

6.1. Exploratory and sensitivity analyses undertaken by the ERG

The ERG conducted a number of additional sensitivity analyses within the company's model, which are summarised below:

- In order to test the influence of AEs costs on the cost-effectiveness of nivolumab vs taxanes, the ERG performed an exploratory analysis by removing AE costs.
- The ERG explored the impact of removing the company's application of background mortality (a multiplicative approach which is expected to lead to some double counting), and applied an alternative approach wherein the hazard of death is assumed to be no greater than that of the age- and sex-adjusted general population.
- Estimation of ToT is a key driver of cost-effectiveness results, yet the specification of any predominantly parametric-based approach does not provide a good fit to the Kaplan-Meier curve. As an alternative means of informing ToT within the model, the ERG implemented a 'pragmatic' approach wherein the Kaplan-Meier curve is used to inform the majority of the ToT curve, followed by an assumed fixed probability of discontinuation per model cycle.
- In two exploratory analyses, the ERG considered a 100% market share for docetaxel and a market share split based on ATTRACTION-3,[2] in order to understand the directional effect of results based on assumed market shares.
- Along with the key time-to-event outcomes (OS and ToT), the estimation of utility values is also a key driver of results. The ERG explored a number of scenarios using different utility values for both treatment arms and health states, and in particular focused on the impact of assuming the same utility values for each treatment arm (to establish the impact of a difference in utilities by arm on the model results).
- Due to a limited description provided in the CS, the ERG is concerned that some of the medical resource use costs have been miscalculated and/or referenced incorrectly. In addition, end-of-life costs may introduce an element of double counting. The ERG considered alternative costings for some medical resource use items. Also, much like the AE resolution analysis, the ERG explored a scenario wherein end-of-life costs were omitted from the model to establish its impact on model results.

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

The analyses described in Section 6.1 are described in turn within each section below.

6.2.1. Removal of adverse event costs

As highlighted in Section 4.2.8.4, the ERG considered the costs of AEs included in the company's base-case analysis to be relatively high, and therefore sought to establish the impact these costs had on the model results. By disabling AE costs, the company's base-case ICER increased from £45,491 to £47,671.

6.2.2. Approaches to reflecting background mortality

Disabling background mortality causes the base-case ICER to decrease from £45,491 to £42,299 (with all other parameters settings and assumptions unchanged). Owing to this relatively-large difference in the ICER, and acknowledging that each approach may be considered upper and lower bounds of the 'true' ICER, the ERG considered it appropriate to consider ICERs with background mortality included and excluded.

In addition, the ERG has explored an alternative application of background mortality within its exploratory analyses. In this alternative application, the per-cycle hazard of an OS or PFS event was capped by the risk of death in the general population such that the risk of an event must always be greater than or equal to background mortality:

$$S(t+1) = S(t) \times \max\left(\hat{h}_{Trial}(t), \hat{h}_{Background}(t)\right)$$

(Note: Estimated hazard based on proportion that experience event between cycles t and t + 1).

This application avoids potential double counting of mortality in the company's base-case analysis. The corresponding survival proportions over the course of the model time horizon are presented in Table 19 to illustrate the impact each approach has on survival extrapolation. As may be inferred from Table 19, all three approaches yield similar estimates on an absolute basis, but the ERG's alternative approach exhibits estimates of OS that are closer to the unadjusted OS model versus the company's base-case analysis.

Table 19: Summary of OS proportions based on background mortality approach

Time (years)	CS base-case analysis	ERG's background mortality approach	No adjustment for background mortality
1	45.6%	46.1%	46.1%
2	21.3%	21.7%	21.7%
3	12.3%	12.8%	12.8%
4	8.2%	8.5%	8.6%
5	5.8%	6.2%	6.2%
6	4.4%	4.8%	4.8%
7	3.4%	3.8%	3.8%
8	2.8%	3.1%	3.1%
9	2.3%	2.6%	2.6%
10	1.9%	2.3%	2.3%
20	0.4%	0.8%	0.8%
30	0.1%	0.2%	0.4%
40	0.0%	0.0%	0.0%

Abbreviations: SE, standard error.

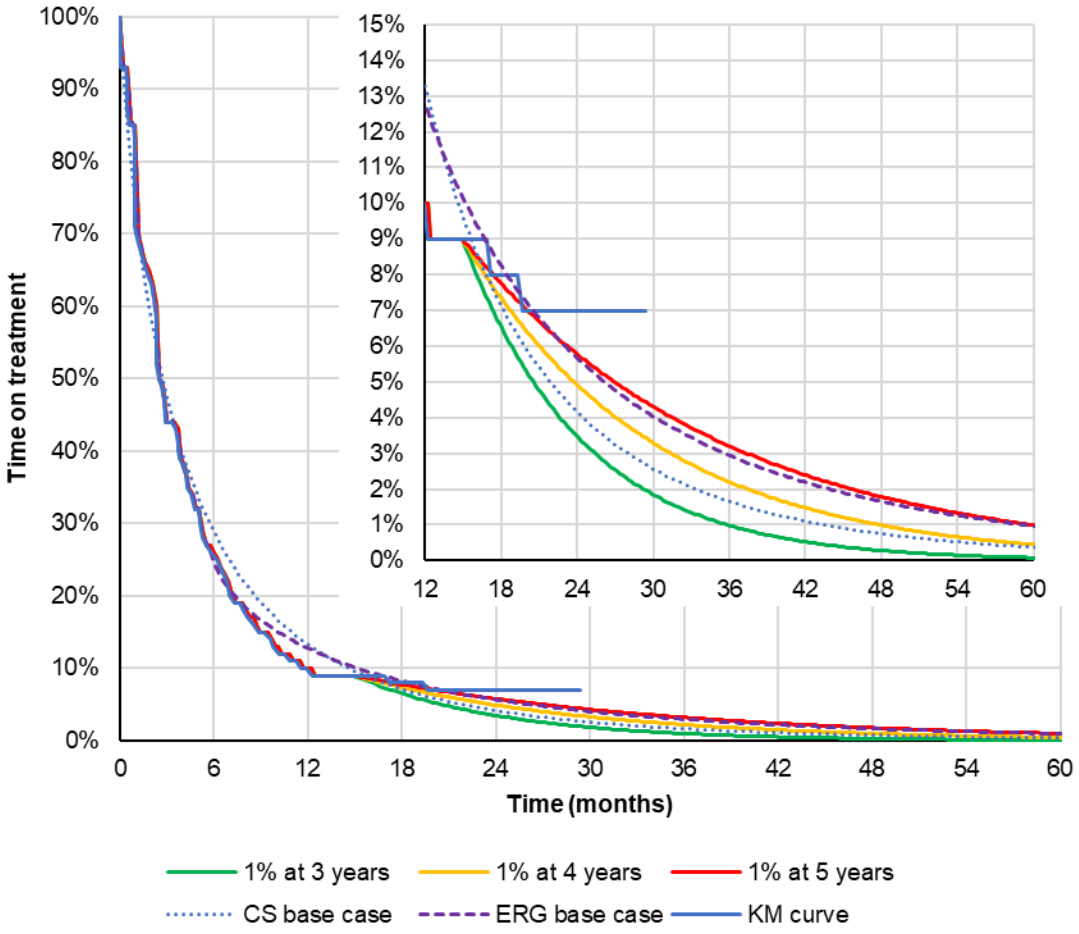
Source(s): CS Table 53.

By implementing the ERG's alternative background mortality approach, the company's base-case ICER decreased from £45,491 to £42,749. However, as highlighted in the CS (Section B.3.3.2.2), patients in ATTRACTION-3[2] are likely to be younger and fitter than the overall OSCC UK patient population. Therefore, even with an element of double counting, the company's base-case approach may still yield an over-estimate of OS. Consequently, the approach to adjusting for background mortality was not changed in the ERG's base-case analysis.

6.2.3. Pragmatic estimation of time on treatment

While the company provided a range of models for use in the economic model for ToT, the ERG considered an exploratory analysis wherein the Kaplan–Meier curve was used directly to inform ToT for nivolumab up until 15 months (an arbitrary time point close to the end of follow-up). After this point, an assumed monthly (constant) discontinuation probability was applied. Three values were considered, which were selected such that the ToT for the nivolumab arm became less than 1% (to two decimal places) at the model cycle following 36, 48, and 60 months (i.e. three, four, and five years). This analysis yielded the extrapolations presented in Figure 9.

Figure 9: Superimposition of ERG's exploratory ToT projections for nivolumab and Kaplan–Meier curve from ATTRACTION-3



Abbreviations: CS, company submission; KM, Kaplan–Meier.

Note(s): Figure produced by the ERG. Inset plot shows times after 12 months (to focus on the tail of the curve)

Source(s): Kaplan–Meier curves digitised based on reporting in the CS.

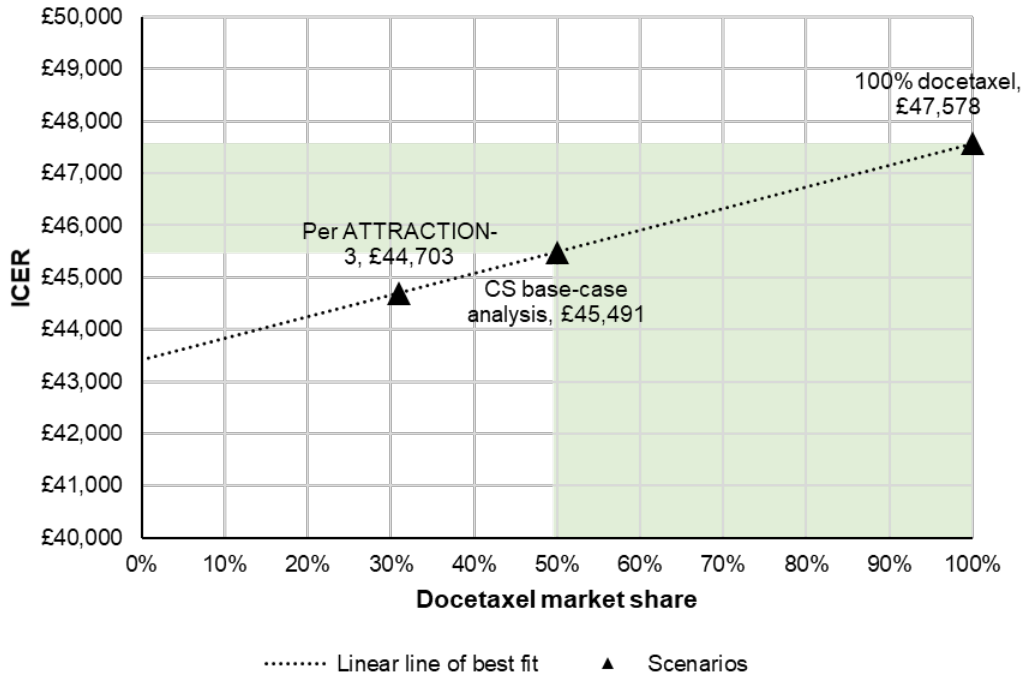
Based on the three values used, the equivalent ICERs were £41,501 (3 years), £45,323 (4 years), and £49,034 (five years); versus the company base-case ICER of £45,491. Therefore, while the three approaches explored by the ERG were identical for the first 15 months, the extrapolated portion of each curve caused the ICER to vary by a large quantity. It can also be seen from Figure 9 that the ERG's base-case analysis produces similar estimates to the '1% at five years' scenario, and so the ERG's selected model may be expected to provide a conservative estimate of the ICER (i.e. estimates a relatively large proportion of patients to continue treatment after the end of follow-up in ATTRACTION-3[2] versus the other projections shown).

6.2.4. Assumed market share of taxanes

Costing for the taxane arm in the CS utilised a 'simple average' of the costs for docetaxel and paclitaxel (i.e. a 50:50 market share). The ERG performed exploratory analysis whereby the taxane costs reflected the proportion of patients receiving docetaxel (n=65) and paclitaxel (n=144) in the pivotal trial ATTRACTION-3.[2] The subsequent ratio of approximately 31:69 in docetaxel versus paclitaxel treatment reduced the ICER to £44,703. However, based on clinical input, docetaxel is often preferred to paclitaxel within UK practice, and so an additional exploratory analysis was also performed whereby docetaxel costs were used to represent taxane therapy. This increased the ICER to £47,578.

Docetaxel is expected to be the most commonly-used taxane in UK practice, yet the ERG does not consider it plausible that exactly 50% or 100% of patients receive docetaxel, and that the true value lies somewhere between these bounds. To demonstrate the approximate relationship between the market share for docetaxel and the ICER, the ERG produced a line of best fit based on the three scenarios covered previously (50:50, 31:69, 100:0), shown in Figure 10.

Figure 10: Assumed market share for docetaxel versus ICER (company's base case)



Abbreviations: CS, company submission; ICER, incremental cost-effectiveness ratio.

Note(s): Figure produced by the ERG. Shaded region shows expected plausible range for docetaxel market share.

For the purpose of informing the ERG's preferred base-case analysis, the 50:50 split specified in the company's base-case analysis is left unchanged, with the understanding that this represents a potentially optimistic scenario (given that any increase in the market share for docetaxel would lead to an increase in the ICER).

6.2.5. Alternative utility values

To explore the impact of using different utility values to inform the model, the ERG considered an exploratory analysis wherein utility values were set to be identical across both arms, and were varied in increments of 0.02 (between 0.50 and 0.90). The results of this analysis are presented in Table 20, centred on the company's base-case analysis.

Table 20: ERG's exploratory analysis of utility values (no difference between arms)

		Progression-free utility (both arms)																				
		0.50	0.52	0.54	0.56	0.58	0.60	0.62	0.64	0.66	0.68	0.70	0.72	0.74	0.76	0.78	0.80	0.82	0.84	0.86	0.88	0.90
Progressed utility value (both arms)	0.50	£78k	£77k	£77k	£76k	£76k	£76k	£75k	£75k	£74k	£74k	£74k	£73k	£73k	£73k	£72k	£72k	£72k	£71k	£71k	£70k	£70k
	0.52		£75k	£74k	£74k	£74k	£73k	£73k	£72k	£72k	£72k	£71k	£71k	£71k	£70k	£70k	£70k	£69k	£69k	£69k	£68k	£68k
	0.54			£72k	£72k	£71k	£71k	£71k	£70k	£70k	£70k	£69k	£69k	£69k	£68k	£68k	£68k	£67k	£67k	£67k	£66k	£66k
	0.56				£69k	£69k	£69k	£68k	£68k	£68k	£67k	£67k	£67k	£66k	£66k	£66k	£66k	£65k	£65k	£65k	£64k	£64k
	0.58					£67k	£67k	£66k	£66k	£66k	£65k	£65k	£65k	£65k	£64k	£64k	£64k	£63k	£63k	£63k	£63k	£62k
	0.60						£65k	£64k	£64k	£64k	£64k	£63k	£63k	£63k	£63k	£62k	£62k	£62k	£61k	£61k	£61k	£61k
	0.62							£63k	£62k	£62k	£62k	£62k	£61k	£61k	£61k	£61k	£60k	£60k	£60k	£60k	£59k	£59k
	0.64								£61k	£60k	£60k	£60k	£60k	£59k	£59k	£59k	£59k	£58k	£58k	£58k	£58k	£58k
	0.66									£59k	£59k	£58k	£58k	£58k	£58k	£57k	£57k	£57k	£57k	£57k	£56k	£56k
	0.68										£57k	£57k	£57k	£56k	£56k	£56k	£56k	£56k	£55k	£55k	£55k	£55k
	0.70											£56k	£55k	£55k	£55k	£55k	£54k	£54k	£54k	£54k	£54k	£53k
	0.72												£54k	£54k	£54k	£53k	£53k	£53k	£53k	£53k	£52k	£52k
	0.74													£53k	£52k	£52k	£52k	£52k	£52k	£51k	£51k	£51k
	0.76														£51k	£51k	£51k	£51k	£50k	£50k	£50k	**£50k**
	0.78															**£50k**	**£50k**	**£49k**	**£49k**	**£49k**	**£49k**	**£49k**
	0.80																**£49k**	**£48k**	**£48k**	**£48k**	**£48k**	**£48k**
	0.82																	**£47k**	**£47k**	**£47k**	**£47k**	**£47k**
	0.84																		**£46k**	**£46k**	**£46k**	**£46k**
	0.86																			**£45k**	**£45k**	**£45k**
	0.88																				**£44k**	**£44k**
	0.90																					**£43k**

Abbreviations: k, thousand(s).

This analysis demonstrated that if utility values are assumed equal between arms, the company's base-case ICER would only be less than £50,000 per QALY gained if the progressed utilities were greater than or equal to 0.76. Considering that a progression-free utility of 0.90 is unlikely to be considered clinically plausible, the ERG noted that a progressed utility of 0.78 or greater would be needed. Therefore, the ERG does not consider there to be a clinically-

plausible scenario wherein the ICER would plausibly be less than £50,000 if utilities are set equal across arms (given that baseline utility in the taxanes arm was [redacted]).

The ERG performed additional exploratory analyses for which a range of other utility values were considered. Some of these analyses were based on adjustments to the values provided by the company (e.g. adjusting the values in the progressed disease state only), whereas others were not based on data from ATTRACTION-3[2] (and should therefore be considered illustrative). The results of these analyses are provided in Table 21.

Table 21: Utility values explored in exploratory analysis

Label	Nivolumab		Taxanes		ICER
	PF	PP	PF	PP	
CS base case	[redacted]	[redacted]	[redacted]	[redacted]	*45,491*
Average PP value	[redacted]	[redacted]	[redacted]	[redacted]	55,449
Minimum PP value	[redacted]	[redacted]	[redacted]	[redacted]	59,215
Custom* small benefit (both states)	[redacted]	[redacted]	[redacted]	[redacted]	58,830
Custom* moderate benefit (both states)	[redacted]	[redacted]	[redacted]	[redacted]	56,119
Custom* large benefit (both states)	[redacted]	[redacted]	[redacted]	[redacted]	53,646

Abbreviations: ICER, incremental cost-effectiveness ratio; PF, progression-free; PP, post-progression.

Note(s): Custom values not based on data from ATTRACTION-3, and should therefore be considered as illustrative.

6.2.6. Alternative medical resource use costs

The ERG could not validate two specific unit costs for medical resource use: clinician consultation and hospitalisation. When changing only the hospitalisation cost (to be consistent with the reporting in the CS in terms of referenced HRG codes), the ICER increased to £62,008, whereas changing only the clinician consultation increased the ICER marginally to £45,575 (CS base-case ICER £45,491). When the ERG replaced the company's values for both costs, this increased the ICER of the company's base-case analysis to £62,092.

The results above illustrated that the ICER was mainly affected by the difference in the unit cost per hospitalisation (£534.07 in the CS, versus £3,379.73 based on the ERG's calculation). The ERG has tentatively used both alternative unit costs to inform its base-case analysis, with the understanding that the incorrect unit costs were specified but the reference was correct.

The ERG also performed an exploratory analysis by repeating the base-case analysis from the CS with the omission of one-time end-of-life costs. This increased the ICER for nivolumab

versus taxanes to £45,853 (compared with the CS base-case analysis of £45,491). The small change in the ICER is due to the application of end-of-life costs as a lump sum upon death for all patients across both treatment arms (and therefore the only difference reflected across treatment arms is based on time preference).

6.2.7. Summary of ERG's additional clinical and economic analyses

A tabulated summary of the ERG's additional clinical and economic analyses is provided in Table 22.

Table 22: ERG's additional clinical and economic analyses summary

Analysis description	Section in ERG report	ICER £/QALY
Company base-case	***5.1.1***	***45,491***
Remove AE costs	*6.2.1*	*47,671*
ERG background mortality	*6.2.2*	*42,749*
Remove background mortality		*42,299*
Pragmatic ToT estimation (1% at 3 years)	*6.2.3*	*41,501*
Pragmatic ToT estimation (1% at 4 years)		*45,323*
Pragmatic ToT estimation (1% at 5 years)		*49,034*
ATTRACTION-3 taxane split	*6.2.4*	*44,703*
100% docetaxel		*47,578*
Average PP value	*6.2.5*	*55,449*
Minimum PP value		*59,215*
Custom* small benefit (both states)		*58,830*
Custom* moderate benefit (both states)		*56,119*
Custom* large benefit (both states)		*53,646*
Change clinician consultation cost	*6.2.6*	*45,575*
Change hospitalisation cost		*62,008*
Change both		*62,092*

Abbreviations: AE, adverse event; ERG, Evidence Review Group; ICER, incremental cost-effectiveness ratio; PP, post-progression; QALY, quality adjusted life year; ToT, time on treatment.

The ERG's preferred base-case settings lead to an ICER of £125,984 per QALY gained. A comparison of the component costs and QALYs that ultimately inform the company's and ERG's base-case results is provided in Table 26.

6.3. ERG's preferred assumptions

The ERG's preferred base-case analysis comprises several alternative model settings and assumptions, which are discussed in turn below. Six changes were made in total.

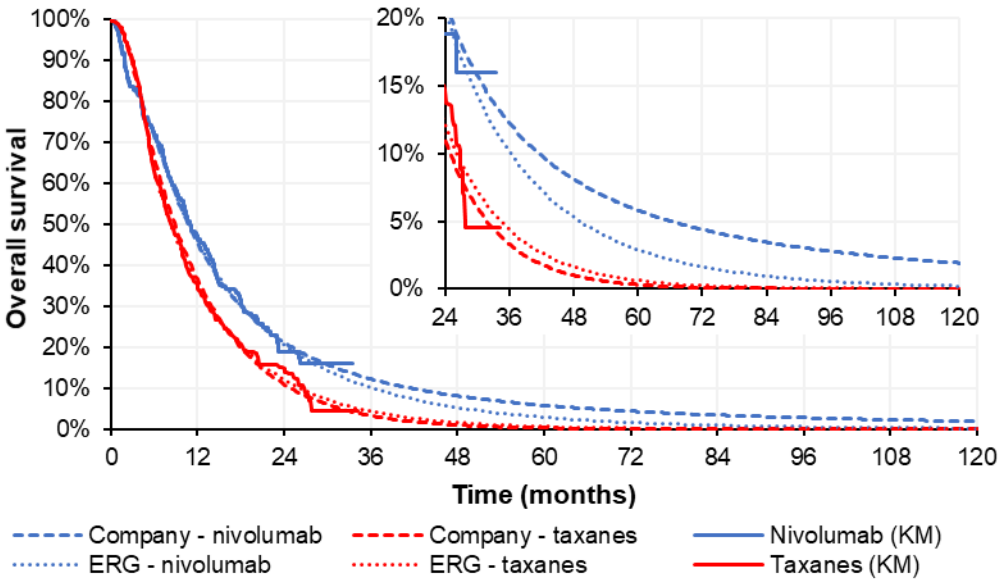
6.3.1. Choice of extrapolation for overall survival

The ERG's base-case analysis included the specification of an SP generalised gamma model for OS (cut point at 5.75 months) for both treatment arms. This model was chosen for both arms based on the following rationale:

- Models for each arm provide a similar, and potentially slightly-improved visual fit to the Kaplan–Meier curves versus the company's base-case analysis.
- The ERG considered the population from ATTRACTION-3[2] to likely exhibit better outcomes (e.g. total life-years accrued on either treatment arm) versus the UK population
  - Accordingly, the ERG considered a more conservative estimate of OS to be more suitable to inform its preferred base-case analysis.
- All patients are expected to have died by 10 years. 10-year OS in the company's base-case was estimated to be 1.92% for the nivolumab arm, versus 0.20% for the ERG's base-case analysis.
- The generalised gamma included the exponential model as a special case, and so similar estimates of OS for the taxanes arm should be expected versus the company's analysis if the 'true' survival of taxane patients is exponentially distributed. However, owing to its increased flexibility, the generalised gamma may better reflect non-constant hazards
  - A generalised gamma parameterisation has also been used to model OS for checkpoint inhibitors and taxanes in previously-published NICE appraisals (e.g. TA525[32]), and in the case of this appraisal yielded AIC scores within two points of the company's base-case models (suggesting similar statistical goodness-of-fit).

A comparison of the ERG's and company's base-case extrapolations of OS are presented in Table 23. The ERG acknowledged that the most appropriate selection of OS model is subject to debate in many appraisals; however, the ERG's preferred model was considered to represent a plausible, yet potentially conservative, estimate of the longer-term survival associated with nivolumab.

Table 23: Comparison of company- and ERG-preferred OS extrapolations

	Company	ERG
Description	SP approach using Kaplan–Meier cure until 2.99 months, followed by a log-logistic (nivolumab) or exponential (taxanes) model	SP approach using Kaplan–Meier cure until 5.75 months, followed by a generalised gamma model (both arms)
Plot		

Time (years)	Nivolumab	Taxanes	Nivolumab	Taxanes
1	45.61%	36.57%	46.07%	35.40%
2	21.27%	11.06%	20.70%	12.20%
3	12.33%	3.34%	10.22%	4.42%
4	8.15%	1.01%	5.36%	1.65%
5	5.84%	0.30%	2.93%	0.63%
6	4.39%	0.09%	1.63%	0.24%
8	2.78%	0.01%	0.56%	0.04%
10	1.92%	0.00%	0.20%	0.01%

Abbreviations: KM, Kaplan–Meier; SP, semi-parametric

Long-term survival with checkpoint inhibitors in an OSCC population has not been established, and the generalisability of the ATTRACTION-3[2] population to the anticipated UK patient population is unclear. The ERG acknowledged that a 'plateau' in the survival curve may be plausible, though currently unclear of what magnitude such a plateau may be.

Given the impact of long-term survival estimation on the ICER, further data collection may be warranted before a more stable estimate of OS (and hence the ICER) may be obtained, as well as to establish the generalisability of outcomes to the UK patient population.

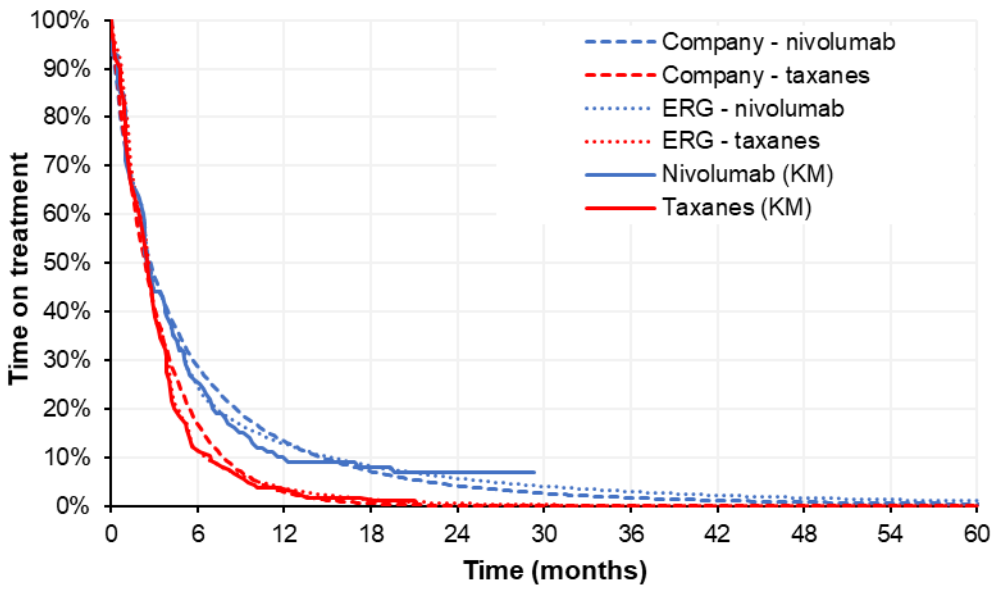
6.3.2. Choice of modelling approach for treatment discontinuation

The ERG's base-case analysis included the specification of an SP Weibull model for ToT (cut point at 5.75 months) for both treatment arms. This model was chosen for both arms based on the following rationale:

- Both models provide an improved visual fit to the Kaplan–Meier curves versus the company's base-case analysis.
- Longer-term extrapolations of ToT are similar to the company's base-case analysis.
- The Weibull model was considered to provide a more realistic pattern of longer-term discontinuation versus the generalised gamma model.

A comparison of the ERG's and company's base-case extrapolations of ToT are presented in Table 24. As with the outcome of OS, the ERG acknowledged that the most appropriate selection of model is subject to debate. Therefore, a range of alternative extrapolations for ToT may be important to consider, and further data collection may be warranted to better understand the likely pattern of treatment discontinuation (both from ATTRACTION-3[2] and in NHS practice).

Table 24: Comparison of company- and ERG-preferred ToT extrapolations

	Company	ERG
Description	FP approach using generalised gamma model (both arms)	SP approach using Kaplan–Meier cure until 5.75 months, followed by a Weibull model (both arms)
Plot		

Time (years)	Nivolumab	Taxanes	Nivolumab	Taxanes
1	13.36%	2.90%	12.80%	3.68%
2	4.19%	0.08%	5.69%	0.51%
3	1.67%	0.00%	2.96%	0.08%
4	0.76%	0.00%	1.66%	0.01%
5	0.38%	0.00%	0.98%	0.00%

Abbreviations: FP, fully-parametric; KM, Kaplan–Meier; SP, semi-parametric

The ERG highlighted that alternative, flexible approaches may produce more credible estimations of the 'true' ToT curve, but were not provided to the ERG. While the ToT Kaplan–Meier curves were relatively complete, any longer-term estimates of ToT may be considered somewhat arbitrary in light of the small number of events beyond 12 months in ATTRACTION-3.[2]

6.3.3. Application of treatment acquisition costs for taxanes

As described in Section 4.2.8.1, the CS included costs for docetaxel and paclitaxel from MIMS. The ERG did not consider these costs to be reflective of the average cost paid by the NHS for

these treatments, and so in its preferred base-case analysis corrected these values based on information reported in the eMIT.

6.3.4. Adjustment of treatment administration costs

Due to the expectation that administration of nivolumab will take place over 30 minutes, versus at least 60 minutes for taxanes, the ERG's preferred base-case analysis includes the specification of a higher administration cost for taxanes versus nivolumab. In addition, the ERG's preferred costs specify that administration is expected to occur in a day case setting.

6.3.5. Specification of health-state utility values

The ERG had a number of concerns with the company's base-case utility analysis. While also subject to limitations, the ERG considered the mixed-effects regression approach to be more appropriate for use in the economic model, and so has applied these values to inform its preferred base-case analysis.

In addition, the ERG did not consider it appropriate to specify utility values that exhibit a large difference in utility for PD patients dependent on initial treatment assignment. The ERG noted that differences in utility by treatment arm after progression may be due to a combination of potential continued benefit from nivolumab after progression, or as a direct consequence of the open-label design of ATTRACTION-3.[2] In the ERG's preferred base-case analysis, the average of the PD utility values per arm is assumed to apply for both treatment arms.

The ERG emphasised that the determination of the most appropriate utility values for use within the model is subject to debate. The approach taken to inform the ERG's base-case analysis may be considered in some respects conservative (with respect to the assumed lack of difference between arms beyond progression) and in others, optimistic (given that the difference between arms in the PF state is unchanged, and is expected in part to be related to the differences seen in utility at screening, as well as the open label design of ATTRACTION-3[2]).

6.3.6. Update of unit costs for outpatient consultation and hospitalisation

The unit costs for outpatient consultation and hospitalisation did not match the references cited in the CS. The ERG considered the references cited to appear sensible, and therefore opted to amend the costs used to inform the model to ensure alignment with the references cited in the CS. This affected the cost of an outpatient consultation (£187.36 in CS, £196.33 in ERG base case) and hospitalisation (£534.07 in CS, £3,379.73 in ERG base case).

6.3.7. Summary of ERG's base-case settings and results

A tabulated summary of the ERG's preferred model settings and assumptions is provided in Table 25, alongside the incremental impact each setting has on the ICER.

Table 25: ERG's preferred model assumptions

Preferred assumption	Section in ERG report	Cumulative ICER £/QALY
Company base-case	***5.1.1***	***45,491***
SP generalised gamma (5.75) OS models	*6.3.1*	*62,440*
SP Weibull (5.75) ToT models	*6.3.2*	*68,343*
Correction of taxanes costs	*6.3.3*	*80,614*
ERG's preferred administration costs	*6.3.4*	*77,198*
ERG's preferred utility values	*6.3.5*	*106,643*
Update of unit costs for MRU	*6.3.6*	*125,984*

Abbreviations: ERG, Evidence Review Group; ICER, incremental cost-effectiveness ratio; MRU, medical resource use; QALY, quality adjusted life year

The ERG's preferred base-case settings lead to an ICER of £125,984 per QALY gained. A comparison of the component costs and QALYs that ultimately inform the company's and ERG's base-case results is provided in Table 26.

Table 26: Comparison of company's and ERG's base case results

Arm	Total			Incremental			ICER (£/QALY)
	Costs (£)	LYs	QALYs	Costs (£)	LYs	QALYs	
Company base-case (deterministic)							
Taxane	[redacted]	[redacted]	[redacted]				
Nivolumab	[redacted]	[redacted]	[redacted]	20,842	0.536	0.458	45,491
ERG base-case (deterministic)							
Taxane	[redacted]	[redacted]	[redacted]				
Nivolumab	[redacted]	[redacted]	[redacted]	27,845	0.302	0.221	125,984

Abbreviations: ICER, incremental cost-effectiveness ratio; LY, life year; QALY, quality adjusted life year.

Source(s): CS Table 75.

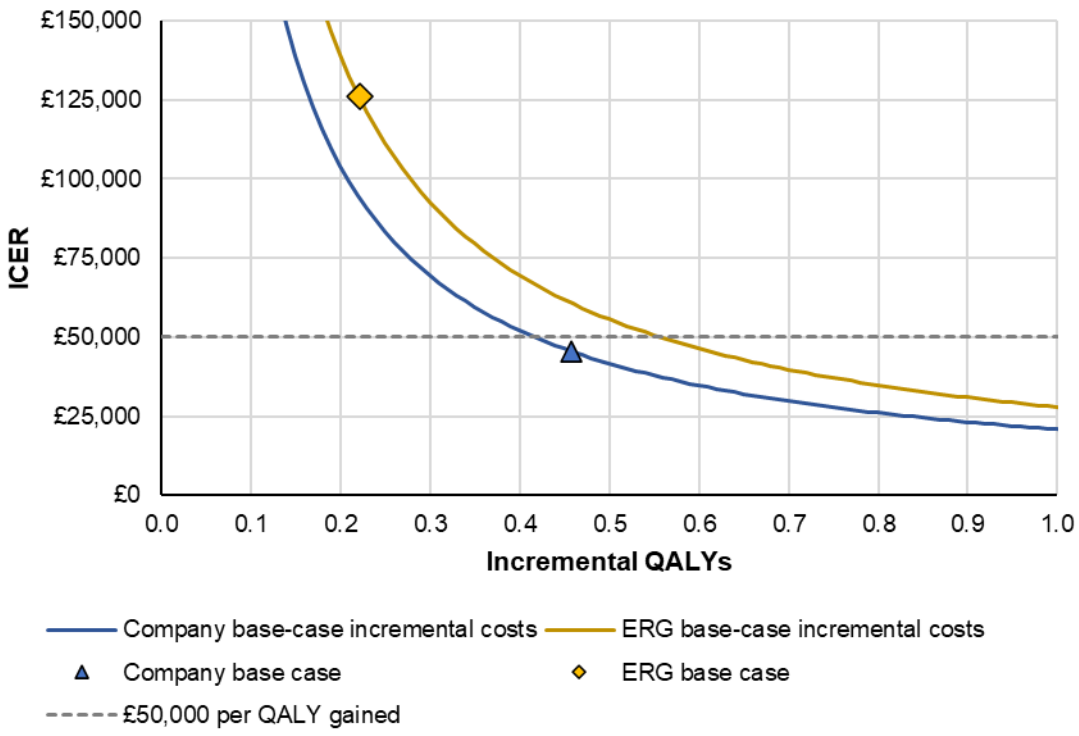
From Table 26, it can be seen that the incremental costs have increased by approximately £7,000, primarily due to the increase in MRU costs for both arms, and the reduction in acquisition costs for taxanes. The (discounted) life-year gain has reduced from 0.536 to 0.302,

due to the specification of alternative models for OS that do not exhibit as large of a survival plateau. Finally, the incremental QALY gain has decreased from 0.458 to 0.221, mostly due to a combination of the OS models selected and the specification of a single utility value for PD.

Two key aspects of the ERG's preferred base-case analysis were particularly influential on cost-effectiveness results: (a) the choice of OS model, and (b) setting the same utility value for the PD state. The ERG's preferred base-case ICER (£125,984) would decrease to £90,758 if the OS models are unchanged from the company's base-case analysis. Equivalently, if the company's original utility values were used, the ERG's preferred base-case ICER would decrease to £91,198. Combining both changes (i.e. the ERG's preferred base-case, but with the company's preferred OS models and utility values), leads to an ICER of £71,964.

The ERG highlighted that a small change in the incremental QALYs could lead to a large change in the ICER (given the magnitude of incremental costs). For context, Figure 11 demonstrates the hypothetical relationship between incremental QALYs and the ICER based on fixed incremental costs from the company's and ERG's base-case ICERs.

Figure 11: Hypothetical QALYs versus ICER relationship



Abbreviations: ERG, Evidence Review Group; ICER, incremental cost-effectiveness ratio; QALY, quality-adjusted life year.

6.4. Conclusions of the cost-effectiveness section

The company's model appropriately reflects the decision problem set out by NICE within its final scope, and any deviations have been adequately justified

The company's PartSA model broadly adhered to the decision problem set out by NICE within the final scope of this appraisal. The key deviation from the scope was to focus predominantly on the comparator of 'taxanes', provide an exploratory comparison to BSC, and disregard irinotecan as a comparator. The ERG did not consider it possible to establish a robust comparison of nivolumab to BSC with available evidence, and agreed with the company's decision to provide this as an exploratory analysis only. For irinotecan, the ERG understood that a small proportion of patients may be treated with this intervention in NHS practice, but that this is not representative of the standard of care, and no robust data are available to appropriately compare to nivolumab.

The systematic literature reviews were appropriately conducted, though discussion surrounding how findings were applied within the submitted model was brief

The company used three separate search strategies to identify existing, relevant cost-effectiveness, HRQoL, and cost and healthcare resource use evidence. The ERG was generally satisfied that the company's literature review identified all potentially-relevant studies, yet it was unclear to the ERG how these studies were subsequently used to inform the CS. At clarification stage, the company provided further information concerning how specific studies were integrated in the model design, inputs and/or related assumptions; though the commentary provided was limited.

ATTRACTION-3 is a well-designed RCT of nivolumab versus taxanes, yet the generalisability of this study to the UK population is unclear

The ERG considered the ATTRACTION-3[2] study to be a high-quality, RCT of nivolumab versus taxanes conducted in an OSCC population. However, as described in relation to the clinical effectiveness evidence earlier in this report (Section 3.6), close to 100% of patients were Asian (of which approximately two-thirds were from Japan). Therefore, the ERG considered it highly likely that UK patients would not achieve equivalent outcomes versus the ATTRACTION-3 population (most notably, the Japanese subgroup). The ERG was unable to perform additional analyses based on the ROW population from ATTRACTION-3,[2] as the necessary data were not provided at clarification stage. However, analyses based on ROW population may constitute a

more suitable basis for decision making, and the generalisability of the ATTRACTION-3[2] population remains a key area of uncertainty.

Estimation of OS and ToT are both key drivers of cost-effectiveness results

As is the case for a number of economic evaluations of cancer interventions, estimation of OS and ToT constitute two of the main drivers of results. Based on the potential generalisability issues highlighted concerning ATTRACTION-3,[2] as well as a lack of established long-term survival in an OSCC population, the ERG considered it appropriate to consider a range of OS extrapolations, though a more conservative approach has been adopted to inform the ERG's base-case analysis. For ToT, a SP approach was preferred to inform the ERG's base-case, based predominantly on improved fit to the Kaplan-Meier curve. Alternative models for both outcomes may be important for consideration in decision making.

The company's approach to elicit health-state utility values is unconventional and subject to substantial uncertainty, especially due to the open-label design of ATTRACTION-3

The company estimated utility values based on EQ-5D data from ATTRACTION-3[2] using a range of imputation methods. The ERG highlighted a number of concerns with the approach taken to estimate utilities, and noted that this approach has not been used extensively in previous appraisals conducted by NICE. The ERG instead preferred utility values estimated through a mixed-effects regression model provided at clarification stage, yet these values are still subject to limitations based on the open-label design of ATTRACTION-3 (risk of bias) and the face validity of large difference in utility by treatment arm and progression status.

Some unit costs used in the company model were not deemed the most relevant to NHS practice or were misaligned with the cited reference(s)

Acquisition costs for the taxanes were identified from MIMS, yet the ERG noted a more appropriate reference would have been the eMIT database which reflects the true price paid by NHS trusts. In addition, the ERG considered it likely that on a per-administration basis, nivolumab is expected to be associated with a reduced administration cost based on less chair time (30 minutes versus one hour or more for taxanes). These edits were reflected in the ERG's preferred base-case analysis, along with the alignment of medical resource use costs that did not match the stated reference source(s).

The company's sensitivity analyses were subject to a number of limitations, and therefore scenario analyses are considered more suitable for informing decision making

The company provided a number of sensitivity analyses to explore the impact of different inputs on model results, though the ERG identified a number of flaws in the approaches taken to quantify parameter uncertainty within the model. The company's DSA and PSA are thus not considered to represent a reliable presentation of the parameter uncertainty inherent within the model. Consequently, the ERG's report focused predominantly on the exploration of a range of scenario analyses to 'stress test' various settings and assumptions. The ERG's preferred base-case analysis has been implemented in one of the models provided by the company, however it was beyond the remit of the ERG to combine analyses provided in 16 different model files in order to fully explore the impact of all potentially-related model settings and assumptions.

The ERG's preferred base-case analysis yields an ICER greater than the company's base-case ICER, and is in excess of £50,000 per QALY gained

The ERG's preferred base-case analysis includes alternative time-to-event models, utility values, and costs for treatment administration, acquisition (taxanes only), and MRU. When combined, these changes resulted in larger total costs and fewer incremental QALYs, causing an increase in the ICER from £45,491 to £125,984. Assuming a willingness-to-pay threshold of £50,000 per QALY gained, nivolumab would not be considered a cost-effective use of NHS and PSS resources based on the ERG's preferred base-case analysis.

7. END OF LIFE

NICE's end of life criteria are said to be met if both of the following apply:

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

As highlighted in the CS, average survival for patients with unresectable, advanced oesophageal cancer when standard chemotherapy has failed is less than 24 months, both in terms of median survival (approximately 8.4 months) and mean survival projected by the company's economic model base-case analysis (approximately 12.0 months). The ERG considered it highly unlikely that any plausible alternative extrapolation of OS for the taxanes arm would yield an estimated mean survival time of more than double the company's base-case analysis. Furthermore, the ERG notes that survival outcomes seen in UK clinical practice may be poorer compared to the ATTRACTION-3[2] trial population (discussed further in Section 3.2.1). Consequently, the ERG considered the first life expectancy criterion to be likely met.

ATTRACTION-3[2] demonstrated an added 2.5 months of median OS benefit for nivolumab versus taxanes. When interpreting this difference in median OS benefit, it is important to note that estimates of median improvement are unlikely to provide an accurate reflection of the average (mean) benefit accrued by an average patient, both in general (owing to the median providing a point estimate of survival) and in the case of ATTRACTION-3[2] specifically due to the OS curves crossing at approximately four months.

Nevertheless, in terms of mean benefit, the company's base-case analysis yielded an estimated 7.8 months of added survival for patients treated with nivolumab (i.e. an incremental, undiscounted life-year gain of 0.653). This benefit is more than double the three-month minimum (i.e. 0.250 life-years) required in order for nivolumab to meet the second criterion. However, as shown in the range of sensitivity analyses presented by the company and the ERG, the choice of survival extrapolation has the potential to reduce the survival benefit associated with nivolumab markedly. In the ERG's base-case analysis, nivolumab is associated with an estimated 4.0 months of added survival (an undiscounted life-year gain of 0.333). Consequently, while the estimation of life extension is uncertain, both the company's and the ERG's preferred base-case analysis yielded an extension to life of at least three months.

Importantly however, the ERG noted that its preferred extrapolation is also subject to substantial uncertainty owing to limited follow-up available from the ATTRACTION-3[2] study at this time. In addition, there are several generalisability issues concerning the ATTRACTION-3[2] study which may impact the expected extension to survival attributable to nivolumab [redacted] The inferences made above are therefore provided for context, and ultimately the decision as to whether or not nivolumab meets NICE's end-of-life criteria rests with the appraisal committee.

References

1. National Institute for Health and Care Excellence (NICE). Nivolumab for previously treated unresectable advanced oesophageal cancer. Final scope, 2020. Available from: https://www.nice.org.uk/guidance/indevelopment/gid-ta10222/documents.
2. Kato K, Cho BC, Takahashi M, Okada M, Lin CY, Chin K, et al. Nivolumab versus chemotherapy in patients with advanced oesophageal squamous cell carcinoma refractory or intolerant to previous chemotherapy (ATTRACTION-3): a multicentre, randomised, open-label, phase 3 trial. Lancet Oncology. 2019;20(11):1506-17.
3. Cancer Research UK. What is oesophageal cancer? 2019. Available from: https://www.cancerresearchuk.org/about-cancer/oesophageal-cancer/about.
4. Cancer Research UK. Oesophageal cancer survival, 2016. Available from: https://www.cancerresearchuk.org/about-cancer/oesophageal-cancer/survival.
5. Cancer Research UK. Oesophageal cancer incidence statistics, 2019. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/oesophageal-cancer/incidence.
6. Cancer Research UK. Oesophageal cancer survival statistics, 2014. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/oesophageal-cancer/survival.
7. National Institute for Health and Care Excellence (NICE). Oesophago-gastric cancer: assessment and management in adults. NICE guideline [NG83], 2018. Available from: https://www.nice.org.uk/guidance/ng83.
8. Kudo T, Hamamoto Y, Kato K, Ura T, Kojima T, Tsushima T, et al. Nivolumab treatment for oesophageal squamous-cell carcinoma: an open-label, multicentre, phase 2 trial. Lancet Oncology. 2017;18(5):631-9.
9. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. Journal of Epidemiology and Community Health. 1998;52(6):377-84.
10. Muro K, Lordick F, Tsushima T, Pentheroudakis G, Baba E, Lu Z, et al. Pan-Asian adapted ESMO Clinical Practice Guidelines for the management of patients with metastatic oesophageal cancer: a JSMO-ESMO initiative endorsed by CSCO, KSMO, MOS, SSO and TOS. Annals of Oncology. 2019;30(1):34-43.
11. Lordick F, Mariette C, Haustermans K, Obermannová R, Arnold D. Oesophageal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of Oncology. 2016;27(Suppl 5):v50-v7.
12. Ono Pharmaceutical Co. Ltd. ONO-4538 Phase III Study. A multicenter, randomized, open-label study in patients with esophageal cancer refractory or intolerant to combination therapy with fluoropyrimidine- and platinum-based drugs - study report. 2019.
13. Moriwaki T, Kajiwara T, Matsumoto T, Suzuki H, Hiroshima Y, Matsuda K, et al. Survival analysis of platinum-refractory patients with advanced esophageal cancer treated with docetaxel or best supportive care alone: a retrospective study. Diseases of the Esophagus. 2014;27(8):737-43.
14. Nakatsumi H, Komatsu Y, Sawada K, Muranaka T, Kawamoto Y, Yuki S, et al. Retrospective comparison of efficacy and safety of docetaxel and weekly-paclitaxel as 2nd-line chemotherapy for patients with unresectable or recurrent esophageal cancer. Annals of Oncology. 2016;27:50.
15. Shirakawa T, Kato K, Nagashima K, Nishikawa A, Sawada R, Takahashi N, et al. A retrospective study of docetaxel or paclitaxel in patients with advanced or recurrent esophageal squamous cell carcinoma who previously received fluoropyrimidine- and platinum-based chemotherapy. Cancer Chemotherapy and Pharmacology. 2014;74(6):1207-15.

16. Janmaat VT, Bruno MJ, Polinder S, Lorenzen S, Lordick F, Peppelenbosch MP, et al. Cost-effectiveness of cetuximab for advanced esophageal squamous cell carcinoma. PLoS One. 2016;11(4):e0153943.
17. Bascoul-Mollevi C, Gourgou S, Galais MP, Raoul JL, Bouche O, Douillard JY, et al. Health-related quality of life results from the PRODIGE 5/ACCORD 17 randomised trial of FOLFOX versus fluorouracil-cisplatin regimen in oesophageal cancer. European Journal of Cancer. 2017;84:239-49.
18. Doherty MK, Leung Y, Su J, Naik H, Patel D, Eng L, et al. Health utility scores from EQ-5D and health-related quality of life in patients with esophageal cancer: a real-world cross-sectional study. Diseases of the Esophagus. 2018;31(12):doy058.
19. Dutton SJ, Ferry DR, Blazeby JM, Abbas H, Dahle-Smith A, Mansoor W, et al. Gefitinib for oesophageal cancer progressing after chemotherapy (COG): a phase 3, multicentre, double-blind, placebo-controlled randomised trial. Lancet Oncology. 2014;15(8):894-904.
20. Shenfine J, McNamee P, Steen N, Bond J, Griffin SM. A randomized controlled clinical trial of palliative therapies for patients with inoperable esophageal cancer. American Journal of Gastroenterology. 2009;104(7):1674-85.
21. Tian D, Wen H, Fu M. Comparative study of self-expanding metal stent and intraluminal radioactive stent for inoperable esophageal squamous cell carcinoma. World Journal of Surgical Oncology. 2016;14(1):18.
22. Xinopoulos D, Dimitroulopoulos D, Tsamakidis K, Korkolis D, Fotopoulou A, Bazinis A, et al. Palliative treatment of advanced esophageal cancer with metal-covered expandable stents. A cost-effectiveness and quality of life study. Journal of BUON. 2005;10(4):523-8.
23. Dallal HJ, Smith GD, Grieve DC, Ghosh S, Penman ID, Palmer KR. A randomized trial of thermal ablative therapy versus expandable metal stents in the palliative treatment of patients with esophageal carcinoma. Gastrointestinal Endoscopy. 2001;54(5):549-57.
24. Malik V, Johnston C, Donohoe C, Claxton Z, Lucey J, Ravi N, et al. (18)F-FDG PET-detected synchronous primary neoplasms in the staging of esophageal cancer: incidence, cost, and impact on management. Clinical Nuclear Medicine. 2012;37(12):1152-8.
25. O'Donnell CA, Fullarton GM, Watt E, Lennon K, Murray GD, Moss JG. Randomized clinical trial comparing self-expanding metallic stents with plastic endoprostheses in the palliation of oesophageal cancer. British Journal of Surgery. 2002;89(8):985-92.
26. O'Donnell CA, Gray J, Hodgson H, Macpherson M, Zammit M, Fullarton G. A cost comparison of photodynamic therapy and metallic stents in the palliation of oesophageal cancer. Photodiagnosis and photodynamic therapy. 2007;4(1):65-70.
27. Russell IT, Edwards RT, Gliddon AE, Ingledew DK, Russell D, Whitaker R, et al. Cancer of Oesophagus or Gastricus - New Assessment of Technology of Endosonography (COGNATE): report of pragmatic randomised trial. Health Technology Assessment. 2013;17(39):1-170.
28. Tralau-Stewart L, Roy R. Radiology-guided oesophageal stenting for the palliation of dysphagia: A single center experience. Annals of Oncology. 2019;30:aa23.
29. Woods B, Sideris E, Palmer S, Latimer N, Soares M. NICE DSU Technical Support Document 19: partitioned survival analysis for decision modelling in health care: a critical review, 2017. Sheffield, UK: Decision Support Unit, ScHARR. Available from: http://nicedsu.org.uk/wp-content/uploads/2017/06/Partitioned-Survival-Analysis-final-report.pdf.
30. Watanabe M. Recent topics and perspectives on esophageal cancer in Japan. JMA Journal. 2018;1(1):30-9.
31. National Institute for Health and Care Excellence (NICE). Technology Appraisal Guidance [TA428]. Pembrolizumab for treating PD-L1-positive non-small-cell lung cancer after chemotherapy. Final appraisal determination, 2016. Available from: https://www.nice.org.uk/guidance/ta428/documents/final-appraisal-determination-document.

32. National Institute for Health and Care Excellence (NICE). Technology appraisal guidance [TA525]. Atezolizumab for treating locally advanced or metastatic urothelial carcinoma after platinum-containing chemotherapy. Final appraisal determination, 2018. Available from: https://www.nice.org.uk/guidance/ta525/documents/final-appraisal-determination-document.
33. Davies A, Briggs A, Schneider J, Levy A, Ebeid O, Wagner S, et al. The ends justify the mean: outcome measures for estimating the value of new cancer therapies. Health Outcomes Research in Medicine. 2012;3(1):e25-e36.
34. Office for National Statistics. National life tables: UK, 2019. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesunitedkingdomreferencetables.
35. Dolan P. Modeling valuations for EuroQol health states. Medical Care. 1997;35(11):1095-108.
36. Simons CL, Rivero-Arias O, Yu LM, Simon J. Multiple imputation to deal with missing EQ-5D-3L data: should we impute individual domains or the actual index? Quality of Life Research. 2015;24(4):805-15.
37. National Institute for Health and Care Excellence (NICE). Technology appraisal guidance [TA378]. Ramucirumab for treating advanced gastric cancer or gastro–oesophageal junction adenocarcinoma previously treated with chemotherapy, 2016. Available from: https://www.nice.org.uk/guidance/TA378.
38. Ara R, Brazier JE. Using health state utility values from the general population to approximate baselines in decision analytic models when condition-specific data are not available. Value in Health. 2011;14(4):539-45.
39. National Institute for Health and Care Excellence (NICE). Technology appraisal guidance [TA519]. Pembrolizumab for treating locally advanced or metastatic urothelial carcinoma after platinum-containing chemotherapy, 2018. Available from: https://www.nice.org.uk/guidance/ta519.
40. National Institute for Health and Care Excellence (NICE). Technology appraisal guidance [TA519]. Pembrolizumab for treating locally advanced or metastatic urothelial carcinoma after platinumcontaining chemotherapy. Final appraisal determination, 2018. Available from: https://www.nice.org.uk/guidance/ta519/documents/final-appraisal-determination-document.
41. National Institute for Health and Care Excellence (NICE). Technology appraisal guidance [TA490]. Nivolumab for treating squamous cell carcinoma of the head and neck after platinum-based chemotherapy. Final appraisal determination, 2017. Available from: https://www.nice.org.uk/guidance/ta490/documents/final-appraisal-determination-document.
42. National Institute for Health and Care Excellence (NICE). Technology appraisal guidance [TA483]. Nivolumab for previously treated squamous non-small-cell lung cancer. Final appraisal determination, 2017. Available from: https://www.nice.org.uk/guidance/ta483/documents/final-appraisal-determination-document.
43. Grothey A, Van Cutsem E, Sobrero A, Siena S, Falcone A, Ychou M, et al. Regorafenib monotherapy for previously treated metastatic colorectal cancer (CORRECT): an international, multicentre, randomised, placebo-controlled, phase 3 trial. Lancet. 2013;381(9863):303-12.
44. British National Formulary (BNF). Docetaxel, 2020. Available from: https://bnf.nice.org.uk/medicinal-forms/docetaxel.html.
45. British National Formulary (BNF). Paclitaxel, 2020. Available from: https://bnf.nice.org.uk/medicinal-forms/paclitaxel.html.
46. Department of Health and Social Care. eMIT national database. Drugs and pharmaceutical electronic market information tool, 2020. Available from: https://www.gov.uk/government/publications/drugs-and-pharmaceutical-electronic-market-information-emit.

47. NHS Improvement. 2018/19 National Cost Collection data, 2020. Available from: https://improvement.nhs.uk/resources/national-cost-collection/#ncc1819.
48. Campbell H, Stokes E, Bargo D, Logan R, Mora A, Hodge R, et al. Costs and quality of life associated with acute upper gastrointestinal bleeding in the UK: cohort analysis of patients in a cluster randomised trial. BMJ Open. 2015;5(4):e007230.
49. White J, Carolan-Rees G. PleurX peritoneal catheter drainage system for vacuum-assisted drainage of treatment-resistant, recurrent malignant ascites. Applied Health Economics and Health Policy. 2012;10(5):299-308.
50. Georghiou T, Bardsley M. Exploring the cost of care at the end of life. London: Nuffield Trust Research Report. 2014.
51. Copley-Merriman C, Stevinson K, Liu FX, Wang J, Mauskopf J, Zimovetz EA, et al. Direct costs associated with adverse events of systemic therapies for advanced melanoma: Systematic literature review. Medicine (Baltimore). 2018;97(31):e11736.
52. Lewis G, Peake M, Aultman R, Gyldmark M, Morlotti L, Creeden J, et al. Cost-effectiveness of erlotinib versus docetaxel for second-line treatment of advanced non-small-cell lung cancer in the United Kingdom. Journal of International Medical Research. 2010;38(1):9-21.
53. Khan I, Morris S, Hackshaw A, Lee SM. Cost-effectiveness of first-line erlotinib in patients with advanced non-small-cell lung cancer unsuitable for chemotherapy. BMJ Open. 2015;5(7):e006733.
54. (NICE) NIfHaCE. Technology appraisal guidance [TA628]. Lorlatinib for previously treated ALK-positive advanced non-small-cell lung cancer, 2020. Available from: https://www.nice.org.uk/guidance/ta628.
55. Wehler E, Zhao Z, Pinar Bilir S, Munakata J, Barber B. Economic burden of toxicities associated with treating metastatic melanoma in eight countries. European Journal of Health Economics. 2017;18(1):49-58.
56. Briggs A. Probabilistic analysis of cost-effectiveness models: statistical representation of parameter uncertainty. Value in Health. 2005;8(1):1-2.
57. Rohatgi A. WebPlotDigitizer, version 4.2, 2019. Available from: https://automeris.io/WebPlotDigitizer.
58. Latimer N. NICE DSU Technical Support Document 14: survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data, 2013. Sheffield, UK: Decision Support Unit, ScHARR. Available from: http://nicedsu.org.uk/wp-content/uploads/2016/03/NICE-DSU-TSD-Survival-analysis.updated-March-2013.v2.pdf.
1. National Institute for Health and Care Excellence (NICE). Nivolumab for previously treated unresectable advanced oesophageal cancer. Final scope, 2020. Available from: https://www.nice.org.uk/guidance/indevelopment/gid-ta10222/documents.
2. Kato K, Cho BC, Takahashi M, Okada M, Lin CY, Chin K, et al. Nivolumab versus chemotherapy in patients with advanced oesophageal squamous cell carcinoma refractory or intolerant to previous chemotherapy (ATTRACTION-3): a multicentre, randomised, open-label, phase 3 trial. Lancet Oncology. 2019;20(11):1506-17.
3. Cancer Research UK. What is oesophageal cancer? 2019. Available from: https://www.cancerresearchuk.org/about-cancer/oesophageal-cancer/about.
4. Cancer Research UK. Oesophageal cancer survival, 2016. Available from: https://www.cancerresearchuk.org/about-cancer/oesophageal-cancer/survival.
5. Cancer Research UK. Oesophageal cancer incidence statistics, 2019. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/oesophageal-cancer/incidence.
6. Cancer Research UK. Oesophageal cancer survival statistics, 2014. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/oesophageal-cancer/survival.

7. National Institute for Health and Care Excellence (NICE). Oesophago-gastric cancer: assessment and management in adults. NICE guideline [NG83], 2018. Available from: https://www.nice.org.uk/guidance/ng83.
8. Kudo T, Hamamoto Y, Kato K, Ura T, Kojima T, Tsushima T, et al. Nivolumab treatment for oesophageal squamous-cell carcinoma: an open-label, multicentre, phase 2 trial. Lancet Oncology. 2017;18(5):631-9.
9. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. Journal of Epidemiology and Community Health. 1998;52(6):377-84.
10. Muro K, Lordick F, Tsushima T, Pentheroudakis G, Baba E, Lu Z, et al. Pan-Asian adapted ESMO Clinical Practice Guidelines for the management of patients with metastatic oesophageal cancer: a JSMO-ESMO initiative endorsed by CSCO, KSMO, MOS, SSO and TOS. Annals of Oncology. 2019;30(1):34-43.
11. Lordick F, Mariette C, Haustermans K, Obermannová R, Arnold D. Oesophageal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of Oncology. 2016;27(Suppl 5):v50-v7.
12. Ono Pharmaceutical Co. Ltd. ONO-4538 Phase III Study. A multicenter, randomized, open-label study in patients with esophageal cancer refractory or intolerant to combination therapy with fluoropyrimidine- and platinum-based drugs - study report. 2019.
13. Moriwaki T, Kajiwara T, Matsumoto T, Suzuki H, Hiroshima Y, Matsuda K, et al. Survival analysis of platinum-refractory patients with advanced esophageal cancer treated with docetaxel or best supportive care alone: a retrospective study. Diseases of the Esophagus. 2014;27(8):737-43.
14. Nakatsumi H, Komatsu Y, Sawada K, Muranaka T, Kawamoto Y, Yuki S, et al. Retrospective comparison of efficacy and safety of docetaxel and weekly-paclitaxel as 2nd-line chemotherapy for patients with unresectable or recurrent esophageal cancer. Annals of Oncology. 2016;27:50.
15. Shirakawa T, Kato K, Nagashima K, Nishikawa A, Sawada R, Takahashi N, et al. A retrospective study of docetaxel or paclitaxel in patients with advanced or recurrent esophageal squamous cell carcinoma who previously received fluoropyrimidine- and platinum-based chemotherapy. Cancer Chemotherapy and Pharmacology. 2014;74(6):1207-15.
16. Janmaat VT, Bruno MJ, Polinder S, Lorenzen S, Lordick F, Peppelenbosch MP, et al. Cost-effectiveness of cetuximab for advanced esophageal squamous cell carcinoma. PLoS One. 2016;11(4):e0153943.
17. Bascoul-Mollevi C, Gourgou S, Galais MP, Raoul JL, Bouche O, Douillard JY, et al. Health-related quality of life results from the PRODIGE 5/ACCORD 17 randomised trial of FOLFOX versus fluorouracil-cisplatin regimen in oesophageal cancer. European Journal of Cancer. 2017;84:239-49.
18. Doherty MK, Leung Y, Su J, Naik H, Patel D, Eng L, et al. Health utility scores from EQ-5D and health-related quality of life in patients with esophageal cancer: a real-world cross-sectional study. Diseases of the Esophagus. 2018;31(12):doy058.
19. Dutton SJ, Ferry DR, Blazeby JM, Abbas H, Dahle-Smith A, Mansoor W, et al. Gefitinib for oesophageal cancer progressing after chemotherapy (COG): a phase 3, multicentre, double-blind, placebo-controlled randomised trial. Lancet Oncology. 2014;15(8):894-904.
20. Shenfine J, McNamee P, Steen N, Bond J, Griffin SM. A randomized controlled clinical trial of palliative therapies for patients with inoperable esophageal cancer. American Journal of Gastroenterology. 2009;104(7):1674-85.
21. Tian D, Wen H, Fu M. Comparative study of self-expanding metal stent and intraluminal radioactive stent for inoperable esophageal squamous cell carcinoma. World Journal of Surgical Oncology. 2016;14(1):18.

22. Xinopoulos D, Dimitroulopoulos D, Tsamakidis K, Korkolis D, Fotopoulou A, Bazinis A, et al. Palliative treatment of advanced esophageal cancer with metal-covered expandable stents. A cost-effectiveness and quality of life study. Journal of BUON. 2005;10(4):523-8.
23. Dallal HJ, Smith GD, Grieve DC, Ghosh S, Penman ID, Palmer KR. A randomized trial of thermal ablative therapy versus expandable metal stents in the palliative treatment of patients with esophageal carcinoma. Gastrointestinal Endoscopy. 2001;54(5):549-57.
24. Malik V, Johnston C, Donohoe C, Claxton Z, Lucey J, Ravi N, et al. (18)F-FDG PET-detected synchronous primary neoplasms in the staging of esophageal cancer: incidence, cost, and impact on management. Clinical Nuclear Medicine. 2012;37(12):1152-8.
25. O'Donnell CA, Fullarton GM, Watt E, Lennon K, Murray GD, Moss JG. Randomized clinical trial comparing self-expanding metallic stents with plastic endoprostheses in the palliation of oesophageal cancer. British Journal of Surgery. 2002;89(8):985-92.
26. O'Donnell CA, Gray J, Hodgson H, Macpherson M, Zammit M, Fullarton G. A cost comparison of photodynamic therapy and metallic stents in the palliation of oesophageal cancer. Photodiagnosis and photodynamic therapy. 2007;4(1):65-70.
27. Russell IT, Edwards RT, Gliddon AE, Ingledew DK, Russell D, Whitaker R, et al. Cancer of Oesophagus or Gastricus - New Assessment of Technology of Endosonography (COGNATE): report of pragmatic randomised trial. Health Technology Assessment. 2013;17(39):1-170.
28. Tralau-Stewart L, Roy R. Radiology-guided oesophageal stenting for the palliation of dysphagia: A single center experience. Annals of Oncology. 2019;30:aa23.
29. Woods B, Sideris E, Palmer S, Latimer N, Soares M. NICE DSU Technical Support Document 19: partitioned survival analysis for decision modelling in health care: a critical review, 2017. Sheffield, UK: Decision Support Unit, ScHARR. Available from: http://nicedsu.org.uk/wp-content/uploads/2017/06/Partitioned-Survival-Analysis-final-report.pdf.
30. Watanabe M. Recent topics and perspectives on esophageal cancer in Japan. JMA Journal. 2018;1(1):30-9.
31. National Institute for Health and Care Excellence (NICE). Technology Appraisal Guidance [TA428]. Pembrolizumab for treating PD-L1-positive non-small-cell lung cancer after chemotherapy. Final appraisal determination, 2016. Available from: https://www.nice.org.uk/guidance/ta428/documents/final-appraisal-determination-document.
32. National Institute for Health and Care Excellence (NICE). Technology appraisal guidance [TA525]. Atezolizumab for treating locally advanced or metastatic urothelial carcinoma after platinum-containing chemotherapy. Final appraisal determination, 2018. Available from: https://www.nice.org.uk/guidance/ta525/documents/final-appraisal-determination-document.
33. Davies A, Briggs A, Schneider J, Levy A, Ebeid O, Wagner S, et al. The ends justify the mean: outcome measures for estimating the value of new cancer therapies. Health Outcomes Research in Medicine. 2012;3(1):e25-e36.
34. Office for National Statistics. National life tables: UK, 2019. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesunitedkingdomreferencetables.
35. Ono Pharmaceutical Co. Ltd. ONO-4538 Phase III Study. A multicenter, randomized, open-label study in patients with esophageal cancer refractory or intolerant to combination therapy with fluoropyrimidine and platinum-based drugs. Protocol (ONO-4538-24). 2016.
36. Dolan P. Modeling valuations for EuroQol health states. Medical Care. 1997;35(11):1095-108.
37. Simons CL, Rivero-Arias O, Yu LM, Simon J. Multiple imputation to deal with missing EQ-5D-3L data: should we impute individual domains or the actual index? Quality of Life Research. 2015;24(4):805-15.
38. National Institute for Health and Care Excellence (NICE). Technology appraisal guidance [TA378]. Ramucirumab for treating advanced gastric cancer or gastro–oesophageal

junction adenocarcinoma previously treated with chemotherapy, 2016. Available from: https://www.nice.org.uk/guidance/TA378.
39. Ara R, Brazier JE. Using health state utility values from the general population to approximate baselines in decision analytic models when condition-specific data are not available. Value in Health. 2011;14(4):539-45.
40. National Institute for Health and Care Excellence (NICE). Technology appraisal guidance [TA519]. Pembrolizumab for treating locally advanced or metastatic urothelial carcinoma after platinum-containing chemotherapy, 2018. Available from: https://www.nice.org.uk/guidance/ta519.
41. National Institute for Health and Care Excellence (NICE). Technology appraisal guidance [TA519]. Pembrolizumab for treating locally advanced or metastatic urothelial carcinoma after platinumcontaining chemotherapy. Final appraisal determination, 2018. Available from: https://www.nice.org.uk/guidance/ta519/documents/final-appraisal-determination-document.
42. National Institute for Health and Care Excellence (NICE). Technology appraisal guidance [TA490]. Nivolumab for treating squamous cell carcinoma of the head and neck after platinum-based chemotherapy. Final appraisal determination, 2017. Available from: https://www.nice.org.uk/guidance/ta490/documents/final-appraisal-determination-document.
43. National Institute for Health and Care Excellence (NICE). Technology appraisal guidance [TA483]. Nivolumab for previously treated squamous non-small-cell lung cancer. Final appraisal determination, 2017. Available from: https://www.nice.org.uk/guidance/ta483/documents/final-appraisal-determination-document.
44. Grothey A, Van Cutsem E, Sobrero A, Siena S, Falcone A, Ychou M, et al. Regorafenib monotherapy for previously treated metastatic colorectal cancer (CORRECT): an international, multicentre, randomised, placebo-controlled, phase 3 trial. Lancet. 2013;381(9863):303-12.
45. British National Formulary (BNF). Docetaxel, 2020. Available from: https://bnf.nice.org.uk/medicinal-forms/docetaxel.html.
46. British National Formulary (BNF). Paclitaxel, 2020. Available from: https://bnf.nice.org.uk/medicinal-forms/paclitaxel.html.
47. Department of Health and Social Care. eMIT national database. Drugs and pharmaceutical electronic market information tool, 2020. Available from: https://www.gov.uk/government/publications/drugs-and-pharmaceutical-electronic-market-information-emit.
48. NHS Improvement. 2018/19 National Cost Collection data, 2020. Available from: https://improvement.nhs.uk/resources/national-cost-collection/#ncc1819.
49. Campbell H, Stokes E, Bargo D, Logan R, Mora A, Hodge R, et al. Costs and quality of life associated with acute upper gastrointestinal bleeding in the UK: cohort analysis of patients in a cluster randomised trial. BMJ Open. 2015;5(4):e007230.
50. White J, Carolan-Rees G. PleurX peritoneal catheter drainage system for vacuum-assisted drainage of treatment-resistant, recurrent malignant ascites. Applied Health Economics and Health Policy. 2012;10(5):299-308.
51. Georghiou T, Bardsley M. Exploring the cost of care at the end of life. London: Nuffield Trust Research Report. 2014.
52. Copley-Merriman C, Stevinson K, Liu FX, Wang J, Mauskopf J, Zimovetz EA, et al. Direct costs associated with adverse events of systemic therapies for advanced melanoma: Systematic literature review. Medicine (Baltimore). 2018;97(31):e11736.
53. Lewis G, Peake M, Aultman R, Gyldmark M, Morlotti L, Creeden J, et al. Cost-effectiveness of erlotinib versus docetaxel for second-line treatment of advanced non-small-cell lung cancer in the United Kingdom. Journal of International Medical Research. 2010;38(1):9-21.

54. Khan I, Morris S, Hackshaw A, Lee SM. Cost-effectiveness of first-line erlotinib in patients with advanced non-small-cell lung cancer unsuitable for chemotherapy. BMJ Open. 2015;5(7):e006733.
55. National Institute for Health and Care Excellence (NICE). Technology appraisal guidance [TA628]. Lorlatinib for previously treated ALK-positive advanced non-small-cell lung cancer, 2020. Available from: https://www.nice.org.uk/guidance/ta628.
56. Wehler E, Zhao Z, Pinar Bilir S, Munakata J, Barber B. Economic burden of toxicities associated with treating metastatic melanoma in eight countries. European Journal of Health Economics. 2017;18(1):49-58.
57. Briggs A. Probabilistic analysis of cost-effectiveness models: statistical representation of parameter uncertainty. Value in Health. 2005;8(1):1-2.
58. Rohatgi A. WebPlotDigitizer, version 4.2, 2019. Available from: https://automeris.io/WebPlotDigitizer.
59. Latimer N. NICE DSU Technical Support Document 14: survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data, 2013. Sheffield, UK: Decision Support Unit, ScHARR. Available from: http://nicedsu.org.uk/wp-content/uploads/2016/03/NICE-DSU-TSD-Survival-analysis.updated-March-2013.v2.pdf.

National Institute for Health and Care Excellence

Centre for Health Technology Evaluation

ERG report – factual accuracy check

Nivolumab for previously treated unresectable advanced oesophageal cancer [ID1249]

You are asked to check the ERG report to ensure there are no factual inaccuracies contained within it.

If you do identify any factual inaccuracies, you must inform NICE by the end of **9 July 2020** using the below comments table. All factual errors will be highlighted in a report and presented to the Appraisal Committee and will subsequently be published on the NICE website with the committee papers.

The factual accuracy check form should act as a method of detailing any inaccuracies found and how and why they should be corrected.

Issue 1 Clinical evidence issues

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Overall document including Section 1.2 page 10 and Section 3.2.4 page 23 The ERG comments on the generalisability of the data to the UK setting, but does not appear to reflect the SLR provided within the submission dossier, undertaken specifically to inform on this issue.	Inclusion of the SLR provided within the submission (Section 2.12.4.1.1 of Document B) when discussing the potential impact of different outcomes between Asian and Western populations.	An SLR evaluating differences in patient characteristics and survival outcomes between Asian and Western populations with treatment experienced advanced OSCC was undertaken. Results of this SLR (as presented in Section 2.12.4.1.1 of the CS) supported the assumption that OS between Asian and Western populations is comparable. Results indicated that OS was comparable between Asian and Western populations with OSCC (median: 7.5 versus 7.4 months); mean one-year OS was 21.1% in Asian and 27.9% in Western patients.	The ERG does not consider this to be a factual error. **No action required.**
Overall document *including Section 1.3 page 11* The ERG state: *"The ERG did not consider the combination of the company's base-case projections of OS for the nivolumab and taxanes arm to be the most appropriate estimates to inform the model, given the follow-up data available and the generalisability issues with ATTRACTION-3.[1]".* The ERG further state: "*The ERG considered any double counting of mortality to partially address the generalisability issues with ATTRACTION-3.*[1]".	The text should be amended to reflect that the ERG has taken a conservative approach to nivolumab OS and an optimistic approach to taxane OS as a result of generalisability concerns.	It is acknowledged that generalisability of ATTRACTION-3 data to the UK setting is uncertain. However, it is unclear how generalisability impacts on extrapolation of Kaplan-Meier data. It appears that the ERG considers that the generalisability should be reflected in more conservative OS estimates for nivolumab only (taxane OS is improved in the ERG base case analysis). If this is the case, the rationale should be provided for why generalisability would result in improving taxane OS versus reducing nivolumab OS.	The ERG agrees that its preferred approach taken to estimate OS for patients treated on the taxanes arm specifically ought not be considered "conservative", as it produces slightly higher survival estimates than the company's. However, the text highlighted by the company is not considered to contain any factual errors, and thus no edits have been made to the ERG's report. **No action required**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Overall document including Section 3.2.3 page 23 and Section 3.2.5.4 page 26 The ERG state: *"..the ERG noted some quality issues specifically relating to the open-label design of ATTRACTION-3,[1] where patients and investigators are not masked to treatment allocation. The ERG noted potential limitations with the open-label treatment of nivolumab in ATTRACTION-3,[1] particularly given that nivolumab, docetaxel and paclitaxel are all intravenously administered drugs. The ERG noted a substantial limitation with the open-label design in respect of internal validity, especially with regard to safety and HRQoL outcomes. Specifically, while for the objective measurement of the main clinical outcomes (PFS, response and OS), the risk of bias arising from lack of blinding is likely to be low, the ERG noted that in subjective measures of HRQoL and some safety data the risk of bias might be higher.".* The ERG further state: *"Clinical advisors to the ERG confirmed that improvements in HRQoL with nivolumab vs chemotherapy are clinically meaningful. However, the ERG noted that the lack of blinding inherent in the ATTRACTION-3[1] study design could bias subjective measures of QoL, potentially inflating the effect of nivolumab on HRQoL.".*	Text should be updated to reflect the rationale for an open-label trial design. Additionally, it is recommended that evidence relating to the impact of blinding on outcomes, particularly patient-report outcomes, is reflected in the report.	The company acknowledged that in certain circumstances an open-label study design means there is a possibility the knowledge of the treatment might have influenced patient responses. However, the intervention and the comparators of interest were clearly administered at different frequencies and are associated with different kinds of toxicities (as outlined in the CS). These clearly recognisable aspects of the interventions would render an attempt at blinding redundant. Furthermore, the primary endpoint of overall survival is an objective measure, which would remain unaffected by any potential bias resulting from an open-label study design. Additionally, involvement of an independent data monitoring committee for safety assessments ensured anonymity of the treatment groups during data review. Based on this, an open-label study design was considered more appropriate. Furthermore, a review of trials to study the impact of blinding on estimated treatment effects in RCTs by Moustgaard et al. (2019) did not find any evidence for a difference in estimated treatment effect between trials applying blinding and open-label study designs.[2] This included patient-reported outcomes. Hence, the impact of the open-label trial design on AEs and HRQoL is likely to be limited. This	The ERG acknowledged the rationale provided by the company for the open-label design; however, it stands by its interpretation of the risk of bias assessment in respect of this aspect and does not consider it to be a factual error. **No action required**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
		further supports the appropriateness of the open-label study design of ATTRACTION-3. It should be noted that similar outcomes have been observed across immunotherapy indications, so that the observed outcomes cannot be considered unexpected or potentially spurious.	
Overall document including Section 3.2.4 page 23 The ERG state: "*Furthermore, only patients with Eastern Cooperative Oncology Group (ECOG) performance status (PS) 0-1 are only included in the trial, which suggests that participants in ATTRACTION-3[1] are fitter and otherwise systematically different than those encountered in routine UK practice.".*	This text should be removed or amended to reflect the clinical advice provided to the ERG.	Within the cost-effectiveness section of the ERG report (Section 4.2.3 page 38), it is stated that: "*Clinical advice provided to the ERG suggested that patients would only be considered as candidates for systemic anticancer therapy (either with a taxane or nivolumab) if they had an ECOG PS of 0 or 1. In addition, clinical advice suggested that some patients may opt for treatment with nivolumab but would otherwise decline to receive a further line of chemotherapy (i.e. a taxane).*" Hence, the ERG statement is contradicted within the report, based on evidence provided by clinical experts.	The ERG accepts the company's suggestion. **Sentence deleted.**
Section 5.2.3.1 page 85 The ERG state: *"In consultation with the ERG, two practising oncologists suggested that the majority of patients undergoing second line taxane therapy would have an estimated survival of less than 12 months. However, as seen in ATTRACTION-3,[1] 12-month OS for the taxanes arm was approximately 34%, indicating that patients in the*	This statement should be removed.	While the Company do not dispute that the "*majority of patients undergoing second line taxane therapy would have an estimated survival of less than 12 months*", this is reflected in ATTRACTION-3, where both arms are associated with median OS less than 12 months (8.38 months in the taxane arm and 10.91 months in the nivolumab arm). Further, within the	The ERG follows the advice of its independent clinical experts and does not consider this to be a factual error. **No action required.**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
study were likely fitter than those that would be eligible for taxanes in NHS practice. This means that inferences concerning the most plausible extrapolations for each treatment arm should be considered with this potential discrepancy in patient population in mind.".		taxane arm, almost two-thirds (66%) of patients had died at twelve months. As such, it is unclear how the data from ATTRACTION-3 does not conform to clinical expert opinion that the majority of patients would have life expectancy less than 12 months.	

Issue 2 Survival analysis

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Overall survival model proposed by ERG Whole document, specifically the ERG base case analysis	Amendment of text to reflect the overall impact of the proposed ERG base case	Relevant survival modelling guidelines indicate that survival extrapolations should reflect the disease pathway and plausible biological explanation for treatment effect. This includes reviewing the overall impact across model inputs, rather than reviewing models in isolation. This is specifically of note in the ERG base case, where a case is made for each input in isolation, but the overall impact is to predict clinically implausible outcomes. As noted by the ERG, the ERG base case applies more conservative extrapolations for nivolumab OS due to the concerns around the generalisability of the evidence. However, taxane OS is assumed to be more optimistic, to the extent that ATTRACTION-3 extrapolations predict	The ERG does not agree that it's base-case analysis includes the estimation of "clinically implausible outcomes". Moreover, the ERG's rationale for its preferred choice of models is based on several factors, including (but not limited to) generalisability concerns with ATTRACTION-3. The ERG's base-case analysis estimates a total of [redacted] LYs gained for nivolumab, versus [redacted] LYs gained for taxanes (ERG report Table 26). The company's statement that extrapolations made are more optimistic for taxanes versus nivolumab is therefore incorrect. The ERG accepts that the choice of models for each outcome (i.e. OS, PFS, and ToT) is subject to debate, but

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
		longer mean OS for taxanes than for nivolumab, which can be considered implausible in the context of the observed data. Further, despite shorter OS assumptions, it is assumed that time on treatment is increased. As the Summary of Product Characteristics (SmPC) for nivolumab specifies that treatment should be administered for as long as there is clinical benefit, derivation of time on treatment should be considered in the context of both PFS and OS. Although it is likely that some patients will receive treatment beyond progression, it is not plausible that there will be extended post-progression treatment period in the absence of clear benefits such as improved quality of life. Further, it is implausible that extended post-progression treatment would be seen in the absence of clinical improvement, which would be reflected in OS. In this context, it is not appropriate to extend time on treatment, reduce OS (for nivolumab only) and assume no post-progression utility differential between treatment arms. Hence, survival modelling in this indication should be considered across model inputs, rather than in isolation, otherwise the results are clinically implausible. This should be noted as a direct consequence of the ERG's proposed base case analysis.	does not consider any specific amends to be required to the ERG's report based on the information provided by the company here. **No action required.**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
OS Generalised Gamma for both arms Section 6.3.1, page 98 Specification of SP generalised gamma for OS (cut point at 5.75 months) for both treatment arms. The Company have several issues with the accuracy of this choice. The ERG justify this for the following reason: *"All patients are expected to have died by 10 years. 10-year OS in the company's base-case was estimated to be 1.92% for the nivolumab Arm, versus 0.20% for the ERG's base-case analysis." Though later go on to state that: "Long-term survival with checkpoint inhibitors in an OSCC population has not been established,"* Additionally, the Company do not believe that the choice of curve is appropriate as the outcomes (mean, median, range) are contradictory with clinical evidence. **Section 4.2.6.1, pages 50-51** The ERG report also describes criticisms of the CS OS taxane curve and their subsequent preferred choice with apparent irregularity. *"The ERG does not consider the taxanes OS curve to be affected by the*	Amendment to reflect additional details around the limitations of the ERG proposed approach versus the company base case analysis	The assumption that patients are expected to have died at 10 years is not considered as justifiable by the company; by the ERG's own admission: "Long-term survival with checkpoint inhibitors in an OSCC population has not been established." Further, this assumption is justified based on the clinician opinion that the majority of patients will have died within 12 months. As this is in agreement with data observed in ATTRACTION-3, it is not possible to use this rationale to distinguish between extrapolations. Additionally, while the reasoning for the ERG's alternative curve selection is not entirely disputed, the specific choice of generalised gamma curves for both arms infer that outcomes for patients are better with taxanes than with nivolumab. The median and mean estimates of survival from the generalised gamma OS model for nivolumab are 8.29 and 11.68 months. While for taxanes these are 8.38 and 12.22 months, respectively. Selection of these curves would suggest that patients will have improved outcomes of survival with taxanes compared with nivolumab, which seems improbable based on observed data. Additionally, the clinical evidence provided in Section B.2 of the CS disproves this. These reasons informed the originally selected base case, as it is important	The assumption that all patients (treated with nivolumab or taxanes) are expected to have died by 10 years was based on clinical expert opinion provided to the ERG. The ERG is unclear how the numbers estimated by the company were produced, and therefore cannot comment on these further. The ERG report states that the taxanes OS curve does not appear to be affected by the same 'elbow' in the curve seen for the nivolumab arm. The ERG did however ultimately select an SP generalised gamma model owing to the fact that this approach provided a good fit to both arms and the ERG considered its extrapolations to be reasonable. The company highlights that due to the similarity between the exponential and generalised gamma models estimations, the generalised gamma model may be considered an over-complication, and that the exponential model may be a reasonable fit. However, this does not mean that the models are identical, and the generalised gamma model is therefore still considered a reasonable choice. As noted above, the ERG accepts model selection may be the subject of debate, but does not consider its

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
same 'elbow', [referring to the "elbow" in data seen in the nivolumab OS curve] and so the specification of an FP approach for the taxanes arm in particular would not seem unreasonable". Later in Section 4.2.6.2, page 53, the report states "For the taxanes arm, the fitted exponential model did not appear to provide a good visual fit to the Kaplan–Meier curve . . . a constant hazard rate is likely too simplistic in order to fully reflect the pattern of survival seen for taxane-treated patients". **Section 6.3.1, page 98** The ERG state: *"owing to its increased flexibility, the generalised gamma may better reflect non-constant hazards.".*		that extrapolatory estimates validate clinical evidence. Finally, if the model selection incorrectly infers that the comparator is more effective than the intervention, the ICER will inevitably increase artificially. The ERG do not accept that there is an "elbow present" and cannot reject an FP model. It is further stated that the ERG do not believe there to be a constant hazard (thus making the exponential model inappropriate). Further, the ERG report intimates that a FP could accurately represent the data, but the ERG preferred base case places the cut point further from the start of the trial than that of the CS. The CS choice therefore includes more data to inform the extrapolation than that of the ERG, with 26 events in the KM period of the CS base case model and 79 in the ERG's. The description of the problem and the solution presented by the ERG seem to be at odds. Importantly, the curve chosen by the ERG to best represent the taxane OS was generalised gamma, citing its flexibility as one of the reasons. However, where cut points were placed at 5.75 months, the generalised gamma curve and the exponential curve are almost indistinguishable from each other on the plots as they are almost completely overlaid. This would suggest that while the generalised	selection to be a factual error, nor does it consider its reasons given to be in error. **No action required.**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
		gamma model has increased flexibility and the ability to fit better to the data presented than an exponential, it is in fact taking on the exact shape of the exponential model. Therefore, there seems to be no particular reason to disregard that the exponential model fits just as well or to move the cut point further from the initially proposed base case.	
ToT SP Weibull for both arms **Section 6.3.2, page 100** The ERG state that their preferred model to represent ToT for both nivolumab and taxanes is a SP Weibull model (cut point at 5.75 months) as it *"was considered to provide a more realistic pattern of longer-term discontinuation versus the generalised gamma model*".	The base case proposed in the CS is considered to be appropriate.	The Company would like to highlight that by moving the cut point to 5.75 months, the ERG's preferred models extrapolatory period is informed by just 25 patients, making it highly uncertain, if not inappropriate. The corresponding nivolumab arm is informed by 55 patients. As noted in the ERG report, *"selection of a suitable point on the* [Kaplan–Meier] *function from which to extrapolate becomes increasingly arbitrary as the effective sample size decreases*".[33] This is in contrast to the suggested models in the CS, which use information about discontinuation from all 209 and 210 patients, respectively. Therefore, the company is concerned that it is not correct to stipulate that the model provides a "more realistic pattern of longer-term discontinuation" where this informed by only 25 patients.	The ERG accepts that the extrapolations are uncertain, but disagrees that they are any less "realistic" because they are based on a smaller number of patients. **No action required.**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Section 6.2.3 page 93 The ERG report states: *"It can also be seen from Figure 9 that the ERG's base-case analysis produces similar estimates to the '1% at five years' scenario, and so the ERG's selected model may be expected to provide a conservative estimate of the ICER (i.e. estimates a relatively large proportion of patients to continue treatment after the end of follow-up in ATTRACTION-3[1] versus the other projections shown)."* However, this does not reflect that the hazard is decreasing faster in the ERG proposed base case analysis, so that the '1% at five years" scenario is actually more optimistic than the ERG proposed base case when extrapolated beyond five years	The text and/or figure should be amended to reflect the implications of the ongoing extrapolations	The current text/figure does not reflect that the hazard is decreasing faster in the ERG proposed base case analysis, so that the '1% at five years" scenario is actually more optimistic than the ERG proposed base case when extrapolated beyond five years	The ERG accepts the company's suggestion. **Sentence edited.** Added sentence to clarify the impact of different hazards between ERG preferred base-case and this exploratory analysis on the implication of choice of ToT model on the ICER.
Section 5.3 page 88 The ERG quote NICE DSU TSD 14 guidance*: "Where parametric models are fitted separately to individual treatment arms it is sensible to use the same 'type' of model, that is if a Weibull model is fitted to one treatment arm a Weibull should also be fitted to the other treatment arm"*.[3] .	This text should be removed or amended to reflect that it may be inappropriate to inform therapies with different mechanisms of actions and hence, significantly different hazard profiles.	It is considered appropriate where models are fitted separately to individual treatment arms to use the same model "type" where the hazards presenting are the same and where the data directs. The Company do not consider that these conditions are satisfied. Appendix M (Figure 48) shows the cumulative hazard profiles from ATTRACTION -3 are quite different. This is largely driven by very different mechanisms of action, which would be expected to have different	The next sentence in the ERG's report acknowledges the point raised by the company: *"While this principle is subject to debate within the context of two treatments with very different mechanistic properties, …"* The ERG notes however that this is published guidance which is being referenced, and caveated where deemed appropriate. **No action required.**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
		hazard profiles that could not be reflected by the same "type" of model. This is also apparent in the responses to clarification question B10 on pages 54-55, figures 46-49 where the SP Weibull (cut point at 5.75 months) for nivolumab can be seen to deviate greatly from the observed data, both over and underestimating. In contrast, the same model fit to the taxane arm adheres much more closely throughout. Bagust and Beale (2014)[4] describe recommendations for pragmatic modelling that were used to guide analysis as reported in the CS Section 3.3.2.1.1 page 96. This report recommends that "*The presumption should be against joint modeling of treatment arms unless modeling the trial arms independently reveals that functional forms and parameter estimates are closely aligned. Nonetheless, the appropriateness of each separate functional form needs careful justification, from both the available data and other sources (such as clinical experience and, if available, patient registries).".* These recommendations were adhered to in the Company base case approach and seem to be at odds with the technique used in the ERG base case selection.	

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Section 4.2.5 page 45 and Section 6.3.1 page 98 The ERG state: *"As a part of the development of the ERG's report, two practising oncologists independently confirmed that the majority of patients undergoing second-line taxane therapy in current UK practice would have an estimated survival of less than 12 months. As such, it may be considered reasonable that by 10 years (that is, ten-times the maximum life expectancy for most patients with current care), the majority of relevant costs and effects would be captured by the model.".* *"All patients are expected to have died by 10 years. 10-year OS in the company's base-case was estimated to be 1.92% for the nivolumab arm, versus 0.20% for the ERG's base-case analysis."*	Text should be amended to reflect the uncertainty around using the median to describe survival at ten years, particularly in patients receiving immunotherapy.	While the Company do not dispute that the "*majority of patients with current care in the UK have an estimated survival of <12 months*", this is accurately reflected in ATTRACTION-3, where standard of care is associated with median OS less than 12 months. However, although the majority of patients (i.e. >50%) will have died by 12 months, this does not inform on the exact percentage that will survive to ten years, particularly in the case of immunotherapies where no evidence exists. Further, it should be noted that both the ERG and company preferred base case predict surviving patients at 10 years for the nivolumab, although the ERG preferred base case predicts 0.2% while the company preferred base case predicts 1.92%. However, it should be noted that the company model predicts 0% patients surviving to ten years in the taxane arm, while the ERG model predicts 0.01% patients surviving, which may be implausible.	The ERG understands that the term "majority" may be misleading within this context, as the ERG's intended meaning was to state that nearly all patients undergoing second-line taxane therapy in current UK practice would have an estimated survival of less than 12 months. With the text in Section 4.2.5 edited to state "nearly all", the ERG's report will be clearer, and the Company's concern is no longer applicable. **Sentence edited.** Changed 'majority' to 'nearly all'.
Section 6.3.1 page 98 The ERG state: *"Accordingly, the ERG considered a more conservative estimate of OS to be more suitable to inform its preferred base-case analysis.".*	Suggest a text change to read " . . a more conservative estimate of nivolumab OS . . ." or removal of this statement.	The Company does not believe this to be accurate reporting as this refers only to the ERG's choice of OS curve for the nivolumab arm. Indeed, the taxane arm OS mean increased from 11.96 to 12.22 demonstrating a non-conservative approach.	The ERG accepts the company's suggestion. **Sentence edited.** Clarified that OS refers to nivolumab

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Section 6.3.1 page 98 The ERG stated: "*A generalised gamma parameterisation has also been used to model OS for checkpoint inhibitors and taxanes in previously-published NICE appraisals (e.g. TA525[5]), and in the case of this appraisal yielded AIC scores within two points of the company's base-case models (suggesting similar statistical goodness-of-fit).*"	Suggest correcting this statement such that it reflects the decisions made in TA525.	TA525 is for a different indication at a different line and with quite different prognosis and assumed pattern of progression than the indication considered in this submission. Additionally, in TA525[5] the company submitted with a generalised gamma parameterised mixture cure model for OS and the ERG disagreed that this was appropriate. Indeed, it is further reported in the final decision that, " *It concluded that modelling overall survival using Kaplan–Meier curves with the tails extrapolated with a log-logistic distribution (the ERG's approach) was more appropriate than the company's approach, because it produced more plausible estimates for the taxanes*". Importantly, as fewer data points are included (use of KM data as in this model), the AIC will be expected to decrease and so this would not necessarily be "suggesting similar statistical goodness-of-fit" as the ERG state.	The Company is correct that ultimately the KM + log-logistic was used in the committee's final decision making in TA525. However, the ERG notes that TA525 serves as an example where this specific modelling approach has been used previously (not necessarily to inform the committee's preferred settings). However, for clarity, the ERG considers it important to note that the TA525 committee ultimately preferred the KM + log-logistic model. **Sentence edited.** Clarified 'not used in committee's preferred base-case'. For the second point, this is correct. For simplicity, the ERG has opted to remove this statement. **Part of sentence removed:** ", and in the case of this appraisal yielded AIC scores within two points of the company's base-case models (suggesting similar statistical goodness-of-fit)".

Issue 3 Utility analysis

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Change in utility over time	It is feasible that the marginal state mean utility may increase through time	The ERG statement is based upon the principle that the mean utility in state is	The Company's comment is theoretically correct – that the utility of

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Section 4.2.7.1., page 65 The ERG state: *"The ERG considered it likely that 'true' utility values at later time points are systematically lower than those seen in earlier time points (based on the general principle that utility declines over time, both related to disease progression and natural health decline as patients age).".*	due to selective effects on the population acting to reduce the proportion of lower-utility patients in state, in contrast to the general patient-level trend.	based upon the mean experience of a single patient through all time, and not accommodating for the selective removal of patients from state (due to progression or death) that tends to increase mean utility when observed over the remaining patients.	the group could increase over time as older, frailer patients die. However, this does not render the statement made in the ERG's report incorrect, as this statement refers to the nature of missingness. **No action required.**
Impact of time-to-death on utility Section 4.2.7.1., page 65 The ERG state*: "Here, it should also be noted that it was assumed that time-to-death would not be considered to have an impact on utility for patients still alive after 18 months.".*	This statement and the remainder of the paragraph should be deleted	The CS clearly describes the chained imputation model as imputing times to death where unobserved, and using these times and observed times to impute utility, with a time until death of greater than 18 months having no impact on utility conditional upon other imputation variables. The ERG's statement that time to death is not considered to have an impact on utility for patients still alive after 18 months (after study initiation) is incorrect.	The ERG accepts the company's correction concerning the 18 months' time point. However, the ERG does not consider it necessary to delete the entire paragraph, so has edited this instead. **Paragraph edited.** Revised text for greater clarity around the impact of time-to-death on utility.
Data imputation Section 4.2.7.4, page 70 The ERG state: *"However, such an analysis avoided the need to rely on data imputation, which (as described in Section 4.2.7.1) the ERG did not consider to have been conducted appropriately, nor did it consider it possible to impute these data appropriately with current-available*	This statement should be appended with a description of the structural assumptions of a linear mixed-effects model, and the implications for the missing data for each health state.	The fitting of a linear mixed effects model assumes that all subjects would have valid observations at all times. If the imputation model were to behave similarly, it would assume that all patients remained in state (i.e. without progression or death) at all observed time points. This is clearly inappropriate, but the assumption is preferred by the ERG for no adequate reason.	Not a factual error. **No action required.**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
methods.".			
Mean utility in PD patients Section 6.3.5., page 102 The ERG state: *"In addition, the ERG did not consider it appropriate to specify utility values that exhibit a large difference in utility for PD patients dependent on initial treatment assignment. The ERG noted that differences in utility by treatment arm after progression may be due to a combination of potential continued benefit from nivolumab after progression, or as a direct consequence of the open-label design of ATTRACTION-3.2 In the ERG's preferred base-case analysis, the average of the PD utility values per arm is assumed to apply for both treatment arms.".*	The ERG must append an explanation of how mean utility values in PD are expected to be consistent between populations with differing average times to death if they accept that a) proximity to death has an impact upon utility, clearly visible in the collected trial data, independent of progression status and b) patients having received nivolumab experience greater post-progression survival than patients having received taxanes.	The ERG's position is logically inconsistent with the evidence provided. They have not provided a refutation to either an extension of survival in post-progression for nivolumab (present in the ERG's base case) or of the profile of decreasing utility with proximity to death (within 18 months of death). Therefore, to assume that utility is equal in the PD states is to assume that nivolumab-receiving patients are in some way *disadvantaged* in utility entering PD, as they, in the mean, have a greater time until death.	The ERG does not agree that its position is "logically inconsistent" with the evidence provided. Benefit in the PD state in terms of QALYs is based on LYs gained in this state, as well as the utility associated with this state. The ERG does not consider the proposed amendment to constitute a factual error, and therefore no change has been made to the ERG's report. However, the ERG accepts that utility valuation is a complex issue, and so further discussion at technical engagement may be useful for the committee. **No action required.**

Issue 4 Cost inputs for cost-effectiveness modelling

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Radiotherapy cost for BSC Section 4.2.8.3, Page 75 The ERG state: "*Radiotherapy is costed as £184.25; however, the ERG*	Table 65 in the CS provides the schedule of administration to be used for each BSC component, where it states radiotherapy is to be administered twice weekly, for 7.5	Using the incorrect number of radiotherapy administrations leads to an overestimate of the ICER as presented in the ERG report.	The ERG thanks the company for clarifying the cost used here. However, for clarity, the ERG's base-case did not include changes made to this cost, and so no edit to the ERG's ICER is

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
calculated this as £92.13 based on the information provided in CS Table 64 (£487.45 [cost of treatment] x 0.189 [proportion of patients requiring treatment]).".	weeks (total 15 visits). This yields a weekly cost of £487.45 [cost of treatment] x 0.189 [proportion of patients requiring treatment] x 2 [number of admission per week] = £184.25.		required. However, the text has been amended to remove this discussion in the ERG's report. **Sentences removed.** Deleted irrelevant text in the costs section
Medication costs to control GI bleeds for BSC Section 4.2.8.3 Page 76 The ERG state: "*Medication costs to control GI bleeds were sourced from Campbell et al. (2015) and provided in CS Table 65 as a component of BSC. The costs from Campbell et al. (2015) were derived from data collected in 2012-13, which were stated as £23.76. The ERG assumed that an inflation factor was applied to obtain the cost of £25.71 presented in the CS (Document B, Table 65) but this is not clearly stated within the CS.".*	The inflation factor used to estimate the current cost of £25.71 was derived using the PSSRU Unit Costs of Health & Social Care indexes.[6] Costs were reported in the literature from 2012-13 and so were inflated to 2018/19 costs using an inflation factor of 1.082 [310.9/287.3].	Clarification of methodology used to inflate BSC component cost to current values.	While the ERG thanks the Company for clarifying the methodology used, the text within the ERG's report is not a factual error based on the CS and is therefore left unchanged. **No action required.**
Radiotherapy cost for BSC Section 4.2.8.3, Page 76 *The ERG state: "The cost for ascites drainage is stated as £3,404.20 in the CS (Document B, Table 65), based on a value of £3,146 obtained from White & Carolan-Rees (2012) inflated to 2015/2016 costs. After reviewing the cited manuscript, it was not clear to the ERG how the value of £3,146 was*	The proposed cost of £3,146 is sourced from Section 5.5 of NICE Medical technologies guidance *PleurX peritoneal catheter drainage for vacuum-assisted drainage of treatment-resistant, recurrent malignant ascites*.[7] The inflation factor used to estimate the current cost of £3,404.20 was derived using the PSSRU Unit Costs of Health & Social Care indexes.[6] Costs were	Clarification of source of BSC component as well as methodology used for inflation to current values.	While the ERG thanks the Company for clarifying the methodology used, the text within the ERG's report is not a factual error based on the CS and is therefore left unchanged. **No action required.**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
obtained for ascites drainage. The CS did not provide any breakdown of how this cost was calculated, nor the method used for applying an inflation factor to this cost."	reported in the literature from 2012-13 and so were inflated to 2018/19 costs using an inflation factor of 1.082 [310.9.287.3].		
Clinician consultation cost Section 4.2.8.3, Page 77 The ERG state: "*The cost of clinical consultation was cited as a weighted average of consultant-led and non-consultant led consultations from the National Cost Collection for the NHS 2018/19. The ERG calculated this as £196.33, in contrast to the value of £187.36 provided in CS Table 70. The ERG suspected the value of £187.36 was provided based on outpatient code 370 (Medical Oncology), and not the average across all HRG codes as stated in the CS. However, it remained unclear to the ERG which cost the company considered most appropriate to inform the model.".*	Table 70 in the CS states the source of costs were derived as a weighted average of costs for consultant led and non-consultant led, using codes WF01A, WF01B, WF01C, WF01D, WF02A, WF02B, WF02C and WF02D.	Given the nature of the indication, using a weighted average of costs for a medical oncologist is deemed appropriate to inform this component cost. Costs sourced from total HRGs is deemed inappropriate and could lead to inaccurate and uninformative costs being applied.	The ERG is still unclear which cost is preferred by the Company. The CS uses a cost of £187.36 which matches outpatient code 370 (Medical Oncology), whereas the Company's response here advocates an alternative cost which the ERG calculated to be £196.33. Further information is required from the Company to determine which cost is preferred and how this was calculated. **No action required.**
Hospitalisation cost Section 4.2.8.3, Page 77 The ERG state: "*The costs for hospitalisation included in the CS (Document B, Table 70) are based on a weighted average of elective and non-elective long-stay hospitalisation (Malignant Gastrointestinal Tract*	Each individual cost from the NHS Cost Collection was divided by the length of stay (using the same codes), sourced from NHS Reference Costs 2017/18.[8] The weighted average of these daily costs was used to calculate the cost of £534.07, as provided in the CS.	Costs were standardised using the length of stay in order to gain a fairer reflection of resource use. Weighting these standardised costs allowed the model to reflect a truer weekly cost for patients requiring hospitalisation.	Not enough information has been provided for the ERG to accurately edit the hospitalisation costs in our base-case. However, the Company's response suggests that the model applies hospitalisation cost as a daily cost, (though this is not explained in the CS). The ERG does not consider it

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Disorders, weighted average of elective and non-elective long-stay FD11A-FD11K). This was costed as £534.07 per hospitalisation; however, the unit costs of hospitalisation included in the weighted average ranged from £1,907 (FD11K, elective) to £9,650 (FD11A, elective). It was therefore unclear to the ERG how a weighted average of £534.07 was estimated. The ERG was able to estimate a different value of £577.11 using the same codes in a non-elective short stay setting, but was unable to calculate a value of £534.07. The ERG calculated the weighted average for hospitalisation costs as £3,379.73, based on the description provided in the CS.".			appropriate to apply a hospitalisation cost based on a length of stay equivalent to 1 day, especially given long-stay cost sources have been cited. Further discussion is required to determine the most appropriate cost to inform the model. However, no edit is required for the ERG's report. **No action required.**
Nerve block cost source Section 4.2.8 page 76 The ERG state: *"The source of costs for nerve blocks pain relief is not specified in Table CS 66, despite the reader being directed to this table to obtain further details of the medication(s) used and cost breakdown.".*	This section should refer to Table 64 in the CS. Based on clinician consultation through the market research survey it was obtained that 10.88% of patients receiving pain relief medication receive nerve blocks. Therefore, 10.88% of 0.459 of the patient population receiving pain medication results in 0.05 of patients receiving nerve blocks.	Table 64 provides the proportion of patients receiving nerve blocks, out of patients receiving pain relief medication. A patient population of 0.005 would underestimate the proportion of patients receiving nerve blocks.	This explains the proposed frequency of administration for nerve blocks but does not address the issue highlighted by the ERG that details of the specific medications and their cost breakdown is not provided. **No action required.**
Nerve block cost calculation Section 4.2.8 page 76 The ERG state*: "Additionally, nerve blocks pain relief was costed as £26.62*	The following calculation was used to derive the cost for nerve blocks (10.88% * 0.459) * £532.96=£26,62 Thus, the costs applied to nerve blocks received by patients in this indication is	The following calculation was used to derive the cost for nerve blocks (10.88% * 0.459) * £532.96=£26,62 The calculation by the ERG would result in an underestimation of the	As explained in the ERG comment, the proportion of patients receiving nerve blocks for the BSC scenario analysis was reported as 0.005 in CS Table 64. The calculation provided in the

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
for the BSC scenario analysis, whereas the ERG calculated this as £2.66 (£532.96 [cost of treatment] x 0.005 [proportion of patients requiring treatment]) based on the information in CS Table 64. The ERG suspected this may be due to a typographical error concerning the number of zeros in the proportion of patients requiring nerve blocks (e.g. 0.005 should perhaps be 0.05), but this is purely speculation.".	correctly reported.	costs applying to nerve blocks in this indication.	company response (10.88% * 0.459) results in the proportion of patients receiving treatment to be 0.05. Consequently, this represents a typographical error in CS Table 64, as suggested in the ERG comment. The BSC comparator is not included in the ERG model and therefore does not require any revision. **No action required.**
Cost source Section 4.2.1., page 35 The ERG state: "*Costs for docetaxel and paclitaxel are not reflective of the average price paid by NHS trusts (obtained through eMIT*)."	The text should be amended to reflect potential uncertainties around eMIT costs, in line with the reference case.	The NICE reference case states that analyses based on price reductions for the NHS will only be considered when the reduced prices are transparent and consistently available across the NHS, and if the period for which the specified price is available is guaranteed. Costs available the electronic Market Information Tool (eMIT) capture volume-based discounts provided to the NHS. It is unclear if all NHS trusts will have access to these medicines at this price, particularly due to the large standard deviation. Further, there is no confirmation these prices will be available for a guaranteed period.	The ERG accepts that eMIT costs are not necessarily "fixed" in nature, but do nevertheless represent the average price paid. The ERG notes that eMIT costs have been used to inform a number of previous appraisals, and does not consider their use to inform this appraisal to be inappropriate. Nevertheless, minimum BNF prices are more aligned with the eMIT costs versus the MIMS costs used in the company's base-case analysis. For example, docetaxel is costed at £720.10 in the company's base-case, £20.96 based on eMIT, or £101.25 based on BNF (docetaxel 160mg/8ml concentrate for solution for infusion vials [Seacross Pharmaceuticals Ltd]). **No action required.**
Patients opting for docetaxel or	Based on a market share assessment	Assuming that the majority of patients	The ERG accepts this is uncertain, but

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
paclitaxel Section 4.2.4 page 44 The ERG state*: "It is expected that the majority of patients in the UK would opt for treatment with docetaxel instead of paclitaxel, and so any differences in the modelled costs for each of these taxanes may be important to consider."*	conducted by the company (including 30 oncologists), 33% of patients receive docetaxel and 31% are treated with paclitaxel (with the remaining patients being treated with BSC, irinotecan, as part of a clinical trial, other active treatment or no treatment at all). Although there appears to be a slight preference for docetaxel, the company disagrees that this translates to the assumption that the majority of patients are treated with docetaxel.	in the UK would opt for treatment with docetaxel instead of paclitaxel contradicts the market share assessed by the company.	the ERG's report states that the 'true' value lies between 50-100%. In this case, this would be approximately 52% (i.e. 33% / 33+31%), which falls within the 50-100% range, and is therefore not an error. **No action required.**
Section 6.2.4 page 94 The ERG state: "*Docetaxel is expected to be the most commonly-used taxane in UK practice, yet the ERG does not consider it plausible that exactly 50% or 100% of patients receive docetaxel, and that the true value lies somewhere between these bounds.".*	Text should be amended to reflect the limitations of modelling docetaxel costs with clinical outcomes primarily derived from paclitaxel	As noted in the ERG report, docetaxel is associated with a poorer safety profile than paclitaxel. Hence, modelling docetaxel costs with clinical outcomes primarily derived from paclitaxel (particularly safety and HRQoL) may be considered extremely conservative and potentially inappropriate. Further, there is the potential for small differences in efficacy profiles, driven by differences in safety profile, which may not be fully reflected in the overall taxane arm	Not a factual error. **No action required.**

Issue 5 Early deaths

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Section 3.2.5.6 page 28 and Section	The text should be deleted	Question A13 states: *"A higher*	The ERG does not consider this to be a

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
1.2 page 11 The following text is misleading: *The ERG asked the company for clarification regarding why the early death rate was so much higher on nivolumab than control (clarification question A13). The company responded (clarification response A13) that this may relate to differences in the mechanism of action of immunotherapy treatments such as nivolumab compared to chemotherapy agents. Potentially of relevance, according to the company, are a longer time to response in immunotherapies compared to chemotherapy agents, and the indirect anti-tumour mechanism associated with immunotherapies, which may result in initial growth of existing lesions or formation of new lesions, prior to potential tumour shrinkage or eradication.* *The company further stated: "…the company commented that this was potentially related to the mechanism of action of immunotherapies versus chemotherapy agents".*		*proportion of patients died within the first 2.5 months in the nivolumab arm (32/210, 15.2%) as compared to the chemotherapy arm (15/209, 7.2%). Please can you provide an explanation for this finding, e.g. do you think it may be related to the mechanism of action of nivolumab?"* This question is not relevant to the endpoint in question '([redacted])'. Conflating the clarification response with the issue of "on treatment deaths" is misleading and inaccurate.	factual error. **No action required.**
Several paragraphs including Section 3.2.5.6 page 28 and Section 1.2 page 10-11 and Section 3.6 page 31 Additional context is required to	The text should be amended to reflect the context of longer time on treatment for nivolumab and the expanded definition.	It should be noted primarily that time on treatment was longer in the nivolumab arm than the taxane arm, in terms of mean ([redacted]) and in proportion of patients receiving long-term treatment,	The ERG does not consider this to be a factual error. **No action required.**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
describe “early deaths”, including the following text: *The ERG noted that early deaths, defined as ‘[redacted] (CS, p.72) were notably higher on nivolumab than control ([redacted] vs [redacted]).* The ERG further state: “*However, the ERG was concerned that early deaths were [redacted] on nivolumab than taxanes ([redacted] vs [redacted] – and the company commented that this was potentially related to the mechanism of action of immunotherapies versus chemotherapy agents. …the ERG was concerned about the fact that early deaths were around [redacted] in the nivolumab arm as the taxane arm.”.* Both statements are misleading		with [redacted] receiving >6 months of treatment (versus [redacted] for taxanes) and [redacted] receiving treatment >12 months (versus [redacted] for taxanes). In combination with the short survival observed for this population, long time on treatment results in additional “on treatment” deaths (defined as ‘[redacted]’). For this reason, the nivolumab arm experienced [redacted] disease-related deaths meeting this definition ([redacted] vs [redacted]). However, when this definition is expanded to ‘[redacted]’, outcomes are slightly lower in the nivolumab arm, both for overall deaths ([redacted] and for disease-related deaths ([redacted]).	
Several paragraphs including Section 3.2.5.6 page 28 and Section 1.2 page 10-11 and Section 3.6 page 31 *The ERG noted that early deaths, defined as ‘[redacted]*	“Early deaths” should be labelled to be more accurate. Suggestions may include “on treatment deaths”.	During ATTRACTION-3, the mean duration of treatment in the nivolumab arm was [redacted], [redacted] received >6 months of treatment and [redacted] received treatment >12 months. As “early deaths” is defined by the ERG as deaths occurring [redacted]	The ERG does not consider this to be a factual error. **No action required.**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
[redacted] [redacted] [redacted]' *(CS, p.72)* The ERG uses the terminology "early death", which was not defined during ATTRACTION-3 or in the company submission. This terminology is inaccurate and misleading, particularly since "early deaths" could occur up to 28 days following last dose.		[redacted]), these deaths could have occurred over 12 months after treatment initiation, which could not be considered early in the context of OSCC survival. Further, the term "early death" implies that these patients are dying earlier than otherwise typical. However, in the majority of patients with an "early death", the cause of death was initial disease (i.e. OSCC). As such, these deaths cannot be considered "early", particularly in the nivolumab arm, where significantly fewer deaths occurred than in the taxane. Finally, labelling these deaths as "early deaths" may have contributed to the ERGs misunderstanding and conflating these events with an increased frequency of deaths in the initial 2.5 months. To add clarity and avoid this misunderstanding in future, "early deaths" should be relabelled.	
Several paragraphs including Section 3.2.5.6 page 28 and Section 1.2 page 10-11 and Section 3.6 page 31	Text around "early deaths" should be marked AIC. This includes magnitude and direction of effect	This information has not yet been published and should be marked as AIC. It should be noted that the OS Kaplan-	The ERG has made the requested amendments. **AIC marking added to requested**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Text around “early deaths” should be marked AIC. This includes magnitude and direction of effect		Meier for ATTRACTION-3 has been published and is publicly available. This is considered distinct from the endpoint under discussion.	**text.**

Issue 6 Indirect treatment comparison

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Meta-analysis between BSC and taxanes combined Section 3.4, page 30 The ERG state: *“It is possible that an alternative meta-analysis including BSC against a pooled taxane arm could have provided a more direct approach to constructing a comparison between nivolumab and BSC”.*	Amendment on the text to describe the rationale provided in the CS that it was not considered reasonable to construct a meta-analysis using the pooled taxane arm	It was not considered reasonable to construct a meta-analysis using the pooled taxane arm and this was detailed in the following sections: In Section 2.9.2, page 60 of the CS, the submission states “There was considerable inconsistency in treatments included. For example, while a number of studies included BSC, all the comparators were different. This would introduce considerable heterogeneity and reduce transitivity if all were to be included in the network.” For this reason, the most appropriate BSC was chosen. Further in Section 2.9.2.1, page 63 of the CS it is stated, “An additional study linked irinotecan to a mixed docetaxel/paclitaxel arm [was available], although this did not report the ratio of docetaxel and paclitaxel received or the dosing regimens.[9] This	The ERG does not regard this to be a factual error. Any approach to ITC requires simplifications and assumptions; indeed, the ERG’s point remains valid, that the ITC itself was not informative and that an alternative approach may also have provided information. Moreover, it is surprising that the company did not regard a combined taxanes arm as relevant for ITC given that it was considered relevant in the primary analysis of the trial. **No action required.**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
		would require the assumption that ratio and dosing of docetaxel/paclitaxel are equivalent to the control arm of ATTRACTION-3. While this would allow a link between the combined control arm and irinotecan, there would be no link to BSC: docetaxel could not be included separately due to the lack of studies comparing docetaxel with combined taxanes and there is only one link from docetaxel to BSC available in the network. The resulting network would be minimal and offer no information about BSC". The study described in the above paragraph was the only study identified, other than ATTRACTION-3, that would compare a mixed taxane arm to any other treatment. The resulting network would include only two studies (less than the submitted network) and no closed loops (a criticism of the submitted network). Additionally, this study was retrospective which was another criticism of the submitted analysis.	
Validity of ITC analysis Section 3.4, page 30 The ERG state: *"… insufficient details regarding burn-in iterations discarded or checks for convergence were provided to provide confidence in the*	Amendment of the text to reflect that the code provided as part of the clarifications response provided these additional details, which were not requested as part of the clarifications stage	During clarification, the ERG requested additional details from the ITC and all of these were provided. Additional details regarding burn-in iterations discarded or checks for convergence specifically were not requested by the ERG. However, the code used and accompanying notations in the code	This is not a factual error. Reporting in the context of code provided for an analysis does not constitute a clear statement of the models estimated, nor of the settings used. **No action required.**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
analysis presented.".		were supplied to the ERG in the CS. In the code, it is clearly commented that the analyst believed convergence was achieved at 50,000 iterations. Finally, analysis results (that match those in the clarification report and the CS) are contained in the files, along with the iterations discarded and number of runs. Therefore, the ERG did have this information at the time of writing their report. Nevertheless, the company would be happy to provide any further details that the ERG felt were insufficiently reported. For clarity, convergence was assessed using the Brooks-Gelman-Rubin convergence statistics and tools within WinBUGs as well as monitoring the history of all relevant parameters.	
Input values for Shirakawa et al. (2014) Section 3.4, page 30 The report states that there could be no confirmation of the input values for one study *" . . . between docetaxel and paclitaxel to the corresponding estimates in the included studies, in particular for Shirakawa et al. (2014)" and then later that " . . . the response to clarification question A18 seems to be at variance with this, noting that reconstructed data 'were not used as*	Amendment of the text to reflect the methodology applied within the NMA, specifically for this instance.	In Section B2.9.4.1, page 67 of the CS, it is stated that: "Where an HR was reported, this [reconstructed] value was used. Only if there was no HR reported, the reconstructed value was used. This is because the reported values in the literature were calculated with PLD and are therefore considerably more accurate than HRs calculated with digitised data." The company would like to clarify that no HR was reported in Shirakawa et al. and thus the reconstructed HR was used. However, for other studies, the HR was reported and reconstructed for	The ERG does not require this to be a factual error. This was a lack of clarity generated by the company's response to the relevant clarification question. **No action required.**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
inputs to the NMA'"		validation of this and the Shirakawa et al value, thus improving confidence in this input.	
Limitations of ITC analysis Section 3.3 page 29 The ERG report discusses the ITC and its limitations, stating*: "For example, it appears that the docetaxel-paclitaxel comparison drawn from ATTRACTION-3 was naïve, without due regard to baseline differences between patients receiving docetaxel and patients receiving paclitaxel; similarly, the additional studies used either multivariable adjustment or naïve comparison to estimate relative effectiveness. Second, included populations were sicker than the population included in ATTRACTION-3,2 including a wider range of ECOG PS scores, suggesting incommensurability with ATTRACTION-3 estimates. Third, outcomes were not measured consistently across included studies. Specifically, Moriwaki et al. (2014) used post-progression survival instead of OS.[10] The exchangeability of this effect estimate with OS is a question of assumption rather than fact, though the company note in response to clarification question A19 that post-progression survival was 'comparable to the definition of OS applied in other*	Amendment of the text to reflect the company acknowledgement of the limitations, along with the acknowledgement that the NMA was undertaken in order to provide all potential evidence for decision making. In the absence of this evidence, no comparison versus BSC would have been feasible	The company and the ERG agree that BSC is a relevant comparator and that the BSC comparison is based on poor quality evidence, due to limitations in the published data for BSC. However, the NMA is provided as an exploratory analysis to aid decision making and should be viewed in this context. As the primary comparators are agreed to be taxanes, the limitations inherent in the BSC comparison do not impact on the base case analysis cost-effectiveness conclusions. Frequently during reporting, the Company described the reasons for attempting to construct a network to include BSC. It also describes the limitations at length and cautions against their use and does not include these in the base case results. Examples of mentions are: In Section 2.9.2, page 60 of the CS – "There was considerable inconsistency in treatments included" In Section 2.9.2, page 61: "All studies reported in the literature applied inclusion criteria that allowed patients with an ECOG PS score of 2 to be included (Table 23). This is	This is not a factual error. **No action required.**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
studies'. Finally, all studies except for ATTRACTION-32 draw on Japan-only populations. As discussed elsewhere (Section 3.2.4), treatment pathways and disease presentation vary significantly between Japan and ROW, including the UK."		contradictory to the inclusion criteria of ATTRACTION-3 (B.2.6.1.2). However, due to the absence of other studies to inform these links, they were all included in this NMA. The impact of including these different populations is discussed in Section B.2.9.3.1" Table 23 outlines the differences between studies. In Section B2.9.3, page 63 of the CS, Moriwaki et al (2014) is discussed, and includes the authors definition of PPS and then states; "This is comparable to the measurements used in other studies and so it was included as if it were a measure of OS." In Section B2.9.3.1. pages 65-66 of the CS there are details of the numerous limitations that the ERG has also reported. In Section B2.9.8 of the CS, details the limitations again and it is stated "Given this limitation, the results should be considered with caution.". Finally, in Section B2.9.8 of the CS the following is stated: "Another important limitation is the quality of the input studies. The included studies were all retrospective, aside from ATTRACTION-3. Therefore, patients included from these trials were not randomised and so this would not be considered high quality input data for	

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
		analysis. While this does not mean they are uninformative, it should be considered while examining the outputs of analysis. This is often a limitation of any evidence synthesis in indications that are sparsely reported on." It is particularly important to recognise that the inclusion of retrospective studies was done out of necessity and highlights the sparsity of evidence with which to construct networks in this indication.	
Section 3.4 Page 29 The ERG state: *"The company presented an ITC comparing docetaxel, paclitaxel and BSC. As described above and in Section 3.3, this network was ultimately sparse, and was only estimated for OS as an outcome.".*	Suggested rewording ". . . and was only possible for OS as an outcome" to reflect the capacity of evidence.	As there was no information available pertaining to the PFS expected with BSC there could be no functional ITC, or indeed any treatment comparison. The sparsity of network is expected in an indication where this	This is not a factual error. **No action required.**

Issue 7 Cost-effectiveness modelling

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Section 4.2.1., page 35 The ERG state: *"A shorter time horizon would likely have been sufficient as the majority of patients with current care in the UK have an estimated survival of*	Suggest removal of this section as there is no evidence to support this assumption.	The Guide to the methods of technology appraisal 2013 states that analysis should be "*Long enough to reflect all important differences in costs or outcomes between the technologies being compared*" and that "*Analyses*	The ERG does not consider this a factual error. The submitted model estimates costs and outcomes over a 40-year time frame. The ERG's report notes that a shorter time horizon would likely have been sufficient, which based

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
<12 months.".		*that limit the time horizon to periods shorter than the expected impact of treatment do not usually provide the best estimates of benefits and costs*.".[11] While the Company do not dispute that the "*majority of patients with current care in the UK have an estimated survival of <12 months*", this is accurately reflected in ATTRACTION-3, where standard of care is associated with median OS less than 12 months. However, although the majority of patients (i.e. >50%) will have died by 12 months, all patients should be modelled in line with NICE guidance. Of note, where the ERG's base case applies a significantly more conservative extrapolation, patients remain alive at 10 years in the nivolumab arm. As such, even under the most conservative assumptions, a ten-year time horizon is insufficient to model outcomes in line with NICE guidance.	on the response provided, the Company appears to agree with. Therefore, no error is included in the ERG's report. **No action required.**
Section 4.2.5 page 44 The ERG state: "*A time horizon of 40 years was used to inform the company's base-case analysis."* This is incorrect and misleading.	Suggested rewording "A lifetime time horizon was used to inform the company's base case analysis"	The Guide to the methods of technology appraisal 2013 states that analysis should be "*Long enough to reflect all important differences in costs or outcomes between the technologies being compared*"[11], which the Company considers to be a lifetime. Indeed, the company model was *up to* 40 years in order to reflect the whole life of every single patient. This was	The model time horizon (per the company's submitted economic model) is 40 years. The ERG agrees that 40 years was definitely sufficient, and that this covered a lifetime horizon. Describing the model time horizon as 40 years is not an error or misleading. **No action required.**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
		considered appropriate in order to facilitate scenario and sensitivity analysis where baseline age is younger. The base case analysis applies a baseline age of 63.82 years, and all patients die by age 100 when all-cause mortality is applied. Hence, in the base case analysis, the maximum life expectancy is 36.18 years, even before disease-specific mortality is applied. In the nivolumab arm of the base case analysis, the last person dies [redacted] years after the model initiates (i.e. significantly less than 40 years, but appropriately described as a lifetime horizon). For this reason, it is inaccurate and misleading to describe the base case analysis as applying a time horizon of 40 years.	

Issue 8 Third-line therapy

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Section 1.3 page 12 and Section 4.2.2 page 38 The ERG state: *"Should any active intervention be used in the third-line*	Amendment of stated text	Clinician advice obtained by BMS suggested BSC is the most common therapy in the third-line setting. However, if active interventions are modelled then this should be applied	No edit needed – the ERG's comment was made on the basis that if the model is appropriate for decision making even if third-line therapy is used, this implication is important to

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
setting, this assumption may lead to an over-estimation of costs incurred." The model is built to reflect the treatment pathway, so that an active treatment in the third line setting would require a model adaptation rather than a model input update.		as separate therapy line, as opposed to the observed BSC state. Further, it should be noted that BSC use is lower in patients also receiving an active therapy, so there will be a decrease in BSC resource use as a result of this amendment. Due to the high cost nature of BSC elements, the current absorbing therapy line may be considered conservative	consider. However, based on the Company's comment here, it may be that the model would not be fit for decision making were third-line therapy used (without model adaptation being needed). Nevertheless, for the purpose of the ERG report communicating the features of the submitted model, this is not an error. **No action required.**
Section 1.3 page 12 and Section 4.2.2 page 39 The ERG state: *"Consequently, the benefits associated with nivolumab may be over-estimated (even if by a small quantity) owing to the specification of a survival model which masks potential benefits accrued concerning the use of treatment(s) after discontinuation of nivolumab (which are not considered standard UK practice). Subsequent therapy costs are discussed further in Section 4.2.8"* This doesn't take into account subsequent treatment use across both arms	Amendment to reflect that subsequent therapies are available in the UK.	Of note, [redacted] As therapies are predominantly available in the UK, this may be reflective of survival in the UK setting.	The ERG accepts that subsequent therapy could affect both arms, and has edited the text accordingly. **Sentence edited in 4.2.2:** Revised text to clarify that subsequent treatments are available in the UK. **There was no text in 1.3 that required updating.**
Section 1.3 page 12 and Section 4.2.2 page 39 The ERG state: "*No adjustment to efficacy was made for any beneficial effects of active third-line therapy."*	Amendment to reflect that an adjustment likely wasn't clinically valid or appropriate.	UK clinical practice in the third line setting is predominantly best supportive care. Hence, it unlikely that any third-line therapy used in clinical practice will impact on survival	This is not a factual error. **No change required.**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
		outcomes	

Issue 9 Typographical/ transcription errors

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Section 1.2, page 10 The ERG state: *"Moreover, 97% of patients in ATTRACTION-3[1] were Asian, while approximately two-thirds of total patients were Japanese."* This percentage is incorrect.	96% of patients in ATTRACTION-3 were Asian.	Of the included 418 patients, 401 were Asian, resulting in 96% of included patients in ATTRACTION-3 to be Asian.	The ERG accepts this is an error. **The ERG has updated the value.**
Section 1.2, page 10 The ERG state: *"Japanese patients receiving nivolumab had considerably longer OS than ROW patients ([redacted] vs [redacted]) and it is notable that Japanese patients on taxanes had superior OS than ROW patients on nivolumab ([redacted] vs [redacted])"* The median OS for ROW patients receiving taxanes is incorrect.	ROW patients receiving taxanes had a median OS of [redacted]	Please refer to page 1958 of the CSR for subgroup analyses for overall survival presented in Figure 14.2.6-1.	Median OS for ROW patients receiving taxanes is not listed in this sentence. All three percentages listed are correct. **No action required**
Section 1.2, page 10 The ERG state: *"Around 64% of OC cases are adenocarcinoma with around 31% being oesophageal*	The correct reference (as used in the CS) is: National Cancer Institute. SEER Cancer Statistics Review (CSR) 1975-2016	The reference provided by the ERG, *Cancer Research UK. Oesophageal cancer incidence statistics, 2019. Available from:*	The ERG has replaced Cancer Research UK (2019) reference with National Cancer Institute (2019) in section 2.1, page 14.

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
squamous cell carcinoma (OSCC).” This statement is supported by an incorrect reference.	2019 [Available from: https://seer.cancer.gov/csr/1975_2016/results_merged/sect_08_esophagus.pdf.	*https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/oesophageal-cancer/incidence*. does not contain the percentages of adenocarcinoma and oesophageal squamous cell carcinoma (OSCC) cases, respectively.	**The ERG has provided the correct reference.**
Section 2.1, page 14 The ERG state: *“While relatively rare in terms of incidence, OC represents the seventh most common cause of cancer death in the United Kingdom (UK), responsible for an estimated 7,295 deaths in the UK in 2017, reflecting extremely poor survival rates, with only around 15% of people diagnosed with OC surviving five years or more”* The ERG did not include a reference to the number of deaths and quoted an incorrect number of OC deaths in the UK in 2017.	OC was responsible for 7,925 death ins the UK in 2017 based on the following reference (as cited in the CS): Cancer Research UK. Oesophageal cancer mortality statistics 2018 [Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/oesophageal-cancer/mortality.	This represents the correct number of OC deaths in the UK in 2017.	The ERG has added a reference to the Cancer Research UK Oesophageal cancer mortality statistics and provided the correct number of OC deaths in the UK in 2017 (7,925). **The ERG has updated the value.** Corrected number of OC deaths.
Section 3.2.1 page 21 The ERG state: “*The data reported in the CS are from planned subgroup comparisons of nivolumab and chemotherapy (paclitaxel or docetaxel) in patients who broadly met the NICE decision problem criteria.”*	Amendment to: The data reported in the CS are from a comparison of nivolumab and chemotherapy (paclitaxel or docetaxel) in patients who broadly met the NICE decision problem criteria, with planned subgroup analyses provided.	The overall population forms the most appropriate information for decision making, although planned subgroup analyses are presented	The ERG accepts the company’s preferred phrasing. **Suggested edit made.** Revised phrasing of population.

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Section 3.2.5.5., page 27 The ERG state: *"Japanese patients receiving nivolumab had considerably longer OS than ROW patients ([redacted] vs [redacted]) and it is notable that Japanese patients on taxanes had superior OS than ROW patients on nivolumab ([redacted] vs [redacted])."* The median OS for ROW patients receiving taxanes is incorrect.	ROW patients receiving taxanes had a median OS of [redacted]	Please refer to page 1958 of the CSR for subgroup analyses for overall survival, as presented in Figure 14.2.6-1.	Median OS for ROW patients receiving taxanes is not listed in this sentence. All three percentages listed are correct. **No action required**
Section 3.6., page 31 The ERG state: *"The generalisability of ATTRACTION-3[1] to UK practice is a concern given the fact that 97% of patients were Asian and approximately two thirds were from Japan."* This percentage is incorrect.	96% of patients in ATTRACTION-3 were Asian.	Of the included 419 patients, 401 were Asian, resulting 96% of included patients in ATTRACTION-3 to be Asian.	The ERG accepts this is an error. **The ERG has corrected % of Asian participants**
Section 4.2.2., page 38-39 The ERG state: *"[redacted] of the nivolumab arm received treatment after progression for a median of [redacted], versus [redacted] of taxanes patients for a median of [redacted]."*. Results are currently no marked as AIC.	Highlight the % of patients receiving treatment beyond progression and the median of days these patients received treatment.	As this data was not part of the ATTRACTION-3 publication, these results should be marked as AIC.	The ERG has made the requested amendment. **AIC marking added to requested text.**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
Section 4.2.3., page 41-42 Subgroup analysis results from ATTRACTION-3 are incorrectly marked as CIC.	Mark the results presented on page 41-42, which are currently marked as CIC, as AIC.	As this data was not part of the ATTRACTION-3 publication, these results should be marked as AIC.	The ERG notes that CSR data are typically CIC by default so this is not an error, but notes and accepts the company's preference for AIC. **Changed marking from CIC to AIC for subgroup results from CSR as suggested.**
Section 4.2.8.1., page 72 The ERG state: *"The model accounted for this adjustment based on the proportion of actual versus expected doses received, which for nivolumab was equivalent to [redacted] of doses (CS Table 58)." .* The proportion of actual doses received is not marked as AIC.	Mark as AIC.	As this data was not part of the ATTRACTION-3 publication, these results should be marked as AIC.	The ERG has made the requested amendment. **AIC marking added to requested text.**
Section 4.2.8.1 page 73-74 The ERG state: *At clarification stage, the ERG asked the company clarify what proportion of patients received post-progression treatment in ATTRACTION-3[1] by treatment arm (clarification question A9), acknowledging that BSC is expected to comprise the mainstay of current NHS practice in this setting. In response, the company stated that 82 of the 210 nivolumab patients (39.0%) received treatment post-progression, with a median of 3 treatments (range: 1-52 treatments) and a median post-discontinuation time on treatment of*	The text should be amended to reflect that the clarification response provided what the ERG requested. However, the clarification response was not worded in such a way to communicate the request. Further, text around discrepancy between data should be deleted.	As context, Question A9 states: Please clarify what proportion of patients received post-progression treatment in ATTRACTION-3 by arm (i.e. the wording did not contain context around BSC comprising subsequent treatment). During ATTRACTION-3, post-progression treatment was defined as patients who maintained current treatment following progression. Progression was not a hard criterion for treatment discontinuation, as acknowledged within the ERG report. The information provided in the clarification response is thus accurate:	The ERG has removed the final sentence for clarity. **Final sentence removed.**

Description of problem	Description of proposed amendment	Justification for amendment	ERG response
32.5 days (95% CI: 28-39 days). For the taxanes arm, three of the 209 patients (1.4%) received treatment post-progression; all patients had one subsequent treatment and a median post-discontinuation time on treatment of one day. *However, in the pivotal trial publication of ATTRACTION-3,[1] it is stated that 119 (57%) of 210 patients in the nivolumab group and 115 (55%) of 209 patients in the taxanes group received subsequent therapy for advanced oesophageal cancer (though is not described in relation to progression status). Furthermore, it is noted that the most common subsequent treatments were taxanes (for 100 [48%] of the 210 patients in the nivolumab group and 43 [21%] of 209 patients in the chemotherapy group), fluoropyrimidine-based chemotherapies (24 [11%] of 210 and 39 [19%] of 209), and platinum-based chemotherapies (20 [10%] of 210 and 22 [11%] of 209).* *The ERG could not establish why these figures differ to such an extent."* This is incorrect and misleading		82 of the 210 nivolumab patients (39.0%) received treatment post-progression [i.e. continued treatment with nivolumab following progression], with a median of 3 treatments (range: 1-52 treatments) and a median post-discontinuation time on treatment of 32.5 days (95% CI: 28-39 days). For the taxanes arm, three of the 209 patients (1.4%) received treatment post-progression [i.e. continued treatment with taxanes following progression]; all patients had one subsequent treatment and a median post-discontinuation time on treatment of one day. The ERG require information regarding subsequent anti-cancer therapy, as acknowledged by the wording within the ERG report (which states: *it is stated that 119 (57%) of 210 patients in the nivolumab group and 115 (55%) of 209 patients in the taxanes group received subsequent therapy for advanced oesophageal cancer (though is not described in relation to progression status).*	

References

1. Kato K, Cho BC, Takahashi M, Okada M, Lin CY, Chin K, et al. Nivolumab versus chemotherapy in patients with advanced oesophageal squamous cell carcinoma refractory or intolerant to previous chemotherapy (ATTRACTION-3): a multicentre, randomised, open-label, phase 3 trial. Lancet Oncology. 2019;20(11):1506-17.
2. Moustgaard H, Clayton GL, Jones HE, Boutron I, Jørgensen L, Laursen DRT, et al. Impact of blinding on estimated treatment effects in randomised clinical trials: meta-epidemiological study. BMJ. 2020;368:l6802.
3. Latimer N. NICE DSU Technical Support Document 14: survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data, 2013. Sheffield, UK: Decision Support Unit, ScHARR; [updated 7 June 2016. Available from: http://nicedsu.org.uk/wp-content/uploads/2016/03/NICE-DSU-TSD-Survival-analysis.updated-March-2013.v2.pdf.
4. Bagust A, Beale S. Survival Analysis and Extrapolation Modeling of Time-to-Event Clinical Trial Data for Economic Evaluation An Alternative Approach. Medical Decision Making. 2014;34(3):343-51.
5. National Institute for Health and Care Excellence (NICE). Technology appraisal guidance [TA525]. Atezolizumab for treating locally advanced or metastatic urothelial carcinoma after platinum-containing chemotherapy. Final appraisal determination, 2018 [Available from: https://www.nice.org.uk/guidance/ta525/documents/final-appraisal-determination-document.
6. Personal Social Services Research Unit. Unit Costs of Health and Social Care 2019 2020 [Available from: https://www.pssru.ac.uk/project-pages/unit-costs/unit-costs-2019/.
7. National Institute for Health and Care Excellence. PleurX peritoneal catheter drainage system for vacuum-assisted drainage of treatment-resistant, recurrent malignant ascites. 2012.
8. Improvement N. Reference costs 2017/18: highlights, analysis and introduction to the data 2018 [Available from: https://improvement.nhs.uk/documents/1972/1_-_Reference_costs_201718.pdf.
9. Auzolle C, Dubreuil O, Pozet A, Coriat R, Dhooge M, Ducreux M, et al. 2316 Efficacy and toxicity of second-line chemotherapy in patients with advanced oesophageal squamous cell carcinoma progressing after a first line of 5-fluorouracil and platinum-based therapy: An AGEO retrospective multicentric study. European Journal of Cancer. 2015;51:S438.
10. Moriwaki T, Kajiwara T, Matsumoto T, Suzuki H, Hiroshima Y, Matsuda K, et al. Survival analysis of platinum-refractory patients with advanced esophageal cancer treated with docetaxel or best supportive care alone: a retrospective study. Dis Esophagus. 2014;27(8):737-43.
11. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013 2013 [Available from: https://www.nice.org.uk/article/pmg9.

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Nivolumab for unresectable, advanced oesophageal cancer when standard chemotherapy has failed [ID1249]

This document is the technical report for this appraisal. It has been prepared by the NICE technical team.

The technical report and stakeholder's responses to it are used by the appraisal committee to help it make decisions at the appraisal committee meeting. Usually, only unresolved or uncertain key issues will be discussed at the appraisal committee meeting.

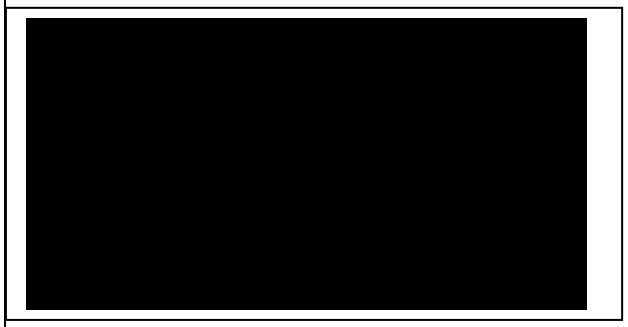
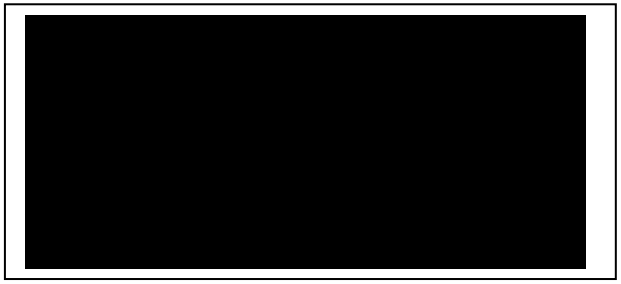
This report is based on:

- the evidence and views submitted by the company, consultees and their nominated clinical experts and patient experts and
- the evidence review group (ERG) report.

The technical report should be read with the full supporting documents for this appraisal.

1 Key issues summary

Issue	Summary	Technical Team Preliminary Judgement
Issues related to clinical evidence		
Clinical effectiveness of nivolumab	Most of the overall survival benefit from nivolumab in ATTRACTION-3 is in the post progression phase, as there is no progression free survival benefit associated with nivolumab (Kaplan-Meier estimate of median PFS was 1.68 months in nivolumab group and 3.35 months in the control group). Furthermore, there is little difference in the overall response rate between nivolumab and taxane therapy (19.3% versus 21.5% with an odds ratio of [redacted] It is not clear how nivolumab is achieving a benefit in overall survival without a benefit in progression-free survival or the overall response rate. What rationale is there for this treatment prolonging post progression survival? Patients on nivolumab treatment also need to survive beyond 3 months (during which they have an increased risk of death) before they can benefit from improved overall survival. To what extent is this observed in other cancer treatments? The company have used a Cox proportional hazards model to estimate hazard ratios and 95% confidence intervals for overall survival and progression free survival. The ERG note that the proportional hazard assumption was violated (the two treatment curves crossed for both overall survival and progression-free survival). See figures 10 and 11 of the company submission, document B below).	The methods used to estimate the efficacy of nivolumab compared with taxanes assume constant relative efficacy of nivolumab compared with taxanes over time. If the proportional hazards assumption does not hold, the effectiveness of nivolumab in the model may be overestimated **<u>What additional information at engagement would help address the issue?</u>** • How is nivolumab improving overall survival without improvements in progression-free survival or the overall response rate? Is this observed with other cancer treatments? • Clinical advice would be useful on whether the relative treatment effect of nivolumab compared with taxanes is likely to be constant over time. • Is the risk of 'early death' with initial nivolumab treatment worth an additional 2.58 months overall survival reported in the trial?

	Figure 1. Kaplan-Meier plot of overall survival in patients receiving nivolumab or taxane  ***Figure 2. Kaplan-Meier plot of progression-free survival in patients receiving nivolumab or taxane*** 	
Is best supportive care a relevant comparator?	• Direct comparative evidence for nivolumab vs taxane chemotherapy was available in ATTRACTION-3 • The company and clinical advisors to the ERG noted that irinotecan is not a relevant comparator because it is only used in 6% of patients • The ERG did not consider comparisons of nivolumab with individual taxanes to be relevant	The ERG and technical team do not consider best supportive care to be a relevant comparator because the pivotal trial was conducted in taxane-eligible patients. The ITC used to derive survival data for best supportive care contained several flaws, which make results from comparisons with best supportive

	because patients were not randomised to taxane treatment, rather this was chosen by the investigator • Best supportive care was included as a comparator in the final scope. Because there is no direct trial evidence comparing nivolumab with best supportive care, the company provided an exploratory indirect treatment comparison (ITC). However, it did consider this analysis to be robust. • The ERG noted several flaws in the ITC including studies that were not randomised, lack of adjustment for differences in baseline characteristics (including ECOG performance status), inconsistent measurement of outcomes (e.g. post progression survival used to approximate overall survival) • The ERG also questioned the generalisibility of the ITC results to NHS practice because all of the trials in the network (except ATTRACTION-3) included Japanese-only patients, who have better survival outcomes • The company note the unmet need in people for whom taxane therapy is unsuitable but have not proposed or asked NICE to consider the clinical and cost effectiveness of nivolumab in this population	care highly uncertain (see Section 3.3. of the ERG report) **<u>What additional information at engagement would help address the issue?</u>** • Clinical expert advice on the most relevant comparator for NHS practice. • Is irinotecan used in clinical practice for patients who would be eligible for nivolumab? • What proportion of people currently receive docetaxel or paclitaxel in NHS practice? • Is best supportive care a relevant comparator? Can the efficacy of nivolumab in ATTRACTION-3 be generalised to people for whom best supportive care is suitable? Is the ITC robust given the heterogeneity between studies? • Are the results of the ITC generalisable to NHS practice?
Generalisability of ATTRACTION-3 participants to UK population	• 96% of ATTRACTION-3 patients were Asian, two-thirds of which were Japanese • Treatment in Asia follows pan-Asian adapted European Society for Medical Oncology guidelines which recommend treatment options that are not available in Europe (for example nedaplatin) • Nivolumab is associated with longer overall survival compared with taxane therapy (10.91 vs 8.38 months, HR 0.77 [95% CI: 0.62, 0.96] P <0.0001) in the intention to treat population. • Clinical advisers to the ERG noted that this is clinically meaningful.	• Differences in the treatment pathway and improved health outcomes for patients in Asia mean that the clinical effectiveness in the intention to treat population is not likely to be realised in NHS practice • The efficacy of nivolumab vs taxanes in the rest of world population may be more suitable to estimate the effectiveness of nivolumab in the UK context, provided that the rest of world

	• The relative efficacy of nivolumab compared with taxanes in the Japanese population (hazard ratio [redacted]) is similar to the rest of the world ([redacted]). However, the absolute overall survival in Japanese patients is higher for both taxanes ([redacted] compared with [redacted]) and nivolumab treatment ([redacted] compared with [redacted]). Overall survival with taxanes in the Japanese population is longer than overall survival with nivolumab in the rest of the world population ([redacted] compared with [redacted]). • Efficacy data in the model was based on the hazard ratio in the intention to treat population from ATTRACTION-3. • Efficacy data based on rest of world data only may be more relevant to NHS practice (although this also included other populations such as Korean) • Only patients with Easten Cooperative Oncology Group (ECOG) performance scores 0-1 were included in the trial • The company also highlighted that individuals randomised into clinical trials are likely to be slightly younger and healthier than the overall oesophageal cancer patient population in the UK	population is more reflective of patients seen in NHS practice • Participants in ATTRACTION-3 are also fitter and more able to tolerate treatment with nivolumab and taxanes than people in UK practice **What additional information at engagement would help address the issue?** • Clinical expert clarification on whether the rest of the world efficacy data is more relevant to the expected effectiveness of nivolumab in the NHS practice • Who is likely to receive nivolumab in clinical practice? Will it be restricted to people with a good performance status (ECOG 0-1)? Is the efficacy of nivolumab in NHS practice (people with a worse performance status) likely to be worse than in the ATTRACTION-3 ? • Is there a subgroup within ATTRACTION-3 which is considered to have similar characteristics to the UK patient population?
Safety data for nivolumab, deaths higher on nivolumab in first 3 months	• The ERG noted that the safety profile of nivolumab is favourable compared to taxanes. However, the number of deaths in the first 3 months of treatment was higher in the nivolumab compared with the taxane arm ([redacted]) • During clarification, the company explained that this effect may relate to the differences in mechanism of action of immunotherapy treatments compared to chemotherapy agents. In particular, that	• The ERG noted that the rationale provided by the company (attributable to its mechanism of action) does not adequately explain the reason for higher deaths on nivolumab compared with taxane therapy in first 3 months. **What additional information at engagement would help address the issue?**

	immunotherapies are associated with a longer time to response which is seen with PDL-1 inhibitors.	• Do clinical experts agree with the company's rationale for higher death rate over first 3 months in the nivolumab arm? • Clinical expert advice is sought on whether the initial higher death rate with nivolumab in ATTRACTION-3 is likely to be seen in NHS practice. • Is it possible to determine in advance which patients are likely to die before they can benefit from treatment with nivolumab (if so please explain)?.
No adjustment to efficacy for any beneficial effects of third-line therapy	• A single overarching OS curve was used to estimate number of deaths per model cycle. The proportion of deaths occurring for patients still on treatment vs off treatment could not be separately calculated • The company model base case assumed that no active therapy is given 3rd line. • The ERG suggests that some of the overall survival benefit in the trial is related to the active 3rd line treatment which is not balanced between the two arms. A higher percentage of people in nivolumab arm received active treatment after progression ([redacted] of the nivolumab arm received treatment after progression for a median of [redacted], versus [redacted] of taxanes patients for a median of one day). Benefits associated with nivolumab may therefore be over-estimated. • The ERG note that no adjustment to efficacy has been made for any beneficial effects of active third-line therapy • The ERG questioned whether the assumption that after progression people would receive active 3rd line treatment until death is relevant to NHS practice	• ATTRACTION-3 allowed active subsequent treatment after progression which contributes to the improved overall survival of nivolumab compared with taxanes • In NHS practice patients are unlikely to receive active treatment at this stage in the treatment pathway. The company did not adjust the efficacy of nivolumab for the efficacy of 3rd line treatment. Therefore, efficacy data which has been used in the economic model is likely to overestimate the benefits of nivolumab **<u>What additional information at engagement would help address the issue?</u>** • Clinical experts to clarify whether patients in NHS practice receive active treatment after progression on nivolumab or taxanes. What proportion of patients receive different 3rd line treatments?

		• Is nivolumab expected to completely replace taxane use for the proposed indication and population? If so, would taxane therapy then be offered to nivolumab patients post-progression? • Should the efficacy of nivolumab in ATTRACTION-3 be adjusted to remove the impact of active subsequent treatment?
Subgroup analysis by taxane was not provided	• The comparator arm in ATTRACTION-3 consisted of either docetaxel or paclitaxel (investigators choice which was recorded pre-randomisation). • Following clarification the company noted that because the choice of taxanes was recorded pre-randomisation this was not a stratification factor, therefore subgroup analyses by taxane would not be provided. • The ERG noted that a potentially more robust comparison of nivolumab vs each taxane could have been made available to support decision analytic modelling. This would allow comparisons to consider the effectiveness of nivolumab in 'docetaxel-preferred' or 'paclitaxel-preferred' populations	**<u>What additional information at engagement would help address the issue?</u>** Clinical advice on the following is required: • Is it reasonable to assume a class effect for taxane therapy? • Are there systematic differences in people who would be suitable for treatment with either docetaxel or paclitaxel? How should these separate populations be defined? • Could post-hoc analysis of the effectiveness of efficacy of nivolumab compared with either docetaxel or paclitaxel have been carried out? • Is a comparison of nivolumab with individual taxanes more relevant to NHS practice than comparisons with a combined taxane arm?
Issues related to cost-effectiveness evidence		
Is the model time horizon (40 years) appropriate?	• The majority of patients undergoing second-line taxane therapy in current UK practice would have an estimated survival of less than 12 months • The ERG consider extrapolated survival benefits of nivolumab over a 40 year time horizon may give	**<u>What additional information at engagement would help address the issue?</u>** • Clinical expert advice is required on the life expectancy of patients with unresectable, advanced oesophageal

	unrealistic mean life years gained depending on the survival models used (ERG report Section 4.2.6.) • The ERG advised that a 10-year time horizon should be long enough to capture differences in the costs and benefits	cancer treated with nivolumab or taxanes • What is the most relevant time horizon for the economic model?
Alternative extrapolations for overall survival	• Kaplan-Meier estimate of median OS in the intention to treat population was 10.91 months in the nivolumab group and 8.38 months in the control group. The OS rates were reported to be numerically higher in the nivolumab group from follow-up at month 6 through to month 30. • Survival modelling methods were used to extrapolate over the lifetime horizon of the model • The company used a semi-parametric model using Kaplan-Meier data to 2.99 months then parametric model (log-logistic model in nivolumab arm and exponential in the taxane arm) for the remainder of the time horizon. The rationale for this was that a semi-parametric model could take into account changes in hazard over time (that is high hazard associated with the early death rate on nivolumab and lower hazard over the longer term. • The ERG agreed that a fully parametric model may not be suitable because of the initial early deaths in the trial It noted that although there is a change in hazard in the nivolumab arm at approximately 3 months there is no change in hazard in the taxane arm. • The company provided alternative analyses in response to a clarification request from the ERG with cut offs of 4.37 (before any deaths) and 5.75 months (before the arms crossed over). See response to question B10 of the clarification response and Figure 3 in the ERG report.	**<u>What additional information at engagement would help address the issue?</u>** • Estimation of OS is a key driver of cost-effectiveness results • Is a fully-parametric or semi-parametric model most appropriate to predict the long term effectiveness of nivolumab and taxane therapy? • If a semi-parametric model is preferred, is 2.99 or 5.75 months an appropriate point to start the extrapolation? • Does the overall survival predicted in the company (1.92% at 10 years) or ERG model (0.2% at 10 years) best represent the likely overall survival in NHS practice?

- The ERG noted that the long term predictions from fully or semi-parametric models may be poor because they may not accurately reflect hazards over time. It further noted that the plausibility of longer-term extrapolations should be carefully considered.
- The ERG base case model used a semi-parametric approach using Kaplan-Meier curve until 5.75 months, followed by a generalised gamma model (both arms) for OS extrapolations because they were a better visual fit.
- 10-year OS in the company's base-case was estimated to be 1.92% for the nivolumab arm, versus 0.20% for the ERG's base-case analysis.

Table 23 ERG report: Comparison of company and ERG-preferred OS extrapolations

	Company		**ERG**	
Description	SP approach using Kaplan–Meier cure until 2.99 months, followed by a log-logistic (nivolumab) or exponential (taxanes) model		SP approach using Kaplan–Meier cure until 5.75 months, followed by a generalised gamma model (both arms)	
Time (years)	**Nivolumab**	**Taxanes**	**Nivolumab**	**Taxanes**
1	45.61%	36.57%	46.07%	35.40%
2	21.27%	11.06%	20.70%	12.20%
3	12.33%	3.34%	10.22%	4.42%
4	8.15%	1.01%	5.36%	1.65%

	5	5.84%	0.30%	2.93%	0.63%
	6	4.39%	0.09%	1.63%	0.24%
	8	2.78%	0.01%	0.56%	0.04%
	10	1.92%	0.00%	0.20%	0.01%

Abbreviations: KM, Kaplan–Meier; SP, semi-parametric

Exploratory analysis of utility values	• Company estimated utility values were based on EQ-5D data from ATTRACTION-3 • The base case utilities in the company model are:	The technical team considers the following: • It is not appropriate to set a pre-progression utility which exceeds the age-adjusted population mean • The magnitude of differences in utility between the nivolumab and taxane arm in the company model are not clinically plausible • A treatment independent approach (e.g. using the same values for pre and post progression across the treatment arms) is more appropriate for the estimation of utility.

Health state	Nivolumab utility: mean (SE)	Control utility: mean (SE)
Pre-progression	[redacted]	[redacted]
Post-progression	[redacted]	[redacted]

Abbreviations: SE, standard error

Table 1: Summary of utility values for cost-effectiveness analysis

- Higher pre-progression utilities in the nivolumab arm were justified by lower serious adverse events in ATTRACTION-3. However, the control arm mean baseline utility was significantly lower than the nivolumab arm ([redacted] versus [redacted]) ERG noted that differences in pre-progression utility could be explained by difference at baseline if people were aware of their treatment allocation at screening visit
- The ERG questioned the use of pre-progression utility with nivolumab that is higher than the mean age-adjusted utility in the UK general population (0.8041 [95% CI: 0.790, 0.817]). It also questioned the use of pre-progression utility which was the

	same as baseline for nivolumab but lower than baseline in the control arm • The ERG considered the large difference in post progression utility ([redacted] for nivolumab vs [redacted] for taxane patients) to lack face validity. In addition, median utility value for nivolumab treated patients was [redacted] from weeks 18 to 54 which is unrealistic given the context of the patient population • The ERG considered a treatment independent approach to be more suitable (see section 4.2.7.3. of the ERG report) • ERG considered the utility vales estimated through mixed-effects regression model which does not use imputed data provided at clarification stage to be appropriate for informing the economic model • ERG conducted an exploratory analysis to determine the effect on ICER of changing utility values, each of the altered utility value sets raised the ICER above £50,000 per QALY	
Alternative extrapolations of time on treatment (TOT)	• The time patient spent on treatment (TOT) was different to the time in the progression-free health state because in ATTRACTION-3 people both discontinued before progression and were able to continue treatment even after progression. • The company used a fully parametric model for both arms in its base case model (generalised gamma model for the nivolumab arm and exponential model for the taxanes arm) • The ERG did not consider fully-parametric models to be a good visual fit to the Kaplan-Meier data • The company provided semi-parametric models with cuts at 4.37 and 5.75 months during clarification • The ERG noted that fully parametric models overestimated TOT in both arms. Although the semi-parametric models are a better fit to the Kaplan-	• The extrapolation of TOT has a greater impact on the ICER than the extrapolation of PFS. This is because TOT is used to determine treatment acquisition and administration costs, which combined are responsible for the majority of incremental costs associated with nivolumab versus taxanes. • The technical team consider a semi-parametric approach with as cut uff at 5.75 years to be most appropriate for extrapolating time on treatment because of an improved visual fit to the Kaplan-Meier data.

<table>
<tr>
<td></td>
<td>
Meier data, a substantial proportion of people (5.69% in the nivolumab arm and 0.51% in the taxane arm) are on treatment at 2 years
• Company sensitivity analysis used a stopping rule at 2 years although there were no stopping rules in ATTRACTION-3 and it is not clear if a stopping rule would be used in clinical practice
• The ERG used a semi-parametric Weibull model for long term extrapolation of the Kaplan-Meier curve to extrapolate TOT with a cut point of 5.75 years without a stopping rule

Table 24 ERG Report: Comparison of company and ERG-preferred ToT extrapolations
<table>
<tr><td></td><td colspan="2">Company</td><td colspan="2">ERG</td></tr>
<tr><td>Description</td><td colspan="2">FP approach using generalised gamma model (both arms)</td><td colspan="2">SP approach using Kaplan–Meier cure until 5.75 months, followed by a Weibull model (both arms)</td></tr>
<tr><td>Time (years)</td><td>Nivolumab</td><td>Taxanes</td><td>Nivolumab</td><td>Taxanes</td></tr>
<tr><td>1</td><td>13.36%</td><td>2.90%</td><td>12.80%</td><td>3.68%</td></tr>
<tr><td>2</td><td>4.19%</td><td>0.08%</td><td>5.69%</td><td>0.51%</td></tr>
<tr><td>3</td><td>1.67%</td><td>0.00%</td><td>2.96%</td><td>0.08%</td></tr>
<tr><td>4</td><td>0.76%</td><td>0.00%</td><td>1.66%</td><td>0.01%</td></tr>
<tr><td>5</td><td>0.38%</td><td>0.00%</td><td>0.98%</td><td>0.00%</td></tr>
</table>
Abbreviations: FP, fully-parametric; KM, Kaplan–Meier; SP, semi-parametric
</td>
<td>
<u>What additional information at engagement would help address the issue?</u>
• It is not clear why a fully-parametric model was used by the company after using semi-parametric for OS and PFS. What was the rationale for this?
• Is the long term extrapolation based on fully-parametric (company) or semi-parametric (ERG) methods most appropriate for estimating time on treatment?
• How long are people likely to remain on treatment with nivolumab in NHS practice?
• Is a stopping rule appropriate? If so, what stopping rule(s) are most relevant for NHS practice?
</td>
</tr>
<tr>
<td>Have the costs of comparator treatment</td>
<td>• The company model used the confidential discounted price of nivolumab and the MIMS</td>
<td>• ERG exploratory sensitivity analysis noted that the market share of taxanes</td>
</tr>
</table>

been appropriately estimated?	(Monthly Index of Medical Specialities) list price of taxanes and subsequent treatment • The ERG stated that eMIT (Electronic Market Information Tool) provides price estimates which are reflective of average price paid by NHS trusts, therefore this should be used as the source for treatment costs • Company's base-case model assumed a 50:50 market share of taxanes between docetaxel and paclitaxel, however the ERG notes that there may be a preference for docetaxel due to its lower frequency of administration • ERG conducted a sensitivity analysis to investigate the impact on ICER by altering the percentage market share of taxanes, the upper and lower bounds fell below the threshold of £50,000 per QALY	does not have a significant impact on ICER. See Section 6.2.4. of ERG report **What additional information at engagement would help address the issue?** • Company to clarify why was MIMS used as a tool to source treatment costs, given that eMIT prices are more reflective of those paid by NHS trusts? • Clinical experts to advise whether in clinical practice docetaxel is preferred over paclitaxel.
Administration and medical resource use costs	• Administration costs for taxanes were higher compared with nivolumab due to the expected time of administration for each treatment • The ERG noted some discrepancy in the costs associated with best supportive care. Radiotherapy is costed as £184.25, however the ERG calculated this as £92.13. Additionally, nerve blocks pain relief was costed as £26.62 whereas the ERG calculated this as £2.66 • Unit costs for outpatient consultation and hospitalisation were calculated by ERG using references cited by the company, which increased the cost of an outpatient consultation from £187.36 to £196.33 and hospitalisation costs from £534.07 to £3379.73 • ERG used alternative medical resource use costs to determine the effect on final ICER	• The ICER is highly sensitive to the unit cost per hospitalisation so most appropriate medical resource use costs must be determined. See Section 6.2.6. of ERG report **What additional information at engagement would help address the issue?** • Are the medical resource costs calculated by the company or ERG most relevant to NHS practice?

Does this technology meet the criteria for end of life?	• NICE criteria for end of life are: treatment is indicated for patients with short life expectancy (normally less than 24 months) and the treatment offers an extension to life (normally of at least 3 months) • Both the company and ERG agreed that the first criterion had been met based on the trial data and base case models (mean OS is 11.48 months in taxane arm of ATTRACTION-3) • For the second criteria, when data is restricted to the observed period there was an extension to life of 2.58 months. However, this was based on the intention to treat population of ATTRACTION-3, a predominantly Asian population known to have better health outcomes. • The modelled overall survival improvements with nivolumab in both the company (7.8 months) and ERG (4.0 months) base-case economic models yielded an extension to life of at least three months. Once again this was based on health outcomes in a predominantely Asian population.	• Limited follow-up data available from ATTRACTION-3 means extrapolation is subject to substantial uncertainty, therefore the estimate of survival gain may not be robust. • Generalisability issues with ATTRACTION-3 may affect the estimated extension of life attributable to nivolumab. A 3 month improvement in overall survival may not be realised in NHS practice. • The improvement in overall survival with nivolumab needs to be considered alongside the increased risk of early death with nivolumab (some patients will not live long enough to benefit from treatment). **What additional information at engagement would help address the issue?** • Is the company or ERG method of survival extrapolation most appropriate? • Could further follow-up data from ATTRACTION-3 be made available to support survival analysis? • Is there a subgroup within ATTRACTION-3 that would best reflect the health outcomes typically seen in NHS practice?

2 Questions for engagement

Clinical effectiveness of nivolumab

1. Is the relative treatment effect of nivolumab compared with taxanes likely to be constant over time?

2. How does nivolumab provide overall survival benefit without improvements in progression-free survival or overall response rate

Is best supportive care a relevant comparator?

3. Is nivolumab expected to completely replace taxane use for the given population?

4. In current NHS practice, what treatment would nivolumab-eligible patients be assigned to?

5. What percentage of patients currently receive docetaxel, paclitaxel and irinotecan in NHS practice?

6. Given the limitations of the indirect treatment comparison, is it appropriate to compare nivolumab with best supportive care in people who are not eligible for further chemotherapy?

Generalisability of ATTRACTION-3 results

7. Given the characteristics of the study population, are the results from the ATTRACTION-3 trial generalisable in a UK decision-making context?

8. Would it be more appropriate to use efficacy data from the rest of world population compared with the intention to treat population to estimate the clinical and cost effectiveness in clinical practice?

Safety data for nivolumab, early deaths higher on nivolumab

9. To what extent does the company's rationale provide an explanation for the early deaths observed in the nivolumab arm?

No adjustment to efficacy for any beneficial effects of third-line therapy

10. In clinical practice, is active third-line therapy administered following progression after the use of a taxane or nivolumab?

11. Should the efficacy of nivolumab be adjusted to account for beneficial effects of third-line therapy?

Subgroup analysis by taxane was not provided

12. Is a comparison of nivolumab with individual taxane or combined taxanes more relevant to NHS practice? Which taxane is most commonly used in the NHS?

13. Are health outcomes expected to be different between ‘docetaxel-preferred’ and ‘paclitaxel-preferred’ populations?

Difference between company and ERG ICER

14. Which modelling methods are most suitable for estimating overall survival and expected time on treatment?

15. Which estimates of treatment, administration and medical resource use costs are most reliable?

Model time horizon (40 years in company base-case)

16. What is the most appropriate time horizon for the economic model?

Alternative extrapolations for overall survival

17. Is a fully-parametric or semi-parametric model most appropriate to predict the long term effectiveness of nivolumab and taxane therapy?

18. If a semi-parametric model is preferred, is 2.99 months an appropriate point to start the extrapolation?

19. Does the overall survival predicted in the company ([redacted] at 10 years) or ERG model (0.2% at 10 years) best represent the likely overall survival in NHS practice?

Exploratory analysis of utility values

20. Are the differences in utility between the nivolumab and taxane arm in the company model clinically plausible?

21. Would a treatment independent approach (e.g. using the same values for pre and post progression across the treatment arms) be more appropriate?

Alternative extrapolations of time on treatment

22. Are fully-parametric or semi-parametric methods most appropriate for estimating time on treatment?

23. How long are people likely to remain on treatment with nivolumab in NHS practice?

24. Is a stopping rule appropriate? If so, what stopping rule(s) are most relevant for NHS practice?

Have the costs of comparator treatment been appropriately estimated?

25. Which source of cost estimates for medical technologies is most reflective of those paid by NHS trusts?

26. Is it appropriate to assume a 50:50 market share of taxanes?

Administration and medical resource use costs

27. Have the most appropriate sources been used to calculate administration and medical resource use costs? If so, which estimates of cost are most reliable?

End of life criteria

28. Is the estimate of extension to life robust? Is an extension to life of at least 3 months expected to be realised in NHS practice?

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Nivolumab for unresectable, advanced oesophageal cancer when standard chemotherapy has failed [ID1249]

This document is the technical report for this appraisal. It has been prepared by the NICE technical team.

The technical report and stakeholder's responses to it are used by the appraisal committee to help it make decisions at the appraisal committee meeting. Usually, only unresolved or uncertain key issues will be discussed at the appraisal committee meeting.

This report is based on:

- the evidence and views submitted by the company, consultees and their nominated clinical experts and patient experts and
- the evidence review group (ERG) report.

The technical report should be read with the full supporting documents for this appraisal.

1 Key issues summary

Issue	Summary	Technical Team Preliminary Judgement
Issues related to clinical evidence		
Clinical effectiveness of nivolumab	Most of the overall survival benefit from nivolumab in ATTRACTION-3 is in the post progression phase, as there is no progression free survival benefit associated with nivolumab (Kaplan-Meier estimate of median PFS was 1.68 months in nivolumab group and 3.35 months in the control group). Furthermore, there is little difference in the overall response rate between nivolumab and taxane therapy (19.3% versus 21.5% with an odds ratio of [redacted] It is not clear how nivolumab is achieving a benefit in overall survival without a benefit in progression-free survival or the overall response rate. What rationale is there for this treatment prolonging post progression survival? Patients on nivolumab treatment also need to survive beyond 3 months (during which they have an increased risk of death) before they can benefit from improved overall survival. To what extent is this observed in other cancer treatments? The company have used a Cox proportional hazards model to estimate hazard ratios and 95% confidence intervals for overall survival and progression free survival. The ERG note that the proportional hazard assumption was violated (the two treatment curves crossed for both overall survival and progression-free survival). See figures 10 and 11 of the company submission, document B below).	The methods used to estimate the efficacy of nivolumab compared with taxanes assume constant relative efficacy of nivolumab compared with taxanes over time. If the proportional hazards assumption does not hold, the effectiveness of nivolumab in the model may be overestimated **<u>What additional information at engagement would help address the issue?</u>** • How is nivolumab improving overall survival without improvements in progression-free survival or the overall response rate? Is this observed with other cancer treatments? • Clinical advice would be useful on whether the relative treatment effect of nivolumab compared with taxanes is likely to be constant over time. • Is the risk of 'early death' with initial nivolumab treatment worth an additional 2.58 months overall survival reported in the trial?

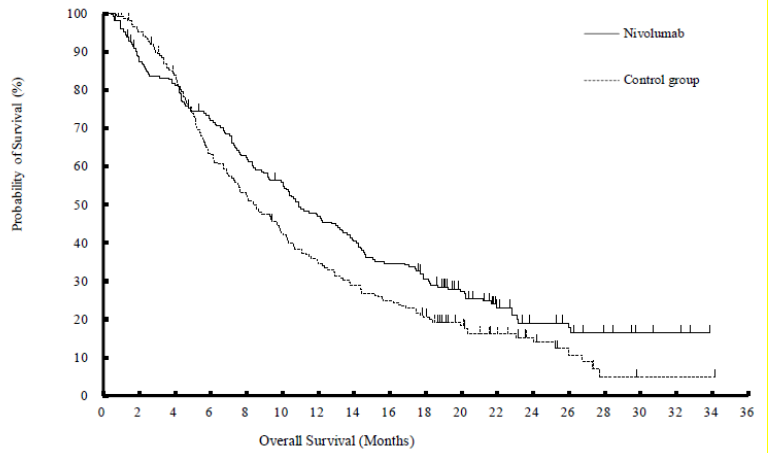


Figure 1. Kaplan-Meier plot of overall survival in patients receiving nivolumab or taxane

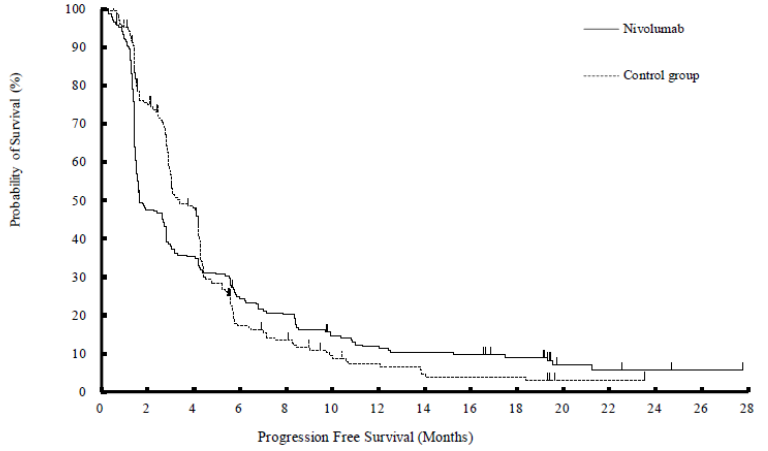


Figure 2. Kaplan-Meier plot of progression-free survival in patients receiving nivolumab or taxane

Is best supportive care a relevant comparator?	• Direct comparative evidence for nivolumab vs taxane chemotherapy was available in ATTRACTION-3 • The company and clinical advisors to the ERG noted that irinotecan is not a relevant comparator because it is only used in 6% of patients • The ERG did not consider comparisons of nivolumab with individual taxanes to be relevant because patients were not randomised to taxane treatment, rather this was chosen by the investigator • Best supportive care was included as a comparator in the final scope. Because there is no direct trial evidence comparing nivolumab with best supportive care, the company provided an exploratory indirect treatment comparison (ITC). However, it did consider this analysis to be robust. • The ERG noted several flaws in the ITC including studies that were not randomised, lack of adjustment for differences in baseline characteristics (including ECOG performance status), inconsistent measurement of outcomes (e.g. post progression survival used to approximate overall survival) • The ERG also questioned the generalisibility of the ITC results to NHS practice because all of the trials in the network (except ATTRACTION-3) included Japanese-only patients, who have better survival outcomes • The company note the unmet need in people for whom taxane therapy is unsuitable but have not proposed or asked NICE to consider the clinical and cost effectiveness of nivolumab in this population	The ERG and technical team do not consider best supportive care to be a relevant comparator because the pivotal trial was conducted in taxane-eligible patients. The ITC used to derive survival data for best supportive care contained several flaws, which make results from comparisons with best supportive care highly uncertain (see Section 3.3. of the ERG report) **<u>What additional information at engagement would help address the issue?</u>** • Clinical expert advice on the most relevant comparator for NHS practice. • Is irinotecan used in clinical practice for patients who would be eligible for nivolumab? • What proportion of people currently receive docetaxel or paclitaxel in NHS practice? • Is best supportive care a relevant comparator? Can the efficacy of nivolumab in ATTRACTION-3 be generalised to people for whom best supportive care is suitable? Is the ITC robust given the heterogeneity between studies? • Are the results of the ITC generalisable to NHS practice?
Generalisability of ATTRACTION-3 participants to UK population	• 96% of ATTRACTION-3 patients were Asian, two-thirds of which were Japanese • Treatment in Asia follows pan-Asian adapted European Society for Medical Oncology guidelines	• Differences in the treatment pathway and improved health outcomes for patients in Asia mean that the clinical effectiveness in the intention to treat

	which recommend treatment options that are not available in Europe (for example nedaplatin) • Nivolumab is associated with longer overall survival compared with taxane therapy (10.91 vs 8.38 months, HR 0.77 [95% CI: 0.62, 0.96] P <0.0001) in the intention to treat population. • Clinical advisers to the ERG noted that this is clinically meaningful. • The relative efficacy of nivolumab compared with taxanes in the Japanese population (hazard ratio [redacted]) is similar to the rest of the world ([redacted]). However, the absolute overall survival in Japanese patients is higher for both taxanes ([redacted] compared with [redacted]) and nivolumab treatment ([redacted] compared with [redacted]). Overall survival with taxanes in the Japanese population is longer than overall survival with nivolumab in the rest of the world population ([redacted] compared with [redacted]). • Efficacy data in the model was based on the hazard ratio in the intention to treat population from ATTRACTION-3. • Efficacy data based on rest of world data only may be more relevant to NHS practice (although this also included other populations such as Korean) • Only patients with Easten Cooperative Oncology Group (ECOG) performance scores 0-1 were included in the trial • The company also highlighted that individuals randomised into clinical trials are likely to be slightly younger and healthier than the overall oesophageal cancer patient population in the UK	population is not likely to be realised in NHS practice • The efficacy of nivolumab vs taxanes in the rest of world population may be more suitable to estimate the effectiveness of nivolumab in the UK context, provided that the rest of world population is more reflective of patients seen in NHS practice • Participants in ATTRACTION-3 are also fitter and more able to tolerate treatment with nivolumab and taxanes than people in UK practice **<u>What additional information at engagement would help address the issue?</u>** • Clinical expert clarification on whether the rest of the world efficacy data is more relevant to the expected effectiveness of nivolumab in the NHS practice • Who is likely to receive nivolumab in clinical practice? Will it be restricted to people with a good performance status (ECOG 0-1)? Is the efficacy of nivolumab in NHS practice (people with a worse performance status) likely to be worse than in the ATTRACTION-3 ? • Is there a subgroup within ATTRACTION-3 which is considered to have similar characteristics to the UK patient population?

Safety data for nivolumab, deaths higher on nivolumab in first 3 months	• The ERG noted that the safety profile of nivolumab is favourable compared to taxanes. However, the number of deaths in the first 3 months of treatment was higher in the nivolumab compared with the taxane arm ([redacted]) • During clarification, the company explained that this effect may relate to the differences in mechanism of action of immunotherapy treatments compared to chemotherapy agents. In particular, that immunotherapies are associated with a longer time to response which is seen with PDL-1 inhibitors.	• The ERG noted that the rationale provided by the company (attributable to its mechanism of action) does not adequately explain the reason for higher deaths on nivolumab compared with taxane therapy in first 3 months. **<u>What additional information at engagement would help address the issue?</u>** • Do clinical experts agree with the company's rationale for higher death rate over first 3 months in the nivolumab arm? • Clinical expert advice is sought on whether the initial higher death rate with nivolumab in ATTRACTION-3 is likely to be seen in NHS practice. • Is it possible to determine in advance which patients are likely to die before they can benefit from treatment with nivolumab (if so please explain)?.
No adjustment to efficacy for any beneficial effects of third-line therapy	• A single overarching OS curve was used to estimate number of deaths per model cycle. The proportion of deaths occurring for patients still on treatment vs off treatment could not be separately calculated • The company model base case assumed that no active therapy is given 3rd line. • The ERG suggests that some of the overall survival benefit in the trial is related to the active 3rd line treatment which is not balanced between the two arms. A higher percentage of people in nivolumab arm received active treatment after progression ([redacted] of the nivolumab arm received treatment after progression for a median of [redacted], versus [redacted] of taxanes patients for a median of one day). Benefits	• ATTRACTION-3 allowed active subsequent treatment after progression which contributes to the improved overall survival of nivolumab compared with taxanes • In NHS practice patients are unlikely to receive active treatment at this stage in the treatment pathway. The company did not adjust the efficacy of nivolumab for the efficacy of 3rd line treatment. Therefore, efficacy data which has been used in the economic model is likely to overestimate the benefits of nivolumab

	associated with nivolumab may therefore be over-estimated. • The ERG note that no adjustment to efficacy has been made for any beneficial effects of active third-line therapy • The ERG questioned whether the assumption that after progression people would receive active 3rd line treatment until death is relevant to NHS practice	**What additional information at engagement would help address the issue?** • Clinical experts to clarify whether patients in NHS practice receive active treatment after progression on nivolumab or taxanes. What proportion of patients receive different 3rd line treatments? • Is nivolumab expected to completely replace taxane use for the proposed indication and population? If so, would taxane therapy then be offered to nivolumab patients post-progression? • Should the efficacy of nivolumab in ATTRACTION-3 be adjusted to remove the impact of active subsequent treatment?
Subgroup analysis by taxane was not provided	• The comparator arm in ATTRACTION-3 consisted of either docetaxel or paclitaxel (investigators choice which was recorded pre-randomisation). • Following clarification the company noted that because the choice of taxanes was recorded pre-randomisation this was not a stratification factor, therefore subgroup analyses by taxane would not be provided. • The ERG noted that a potentially more robust comparison of nivolumab vs each taxane could have been made available to support decision analytic modelling. This would allow comparisons to consider the effectiveness of nivolumab in 'docetaxel-preferred' or 'paclitaxel-preferred' populations	**What additional information at engagement would help address the issue?** Clinical advice on the following is required: • Is it reasonable to assume a class effect for taxane therapy? • Are there systematic differences in people who would be suitable for treatment with either docetaxel or paclitaxel? How should these separate populations be defined? • Could post-hoc analysis of the effectiveness of efficacy of nivolumab compared with either docetaxel or paclitaxel have been carried out? • Is a comparison of nivolumab with individual taxanes more relevant to NHS practice than comparisons with a combined taxane arm?

Issues related to cost-effectiveness evidence		
Is the model time horizon (40 years) appropriate?	• The majority of patients undergoing second-line taxane therapy in current UK practice would have an estimated survival of less than 12 months • The ERG consider extrapolated survival benefits of nivolumab over a 40 year time horizon may give unrealistic mean life years gained depending on the survival models used (ERG report Section 4.2.6.) • The ERG advised that a 10-year time horizon should be long enough to capture differences in the costs and benefits	**What additional information at engagement would help address the issue?** • Clinical expert advice is required on the life expectancy of patients with unresectable, advanced oesophageal cancer treated with nivolumab or taxanes • What is the most relevant time horizon for the economic model?
Alternative extrapolations for overall survival	• Kaplan-Meier estimate of median OS in the intention to treat population was 10.91 months in the nivolumab group and 8.38 months in the control group. The OS rates were reported to be numerically higher in the nivolumab group from follow-up at month 6 through to month 30. • Survival modelling methods were used to extrapolate over the lifetime horizon of the model • The company used a semi-parametric model using Kaplan-Meier data to 2.99 months then parametric model (log-logistic model in nivolumab arm and exponential in the taxane arm) for the remainder of the time horizon. The rationale for this was that a semi-parametric model could take into account changes in hazard over time (that is high hazard associated with the early death rate on nivolumab and lower hazard over the longer term. • The ERG agreed that a fully parametric model may not be suitable because of the initial early deaths in the trial It noted that although there is a change in hazard in the nivolumab arm at approximately 3 months there is no change in hazard in the taxane arm.	**What additional information at engagement would help address the issue?** • Estimation of OS is a key driver of cost-effectiveness results • Is a fully-parametric or semi-parametric model most appropriate to predict the long term effectiveness of nivolumab and taxane therapy? • If a semi-parametric model is preferred, is 2.99 or 5.75 months an appropriate point to start the extrapolation? • Does the overall survival predicted in the company (1.92% at 10 years) or ERG model (0.2% at 10 years) best represent the likely overall survival in NHS practice?

- The company provided alternative analyses in response to a clarification request from the ERG with cut offs of 4.37 (before any deaths) and 5.75 months (before the arms crossed over). See response to question B10 of the clarification response and Figure 3 in the ERG report.
- The ERG noted that the long term predictions from fully or semi-parametric models may be poor because they may not accurately reflect hazards over time. It further noted that the plausibility of longer-term extrapolations should be carefully considered.
- The ERG base case model used a semi-parametric approach using Kaplan-Meier curve until 5.75 months, followed by a generalised gamma model (both arms) for OS extrapolations because they were a better visual fit.
- 10-year OS in the company's base-case was estimated to be 1.92% for the nivolumab arm, versus 0.20% for the ERG's base-case analysis.

Table 23 ERG report:Comparison of company and ERG-preferred OS extrapolations

	Company		**ERG**	
Description	SP approach using Kaplan–Meier cure until 2.99 months, followed by a log-logistic (nivolumab) or exponential (taxanes) model		SP approach using Kaplan–Meier cure until 5.75 months, followed by a generalised gamma model (both arms)	
Time (years)	**Nivolumab**	**Taxanes**	**Nivolumab**	**Taxanes**
1	45.61%	36.57%	46.07%	35.40%

2	21.27%	11.06%	20.70%	12.20%
3	12.33%	3.34%	10.22%	4.42%
4	8.15%	1.01%	5.36%	1.65%
5	5.84%	0.30%	2.93%	0.63%
6	4.39%	0.09%	1.63%	0.24%
8	2.78%	0.01%	0.56%	0.04%
10	1.92%	0.00%	0.20%	0.01%

Abbreviations: KM, Kaplan–Meier; SP, semi-parametric

Exploratory analysis of utility values

- Company estimated utility values were based on EQ-5D data from ATTRACTION-3
- The base case utilities in the company model are:

Health state	Nivolumab utility: mean (SE)	Control utility: mean (SE)
Pre-progression	[redacted]	[redacted]
Post-progression	[redacted]	[redacted]

Abbreviations: SE, standard error

Table 1: Summary of utility values for cost-effectiveness analysis

- Higher pre-progression utilities in the nivolumab arm were justified by lower serious adverse events in ATTRACTION-3. However, the control arm mean baseline utility was significantly lower than the nivolumab arm ([redacted] versus [redacted]) ERG noted that differences in pre-progression utility could be explained by difference at baseline if people were aware of their treatment allocation at screening visit
- The ERG questioned the use of pre-progression utility with nivolumab that is higher than the mean

The technical team considers the following:

- It is not appropriate to set a pre-progression utility which exceeds the age-adjusted population mean
- The magnitude of differences in utility between the nivolumab and taxane arm in the company model are not clinically plausible
- A treatment independent approach (e.g. using the same values for pre and post progression across the treatment arms) is more appropriate for the estimation of utility.

	age-adjusted utility in the UK general population (0.8041 [95% CI: 0.790, 0.817]). It also questioned the use of pre-progression utility which was the same as baseline for nivolumab but lower than baseline in the control arm • The ERG considered the large difference in post progression utility ([redacted] for nivolumab vs [redacted] for taxane patients) to lack face validity. In addition, median utility value for nivolumab treated patients was [redacted] from weeks 18 to 54 which is unrealistic given the context of the patient population • The ERG considered a treatment independent approach to be more suitable (see section 4.2.7.3. of the ERG report) • ERG considered the utility vales estimated through mixed-effects regression model which does not use imputed data provided at clarification stage to be appropriate for informing the economic model • ERG conducted an exploratory analysis to determine the effect on ICER of changing utility values, each of the altered utility value sets raised the ICER above £50,000 per QALY	
Alternative extrapolations of time on treatment (TOT)	• The time patient spent on treatment (TOT) was different to the time in the progression-free health state because in ATTRACTION-3 people both discontinued before progression and were able to continue treatment even after progression. • The company used a fully parametric model for both arms in its base case model (generalised gamma model for the nivolumab arm and exponential model for the taxanes arm) • The ERG did not consider fully-parametric models to be a good visual fit to the Kaplan-Meier data • The company provided semi-parametric models with cuts at 4.37 and 5.75 months during clarification	• The extrapolation of TOT has a greater impact on the ICER than the extrapolation of PFS. This is because TOT is used to determine treatment acquisition and administration costs, which combined are responsible for the majority of incremental costs associated with nivolumab versus taxanes. • The technical team consider a semi-parametric approach with as cut uff at 5.75 years to be most appropriate for extrapolating time on treatment because

- The ERG noted that fully parametric models overestimated TOT in both arms. Although the semi-parametric models are a better fit to the Kaplan-Meier data, a substantial proportion of people (5.69% in the nivolumab arm and 0.51% in the taxane arm) are on treatment at 2 years
- Company sensitivity analysis used a stopping rule at 2 years although there were no stopping rules in ATTRACTION-3 and it is not clear if a stopping rule would be used in clinical practice
- The ERG used a semi-parametric Weibull model for long term extrapolation of the Kaplan-Meier curve to extrapolate TOT with a cut point of 5.75 years without a stopping rule

Table 24 ERG Report: Comparison of company and ERG-preferred ToT extrapolations

	Company		**ERG**	
Description	FP approach using generalised gamma model (both arms)		SP approach using Kaplan–Meier cure until 5.75 months, followed by a Weibull model (both arms)	
Time (years)	**Nivolumab**	**Taxanes**	**Nivolumab**	**Taxanes**
1	13.36%	2.90%	12.80%	3.68%
2	4.19%	0.08%	5.69%	0.51%
3	1.67%	0.00%	2.96%	0.08%
4	0.76%	0.00%	1.66%	0.01%
5	0.38%	0.00%	0.98%	0.00%

Abbreviations: FP, fully-parametric; KM, Kaplan–Meier; SP, semi-parametric

of an improved visual fit to the Kaplan-Meier data.

What additional information at engagement would help address the issue?

- It is not clear why a fully-parametric model was used by the company after using semi-parametric for OS and PFS. What was the rationale for this?
- Is the long term extrapolation based on fully-parametric (company) or semi-parametric (ERG) methods most appropriate for estimating time on treatment?
- How long are people likely to remain on treatment with nivolumab in NHS practice?
- Is a stopping rule appropriate? If so, what stopping rule(s) are most relevant for NHS practice?

Have the costs of comparator treatment been appropriately estimated?	• The company model used the confidential discounted price of nivolumab and the MIMS (Monthly Index of Medical Specialities) list price of taxanes and subsequent treatment • The ERG stated that eMIT (Electronic Market Information Tool) provides price estimates which are reflective of average price paid by NHS trusts, therefore this should be used as the source for treatment costs • Company's base-case model assumed a 50:50 market share of taxanes between docetaxel and paclitaxel, however the ERG notes that there may be a preference for docetaxel due to its lower frequency of administration • ERG conducted a sensitivity analysis to investigate the impact on ICER by altering the percentage market share of taxanes, the upper and lower bounds fell below the threshold of £50,000 per QALY	• ERG exploratory sensitivity analysis noted that the market share of taxanes does not have a significant impact on ICER. See Section 6.2.4. of ERG report **What additional information at engagement would help address the issue?** • Company to clarify why was MIMS used as a tool to source treatment costs, given that eMIT prices are more reflective of those paid by NHS trusts? • Clinical experts to advise whether in clinical practice docetaxel is preferred over paclitaxel.
Administration and medical resource use costs	• Administration costs for taxanes were higher compared with nivolumab due to the expected time of administration for each treatment • The ERG noted some discrepancy in the costs associated with best supportive care. Radiotherapy is costed as £184.25, however the ERG calculated this as £92.13. Additionally, nerve blocks pain relief was costed as £26.62 whereas the ERG calculated this as £2.66 • Unit costs for outpatient consultation and hospitalisation were calculated by ERG using references cited by the company, which increased the cost of an outpatient consultation from £187.36 to £196.33 and hospitalisation costs from £534.07 to £3379.73	• The ICER is highly sensitive to the unit cost per hospitalisation so most appropriate medical resource use costs must be determined. See Section 6.2.6. of ERG report **What additional information at engagement would help address the issue?** • Are the medical resource costs calculated by the company or ERG most relevant to NHS practice?

	• ERG used alternative medical resource use costs to determine the effect on final ICER	
Does this technology meet the criteria for end of life?	• NICE criteria for end of life are: treatment is indicated for patients with short life expectancy (normally less than 24 months) and the treatment offers an extension to life (normally of at least 3 months) • Both the company and ERG agreed that the first criterion had been met based on the trial data and base case models (mean OS is 11.48 months in taxane arm of ATTRACTION-3) • For the second criteria, when data is restricted to the observed period there was an extension to life of 2.58 months. However, this was based on the intention to treat population of ATTRACTION-3, a predominantly Asian population known to have better health outcomes. • The modelled overall survival improvements with nivolumab in both the company (7.8 months) and ERG (4.0 months) base-case economic models yielded an extension to life of at least three months. Once again this was based on health outcomes in a predominantely Asian population.	• Limited follow-up data available from ATTRACTION-3 means extrapolation is subject to substantial uncertainty, therefore the estimate of survival gain may not be robust. • Generalisability issues with ATTRACTION-3 may affect the estimated extension of life attributable to nivolumab. A 3 month improvement in overall survival may not be realised in NHS practice. • The improvement in overall survival with nivolumab needs to be considered alongside the increased risk of early death with nivolumab (some patients will not live long enough to benefit from treatment). **<u>What additional information at engagement would help address the issue?</u>** • Is the company or ERG method of survival extrapolation most appropriate? • Could further follow-up data from ATTRACTION-3 be made available to support survival analysis? • Is there a subgroup within ATTRACTION-3 that would best reflect the health outcomes typically seen in NHS practice?

2 Questions for engagement

Clinical effectiveness of nivolumab

1. **Is the relative treatment effect of nivolumab compared with taxanes likely to be constant over time?**

This is unlikely to be true – based on observed data, the hazard profiles for OS and PFS diverge over time for nivolumab versus taxanes. Hence, it can be extrapolated that this will continue to diverge over time.

It should also be noted that this is in line with other studies for immuno-oncology agents, including nivolumab, wherein hazard profiles diverge over time. Further data are provided in response to Question 2.

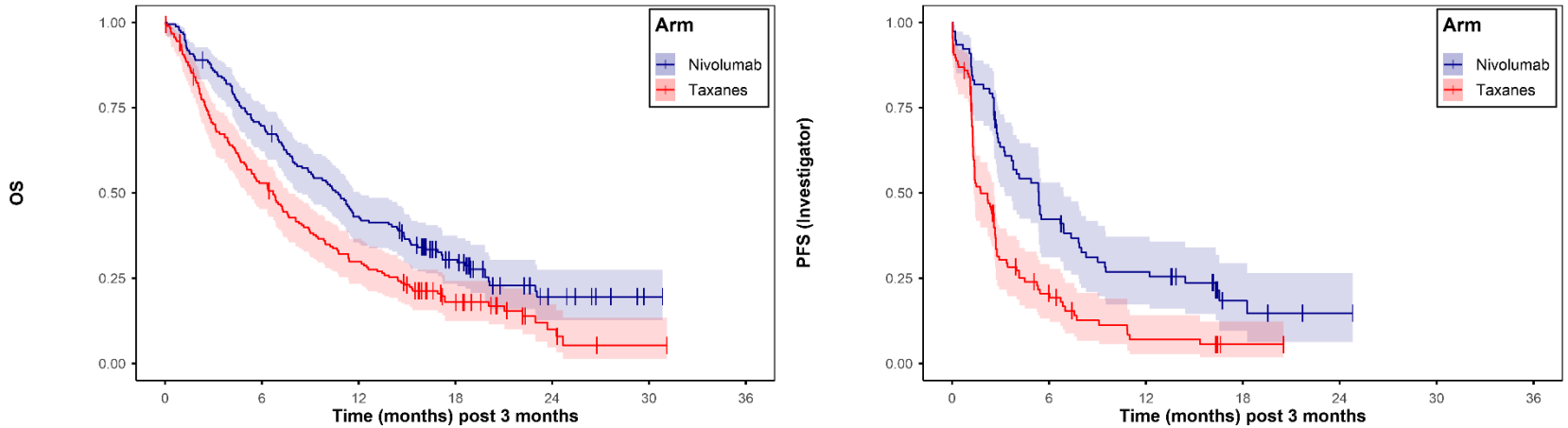
2. **How does nivolumab provide overall survival benefit without improvements in progression-free survival or overall response rate**

    a. **How is nivolumab improving overall survival without improvements in progression-free survival or the overall response rate? Is this observed with other cancer treatments?**

As noted previously, immunotherapies such as nivolumab have a different mechanism of action than conventional anti-cancer therapies, which typically aim to reduce the tumour burden through direct disruption of tumour cell proliferation or induction of apoptosis. By contrast, immunotherapy agents such as nivolumab, often have a delayed clinical responses[1] and differences in response patterns after immunotherapy may potentially be prematurely misclassified as disease progression under the WHO or RECIST criteria.[1, 2] For the same reasons, PFS may not be an adequate endpoint in immunotherapy trials and may not be considered a surrogate for OS for the achievement of clinical efficacy.

Additionally, it is incorrect to say that there are no improvements in PFS; while median PFS is lower for nivolumab than for taxanes, PFS rates from 6 months are improved for the nivolumab arm, as shown by the Kaplan-Meier data (PFS rates were notably higher in the nivolumab group than the taxane group at 6 months [24.2% vs 17.2%], 12 months [11.9% vs 7.2%], and 18 months [9.0% vs 4.0%]).[3] Hence, it is inaccurate to say that there is no improvement in PFS. A similar profile is observed in the OS Kaplan-Meier data, wherein there is initial crossover but median OS and OS rates from 6 months to end of follow up show a beneficial impact for nivolumab therapy versus taxanes (OS rates were notably higher in the nivolumab group than the taxane group at 12 months [46.9% vs 34.4%] and 30 months [[redacted]%]).[3]

In further support of this evidence, a landmark analysis of ATTRACTION-3 evidence was undertaken based on patients alive at three months. As can be seen in Figure 3, for patients who remain alive at three months, outcomes are vastly improved for those in the nivolumab arm for both PFS and OS, which remain significantly higher across the observed data.



NAR (Cumulative Events)

Nivolumab	172 (0)	119 (52)	73 (97)	38 (117)	8 (126)	1 (126)	0 (126)
Taxanes	183 (0)	95 (85)	53 (126)	21 (145)	5 (150)	1 (152)	0 (152)

NAR (Cumulative Events)

Nivolumab	77 (0)	32 (44)	19 (55)	5 (59)	1 (60)	0 (60)	0 (60)
Taxanes	107 (0)	18 (80)	5 (90)	1 (91)	0 (91)	0 (91)	0 (91)

Figure 3. ATTRACTION-3 landmark analysis based on patients alive at three months

For this reason, in the oesophageal cancer setting, where there is short life expectancy and poor prognosis, Kaplan-Meier curves for patients receiving nivolumab monotherapy often demonstrate a high initial hazard, followed by decreasing hazard over time. By contrast, Kaplan-Meier data describing patients receiving conventional chemotherapies have a lower initial hazard followed by increasing hazard over time. This is demonstrated when the hazard rate is plotted against time for both arms from ATTRACTION-3 (Figure 4). The hazard in the taxane arm is initially lower than in the nivolumab arm and increases over time. The hazard for taxanes remains higher than for nivolumab throughout the observed period and demonstrates a decrease towards the end, where this is informed by fewer patients. In contrast, the nivolumab arm

demonstrates a higher initial hazard that flattens and decreases gradually over time as it would be expected. Kernel based methods and B-spline were modelled and display broadly similar findings, variation between these is due to the underlying methods employed. Specifically, Kernel based smoothing can result in distortion at boundaries which results in apparent deviation from the B-spline estimates. Analysis of the hazard rates observed in ATTRACTION-3 are therefore in line with clinical knowledge about how the mechanism of action may result in a slightly delayed response from immunotherapies and apparently limited impact on progression free survival.

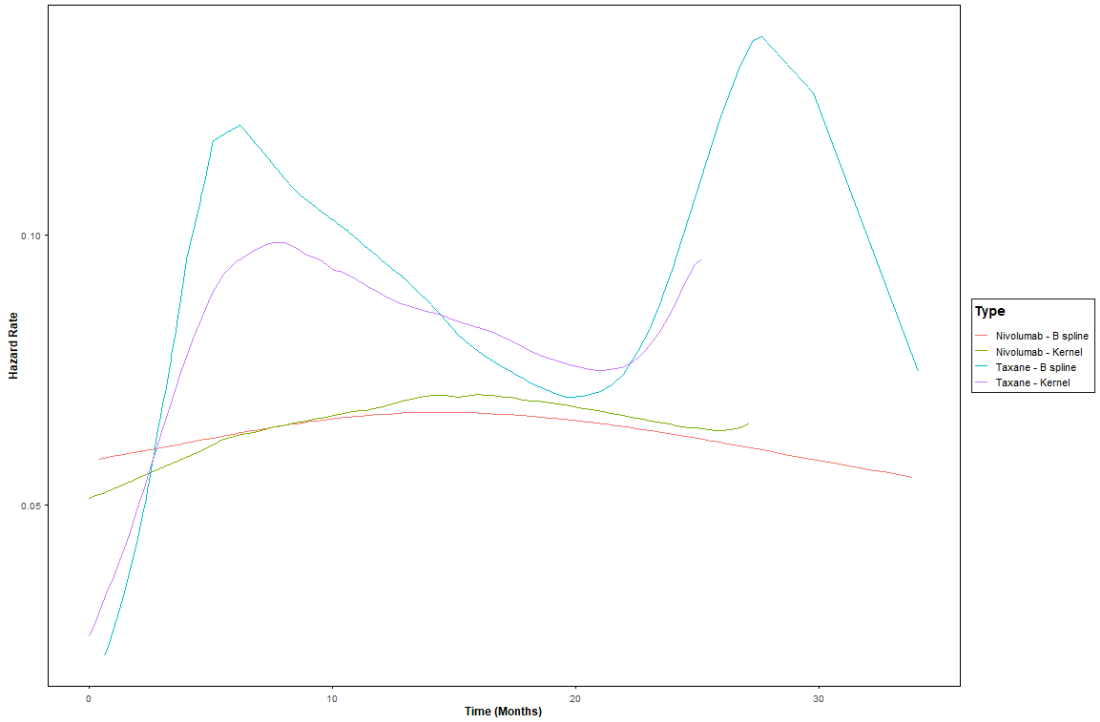


Figure 4: Overall Survival Hazard Rate for Nivolumab and Taxane arms in ATTRACTION-3

It should be noted that this benefit profile is comparable to that observed for all immuno-oncology therapies assessed in indications where survival is short and evidence is versus an active comparator. A short summary of available evidence is provided below, focusing on trials where taxanes are relevant comparators:

- Nivolumab studies (summarised in Table 2):
    - **CheckMate 017** and **CheckMate 057** assessed nivolumab versus docetaxel in patients with previously treated non-small cell lung cancer with squamous and non-squamous histology, respectively.[4] Long-term survival outcomes are provided in Figure 4. Similar to ATTRACTION-3, median PFS in the pooled analysis was lower in the nivolumab arm than the docetaxel arm (2.56 months versus 3.52 months). However, this reflected crossover in the Kaplan-Meier data, so that PFS at all subsequent timepoints was higher for nivolumab compared with docetaxel (one year: 20% versus 9%; two years: 13% versus 2%; three years: 10% versus <1%; five years: 8% versus 0). Reflecting this benefit, median OS was higher for nivolumab than docetaxel (11.1 months versus 8.1 months); outcomes were also increased for all subsequent timepoints (one year: 48% versus 34%; two years: 27% versus 14%; three years: 17% versus 8%, five years: 13.4% versus 2.6%), providing long-term evidence indicating that this benefit was sustained over time.[4, 5]
    - **CheckMate 141** assessed nivolumab versus standard therapy (methotrexate, docetaxel, or cetuximab) for the treatment of patients with recurrent squamous-cell carcinoma of the head and neck whose disease had progressed within 6 months after platinum-based chemotherapy. Long-term outcomes are provided in Figure 6 and Figure 7. No significant difference was observed in terms of median PFS (2.0 months for nivolumab versus 2.3 months for standard therapy); however, a late separation in the Kaplan–Meier curves was observed, and the estimated rates of PFS at 6 months were 19.7% in the nivolumab group and 9.9% in the standard therapy group.[6] Following a minimum follow-up of 24.2 months, median OS was 7.7 months in the nivolumab arm versus 5.1 months in the standard therapy arm. The Kaplan-Meier–estimated 24-month OS rate with nivolumab was 16.9% versus 6.0% for standard therapy and estimated 48-month OS rate was [redacted] versus [redacted] for nivolumab versus standard therapy, respectively. This evidence supports prolonged OS benefit despite initial PFS crossover.[6-9]
- Other immuno-oncology therapies:
    - **KEYNOTE-010** assessed pembrolizumab (2 mg/kg or 10 mg /kg) versus docetaxel for patients with previously treated, PD-L1-positive, advanced non-small-cell lung cancer. Median PFS was comparable between arms (3·9 months with pembrolizumab 2 mg/kg, 4·0 months with pembrolizumab 10 mg/kg, and 4·0 months with docetaxel). However, median OS was extended in the

pembrolizumab arms (10·4 months with pembrolizumab 2 mg/kg, 12·7 months with pembrolizumab 10 mg/kg, and 8·5 months with docetaxel).[10]

- **OAK** assessed atezolizumab versus docetaxel for patients with previously treated NSCLC. Median PFS was lower for atezolizumab versus docetaxel (2·8 months versus 4·0 months). However, median OS was significantly improved (13.8 months versus 9.6 months), while OS at 12 months for atezolizumab and docetaxel were 55% versus 41%, respectively, and 40% versus 27% at 18 months, indicating that this benefit was maintained.[11]

- **KEYNOTE-045** assessed pembrolizumab versus chemotherapy (paclitaxel, docetaxel, or vinflunine) for treatment of advanced urothelial carcinoma that progresses after platinum-based chemotherapy. Similar to other immuno-oncology studies, median PFS was lower for pembrolizumab (2.1 months versus 3.3 months); however, one-year PFS was 16.8% for pembrolizumab versus 6.2% for chemotherapy, demonstrating the impact of the changing hazard profile on outcomes. As can be expected, similar crossover in Kaplan-Meier data is observed for OS, but outcomes remain improved for the pembrolizumab arm. Median OS was 10.3 months in pembrolizumab arm versus 7.4 months for chemotherapy arm, while OS at one year was 43.9% for pembrolizumab versus 30.7% in the chemotherapy group.[12]

- **KEYNOTE-181** assessed pembrolizumab versus chemotherapy in previously treated patients with advanced/metastatic adenocarcinoma or squamous cell carcinoma (SCC) of the oesophagus. Similar to ATTRACTION-3, median PFS was lower for pembrolizumab in SCC patients (2.2 months versus 3.1 months; ITT population: 2.1 months versus 3.4 months).[13] However, outcomes in the pembrolizumab arm were improved for SCC patients in terms of median OS (8.2 months versus 7.1 months; ITT population: 7.1 months versus 7.1 months) and one-year OS (39% versus 25%; ITT population: 32% versus 24%).[14]

- **KEYNOTE-040** assessed pembrolizumab versus standard therapy (methotrexate, docetaxel, or cetuximab) for treatment of recurrent or metastatic head and neck SCC. Median PFS was comparable between the treatment arms (2.1 months for pembrolizumab versus 2.3 months for standard therapy) while median OS reflected improved outcomes for pembrolizumab (8.4 months versus 6.9 months).[15, 16]

Hence, it should be noted that OS benefit with limited median PFS benefit is commonly observed across immunotherapies. However, studies with longer follow-up demonstrate that this benefit is maintained long term.

Table 2. Comparison of outcomes for nivolumab in active-controlled studies in diseases with short survival

		Previously treated OSCC			Previously treated NSCLC			Previously treated HNSCC		
		ATTRACTION-3			CheckMate 017/057			CheckMate 141		
		Nivolumab	Taxane	Source	Nivolumab	Docetaxel	Source	Nivolumab	Standard therapy	Source
PFS	Median	1.68	3.35	Kato 2019[17]	2.56	3.52	Vokes 2018[4], Gettinger 2019[5]	2	2.3	Ferris 2016[6]
	Six months	24.2	17.2		NR	NR		19.7	9.9	
	One year	11.9	7.2		20	9		NR	NR	
	Two years	[redacted]	[redacted]	ATTRACTION -3 PLD	13	2		NR	NR	
	Three years	NR	NR		10	<1		NR	NR	
	Five years	NR	NR		8	0		NR	NR	
OS	Median	10.91	8.38	Kato 2019[17]	11.1	8.1		7.7	5.1	Yen 2020[9]; Ferris 2018[7], BMS 2019 [8]
	One year	46.9	34.4		48	34		33.6	19.8	
	Two years	[redacted]	[redacted]	ATTRACTION -3 CSR and PLD	27	14		16.9	6	
	Three years	NR	NR		17	8		NR	NR	
	Four years	NR	NR		NR	NR		8.0	1.7	
	Five years	NR	NR		13.4	2.6		NR	NR	

CSR: clinical study report; HNSCC: head and neck squamous cell carcinoma; NR: not reported; NSCLC: non-small cell lung cancer; OS: overall survival; OSCC: oesophageal squamous cell carcinoma; PLD: patient-level data; PFS: progression-free survival.

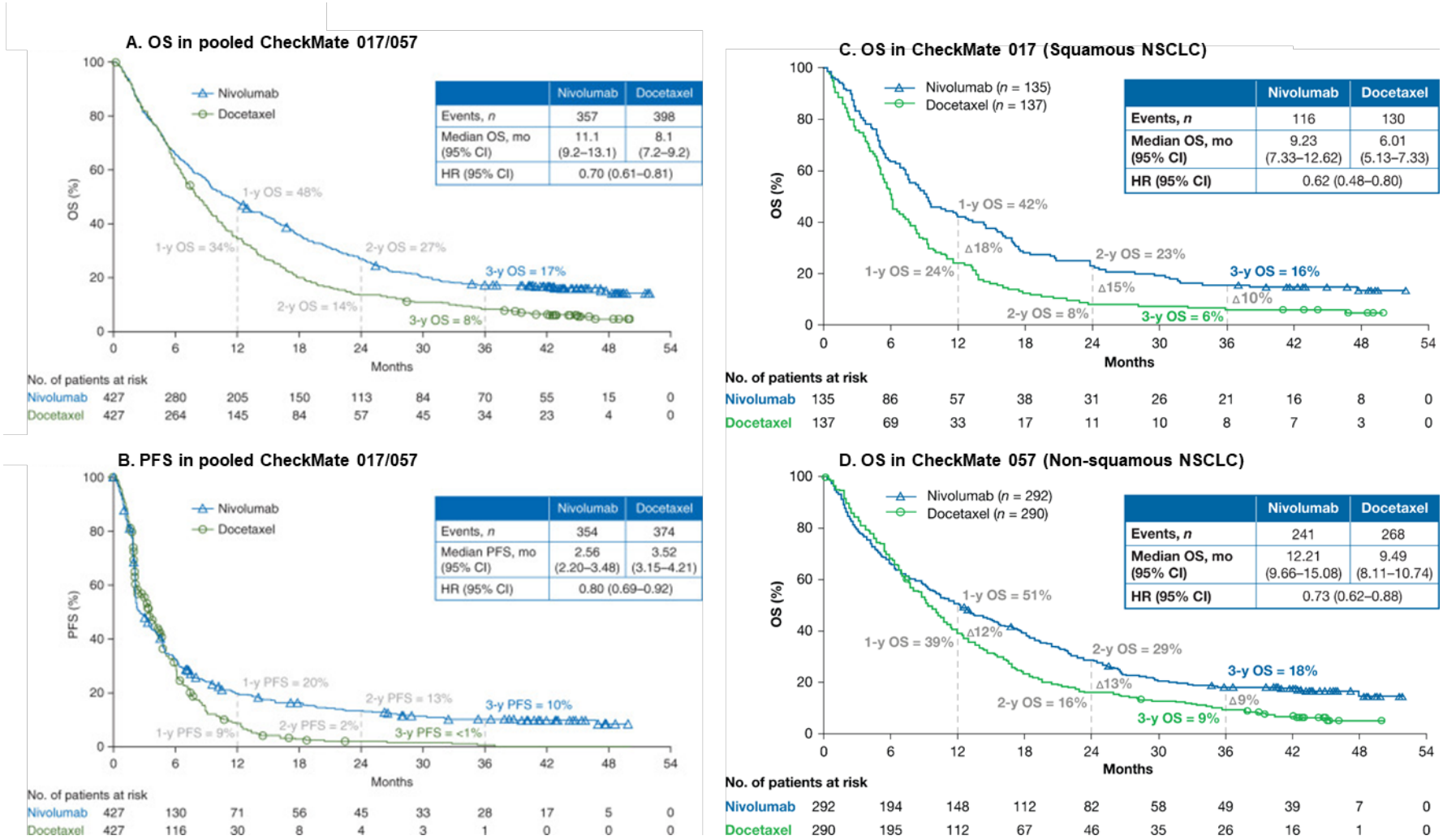


Figure 5. Survival outcomes from CheckMate 017 and CheckMate 057[4]

CI: confidence interval; HR: hazard ratio; NSCLC: non-small-cell lung cancer; OS: overall survival; PFS: progression-free survival.

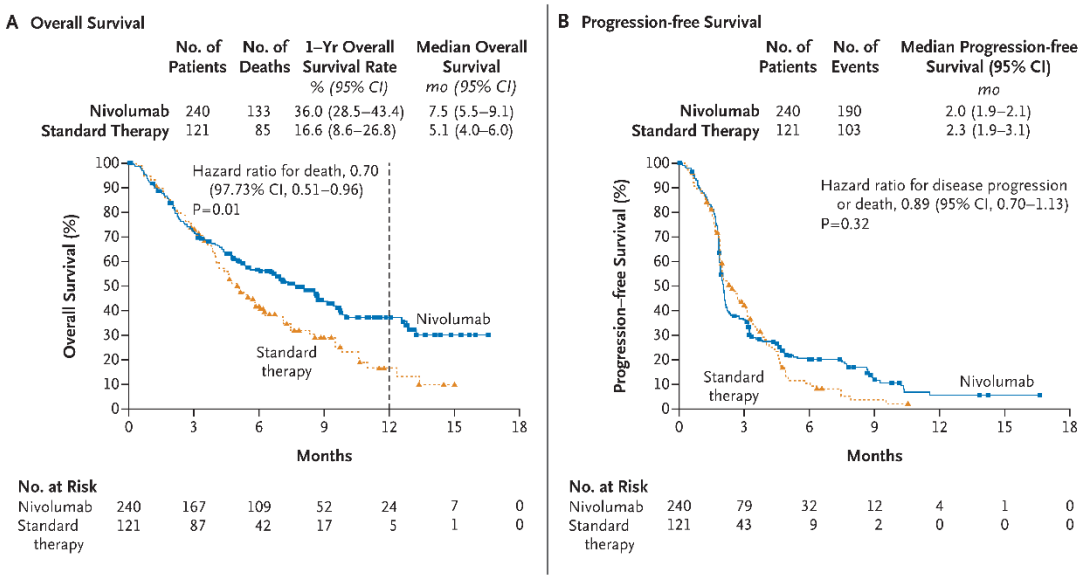


Figure 6. Survival outcomes from CheckMate 141 (Database lock December 2015)[6]

Figure 7. Long-term survival outcomes from CheckMate 141 (Database lock October 2019)[8]

b. Is the risk of 'early death' with initial nivolumab treatment worth an additional 2.58 months overall survival reported in the trial?

It should be noted that nivolumab treatment was associated with an additional 2.58 month in median OS in the context of a disease where median OS for standard of care is 8.38 months, which is highly clinically relevant. Further, it should be noted that OS remains higher at all time points after six months, indicating that more patients are living longer, which is also important to patients and clinicians.

In addition to the above, the company would like to clarify the definition of "early deaths". As acknowledged previously, the OS curve for nivolumab reflects a higher hazard in the initial three months, despite vastly improved outcomes following this period. However, the ERG report uses the terminology "early death". This was not defined during ATTRACTION-3 or in the company submission, but is defined in the ERG report as '[redacted]'. Labelling these deaths as "early deaths" may have contributed to a significant misunderstanding and misquoting of the data by the ERG report and Technical Engagement report, where this endpoint is quoted in place of data for deaths occurring in the first three months. "Early deaths" as defined by the ERG report may have occurred in the later period of the trial, as [redacted] received nivolumab >6 months (versus [redacted] for taxanes) and [redacted] received >12 months of treatment (versus [redacted] for taxanes). To add clarity and avoid this misunderstanding in future, "early deaths" should be relabelled; it is suggested that "on treatment deaths" is more appropriate. Deaths occurring in the first three months of the study could be labelled as such, in order to avoid confusion with the ERG definition of "early deaths".

Is best supportive care a relevant comparator?

3. Is nivolumab expected to completely replace taxane use for the given population?

Based on clinical expert opinion, patients with OSCC currently receive taxanes as second-line therapy where the patient is healthy enough and can tolerate this treatment. Patients who are unable to receive taxanes are currently receiving best supportive care (BSC), in the absence of a clinically effective therapy that is less toxic than taxanes. Clinical experts suggest that common reasons for patients to receive BSC include poor performance status, patient preference, co-morbidities and advanced age. This is in line with clinical advice provided to the ERG, where it was suggested that patients would only be considered as candidates for systemic anticancer therapy if they had an ECOG PS of 0 or 1.

If nivolumab becomes available, it is likely to replace the majority of taxane use, except for those patients with a clear contraindication to nivolumab. However, it is noted that some patients currently receiving BSC may switch to nivolumab: as nivolumab is less toxic than taxanes, patients may be considered healthy enough to receive nivolumab but not taxanes. Further, as noted by the ERG, some patients may opt for treatment with nivolumab but would otherwise decline to receive a further line of chemotherapy (i.e. a taxane).

Hence, nivolumab will predominantly replace taxane use but some replacement of BSC use will be plausible.

4. **In current NHS practice, what treatment would nivolumab-eligible patients be assigned to?**

In line with the response to Question 3, it is anticipated that nivolumab-eligible patients are predominantly receiving taxanes (docetaxel and paclitaxel), but some patients may be receiving BSC.

5. **What percentage of patients currently receive docetaxel, paclitaxel and irinotecan in NHS practice?**

    a. **Is irinotecan used in clinical practice for patients who would be eligible for nivolumab?**

    b. **What proportion of people currently receive docetaxel or paclitaxel in NHS practice?**

Market research conducted in 2019 indicates that taxanes encapsulated 100% of active second-line treatment for OSCC, reflecting 93.7% paclitaxel usage and 6.3% docetaxel usage. There is no usage of irinotecan in the second-line setting. In the third-line setting, market research indicates that 4% of all patients receive irinotecan, while 16% receive other therapies or enrol in clinical trials; the remainder received BSC.

Although it reflects gastro-oesophageal adenocarcinoma, as opposed to OSCC, a retrospective review conducted by the Royal Marsden hospital may be informative, although it should be noted that this focuses on patients receiving an active treatment (i.e. patients receiving BSC are not reflected).[18] Of the 511 patients, 200 (39%) received a second-line treatment, which was most commonly paclitaxel (35% of second-line treatment usage), while other therapies included fluoropyrimidine plus platinum doublet (17%), doublet plus anthracycline (9%), clinical trials (29%) or other (10%, which included regimens incorporating trastuzumab, raltitrexed, irinotecan and docetaxel). Of these 200 patients, 71 received third-line treatment, equivalent to 36% of patients receiving second-line treatment; this was most commonly clinical trial agents (36% of patients receiving third-line active treatment), paclitaxel (23%), FOLFIRI (17%), fluoropyrimidine plus platinum (11%) or other (10%, which included docetaxel and irinotecan monotherapy).[18]Given that fluoropyrimidine-based regimens are not recommended by oesophageal cancer guidelines in this setting, it can be anticipated that this usage may reflect usage in gastric and gastro-oesophageal junction cancer patients. However, this information broadly supports the market research data, reflecting the following:

- Paclitaxel has the highest usage of single-agent regimens.
- BSC is received by a large number of patients in both second-line and third-line settings.
- Irinotecan usage is low, across settings.

It is important to note that the evidence base is smaller for irinotecan than for taxanes, as suggested in the ERG report, which is reflected in the small amount of usage in clinical practice.

6. **Given the limitations of the indirect treatment comparison, is it appropriate to compare nivolumab with best supportive care in people who are not eligible for further chemotherapy?**

    a. **Is best supportive care a relevant comparator? Can the efficacy of nivolumab in ATTRACTION-3 be generalised to people for whom best supportive care is suitable? Is the ITC robust given the heterogeneity between studies?**

    b. **Are the results of the ITC generalisable to NHS practice?**

Relevance of best supportive care as a comparator

As outlined in the responses to Questions 3, 4 and 5, nivolumab will predominantly replace taxane use but some displacement of BSC use will be plausible. Hence, BSC cannot be considered a primary comparator in this setting, but an indirect comparison is provided in order to ensure that all evidence relevant to the decision problem is available.

Relevance of the NMA to the decision problem

The relevance of BSC as a comparator cannot be assessed based on the feasibility of a robust ITC. Where NICE has identified a relevant comparator, the company has made every effort to provide an informative evidence base for decision making. In line with the NICE reference case, a network meta-analysis (NMA) has been attempted based on a flawed evidence base. Hence, the submission includes this NMA and provides an accurate overview of the inherent limitations.

It should be noted that the NICE reference case includes provision for a scenario where valid data are not available to inform an indirect comparison: When sufficient relevant and valid data are not available for including in pairwise or network meta-analyses, the analysis may have to be restricted to a narrative overview that critically appraises individual studies and presents their results. In these circumstances, the Appraisal Committee will be particularly cautious when reviewing the results and in drawing conclusions about the relative clinical effectiveness of the treatment options.[19] Hence, it may be more informative to discuss the uncertainty in the comparison, as opposed to the appropriateness of making the comparison.

The Company Submission identifies and describes the limitations of the input data and the methods used to try and accommodate these; baseline characteristics were examined and documented extensively in the included populations and the most appropriate inputs were used such that the populations were as similar as possible. Specifically, the adjusted HR was used from Moriwaki et al.[20] as it was recognised that this was both more representative of the other studies included and more representative of the ATTRACTION-3 population. This would facilitate a more reliable NMA and a more generalisable output. In addition to this, vague priors were used so as not to overly affect any measure of between study variance generated by the model. Both fixed and random effects models were examined as

recommended by the NICE DSU TSD and results were broadly similar in terms of the absolute estimate of relative efficacy and the apparent heterogeneity and fit of the model.

When considered together, this indicates that while there may be some heterogeneity present, every reasonable action was taken to adjust for this, to examine the potential presence, to examine the potential impact and most importantly to provide the most representative value, which in turn facilitated an assessment of the comparative efficacy of BSC to docetaxel.

In addition, the Company Submission documented validation of the NMA result, which show that the outcomes of the NMA are credible and in line with the variation seen in reporting. The assessment of the robustness of the ITC is subjective; the methods, justification and transparency can be used to make judgement, and these are all provided in the Company Submission and further communications. Indeed, other options and networks were presented and discussed with their limitations identified, and justification for exclusion documented. The aim of the NMA was to facilitate the comparison of BSC with nivolumab within the evidence constraints but with the highest possible scientific credibility and method according to the NICE TSD guidelines such that a comparison could be made for scenario analysis should BSC be considered a relevant comparison.

Relevance of the NMA results to clinical practice

Clinical experts suggest that common reasons for patients to receive BSC include poor performance status, patient preference, co-morbidities and advanced age. ATTRACTION-3 did not restrict patient entry based on patient age but did specify patients with good performance status (ECOG 0-1) and excluded patients with contraindications to docetaxel and paclitaxel. These restrictions allowed recruitment of patients who could ethically be randomised to receive docetaxel and paclitaxel. Further, as noted in the response to Question 7, there is limited evidence to suggest different outcomes between patients with poorer performance scores. One published SLR and meta-analysis identified no difference in outcomes for patients with improved performance scores versus those with worse scores.[21]

In summary, if BSC is a relevant comparator, then a formal NMA complying with NICE TSD guidelines is the most appropriate format to inform that comparison, acknowledging the inherent limitations of the evidence base and the uncertainties that may arise as a result. In this context, it can be observed that nivolumab provides a significant improvement in outcomes versus BSC.

Generalisability of ATTRACTION-3 results

7. **Given the characteristics of the study population, are the results from the ATTRACTION-3 trial generalisable in a UK decision-making context?**

a. Clinical expert clarification on whether the rest of the world efficacy data is more relevant to the expected effectiveness of nivolumab in the NHS practice

Relevance of ATTRACTION-3 to the UK population, based on high prevalence of Asian patients

Advanced OC has a relatively low incidence in the UK and a very poor prognosis, as demonstrated in ATTRACTION-3.

However, there is notable global variation in the disease burden of OC as well as in the distribution of histological types of OC. In Asian countries the majority of OC cases are SCC (79% of the global SCC cases), while the highest burden of adenocarcinoma can be found in Western countries (46% of the total global adenocarcinoma cases).[22-24] Additionally, around 80% of the worldwide OC cases are diagnosed in Asia.[22, 25, 26] This global variation in OC burden and histology means that OSCC is a greater public health issue in Asia, while the Western population is more focused on adenocarcinoma cases. Hence, Asian populations are key evidence generators in OSCC, so that relevant treatment guidelines are based on evidence that is predominantly based on outcomes in Asian populations. However, this evidence is considered generalisable because the OSCC risk factors are relatively comparable across ethnicities and there is less ethnicity-specific impact on prognosis than observed for oesophageal adenocarcinoma. Further, clinical experts agree that the biology of OSCC is comparable between Asian and Western patients.

In support of this, an SLR was undertaken to identify studies reporting patient characteristics and treatment outcomes of oesophageal cancer patients in Asian and Western countries, specifically adult patients with advanced, metastatic or recurrent unresectable OSCC who were refractory, intolerant or resistant to first line therapy; outcomes were also assessed in the overall OC population and the adenocarcinoma subgroup. Median survival (OS) in OSCC patients was comparable between populations, based on a mean of the reported values of 7.5 months (range: 5.1-10.9 months; 3 treatment groups) in the Asian population versus 7.4 months (range: 6.0-8.7 months; 23 treatment groups) in the Western population. OS at one year was slightly improved in Western populations, with a mean reported value of 27.9% (range: 14.4-41.3%; 2 treatment groups) versus 21.1% (range: 16.7-26.7%; 3 treatment groups) in an Asian population. However, outcomes for Western patients with oesophageal adenocarcinoma were dramatically lower, based on a mean of reported values for median OS of 5.6 months (range: 4.0-7.2 months; 2 studies); however, there were no studies in Asian populations to provide a comparison. The high prevalence of adenocarcinoma in a Western population are likely driving outcomes in the overall OC population, where median OS was lower than for Asian patients (median OS: 8.1 months versus 5.7 months).

Problems with recruitment of OSCC patients is not unique to ATTRACTION-3; other gastro-oesophageal cancer therapies recently appraised by NICE have enrolled large proportions of Asian patients.[27-32] Further, NCCN and ESMO evidence cited to inform treatment decisions is highly limited. [33, 34]Although three small non-randomised, uncontrolled studies are cited, limited survival data are reported and median OS ranged from 274 days (approximately 9 months) to 13.2 months. One large RCT is cited by NCCN (COUGAR-2), which reflects

evidence in gastro-oesophageal adenocarcinoma patients and included only 33 oesophageal adenocarcinoma patients out of 168 total cohort. In this study, median OS was 5.2 months in the docetaxel arm compared with 3.6 months in the supportive care arm (OS at 6 months: 82% versus 39%).[35] These data are in line with that from the SLR outlined above, indicating that it is not be appropriate to use adenocarcinoma data to inform SCC outcomes.

Supportive evidence is available from other indications; **CheckMate 078** assessed nivolumab versus docetaxel in an Asian patient population with previously treated advanced non-small cell lung cancer.[36] Median OS was broadly comparable between studies for nivolumab treated patients (12.0 months for CheckMate 078 versus 11.1 months for CheckMate 017/057), although there was a slight difference in docetaxel- treated patients (9.6 months for CheckMate 078 versus 8.1 months for CheckMate 017/057). This was reflected in OS outcomes at one year (nivolumab: 50% for CheckMate 078 versus 48% for CheckMate 017/057; docetaxel: 39% for CheckMate 078 versus 34% for CheckMate 017/057). Hence, within an indication, outcomes for Asian patients may be broadly comparable to Western patients, particularly for patients receiving nivolumab.

BMS has discussed these data limitations extensively with the EMA, with the aim of ensuring that all avenues are explored to ensure rapid access to nivolumab for this disease with significant unmet need. Further, BMS remains committed to providing nivolumab to UK patients via the EAMS, and data will be made available to both NICE and the EMA.

Relevance of ATTRACTION-3 to the UK population, based on patient characteristics

Given the limited data to inform UK patient characteristics in patients with previously treated OSCC, a comparison is provided versus UK-specific published studies for gastro-oesophageal adenocarcinoma. It is highlighted that this comparison should be treated with caution, as patients with gastric adenocarcinoma may differ from OSCC. However, this comparison may be used to highlight uncertainties.

As can be seen in Table 3, ATTRACTION-3 enrolled a slightly higher proportion of male patients than the Royal Marsden retrospective review and the COUGAR-2 clinical study. However, baseline age broadly aligned across all sources. Slightly fewer patients with ECOG status of 1 were enrolled and no patients with ECOG status of 2 were enrolled.

Table 3. Comparison of ATTRACTION-3 baseline characteristics versus those from UK-specific studies

	ATTRACTION 3[17]		Cougar-2[35]		Royal Marsden retrospective review[18]
	Nivolumab	Taxane	Docetaxel	Active symptom control	
N	210	**209**	**84**	**84**	**511**
Sex, male (%)	179 (85.2%)	185 (88.5%)	69 (82%)	67 (80%)	384 (75%)
Median age (range), years	64.0 (37-82)	67.0 (33-87)	65 (28–84)	66 (36–84)	66 (24-90)***

Eastern Cooperative Oncology Group performance status	0	101 (48.1%)	107 (51.2%)	24 (28%)	22 (26%)	64 (13%)
	1	109 (51.9%)	102 (48.8%)	46 (55%)	50 (60%)	276 (54%)
	2	0	0	14 (17%)	12 (14%)	87 (17%)
Disease status	Locally advanced	[redacted]	[redacted]	11 (13%)	10 (12%)	68 (13%)**
	Metastatic disease	[redacted]	[redacted]	73 (87%)	74 (88%)	335 (66)**
Site of primary disease	Oesophagus	100%	100%	18 (22%)	15 (18%)	148 (29%)
	Oesophagogastric junction	0	0	27 (32%)	32 (38%)	173 (34%)
	Stomach	0	0	39 (46%)	37 (44%)	190 (37%)
Histology	Adenocarcinoma	0	0	100%	100%	100%
	Squamous cell carcinoma	100%	100%	0	0	0
* summarised for patients with non-recurrent oesophageal cancer ** 21% of patients had relapsed metastatic disease after radical treatment *** Age at diagnosis, not study baseline						

It should also be noted that EAMS data will be made available to NICE and can be used to validate generalisability of ATTRACTION-3 outcomes to the UK setting.

b. Who is likely to receive nivolumab in clinical practice? Will it be restricted to people with a good performance status (ECOG 0-1)? Is the efficacy of nivolumab in NHS practice (people with a worse performance status) likely to be worse than in the ATTRACTION-3 ?

Clinical trials commonly specify performance scores as an inclusion criterion, typically based on either ECOG or Karnofsky scale. This leads to limited evidence of net clinical benefit for patients with certain performance scores, typically those with worse scores. This absence of evidence contributes to a reluctance to provide certain treatments to patients of reduced performance score. However, this is limited evidence to suggest different outcomes between patients with different performance score.

A 2017 SLR and meta-analysis of RCTs assessed clinical benefit by performance score subgroups. This identified 110 RCTs, with 66 (60%) reporting performance score subgroups for efficacy and none reporting subgroups for toxicity. For these 66 RCTs, pooled HRs for good performance score and reduced performance score subgroups were 0.65 (95% CI 0.61 to 0.70) and 0.67 (95% CI 0.62 to 0.72), respectively, with no difference between the two groups (p=0.68). Sensitivity analyses based on drug or cancer type and type of endpoints (OS or PFS) demonstrated similar results.[21]

It should be noted that taxane use is not currently restricted by performance score. Although poor performance score is a common reason for patients to receive BSC, patients with ECOG performance score 2 may receive taxanes under clinician supervision. As nivolumab has improved efficacy and reduced toxicity compared with taxanes, it may be illogical to restrict nivolumab use to patients who could otherwise receive a taxane.

As outlined in response to Question 3, nivolumab will predominantly be prescribed in patients currently receiving taxanes. However, if nivolumab becomes available, some patients currently receiving BSC may switch to nivolumab: as nivolumab is less toxic than taxanes, patients may be considered healthy enough to receive nivolumab but not taxanes.

8. **Would it be more appropriate to use efficacy data from the rest of world population compared with the intention to treat population to estimate the clinical and cost effectiveness in clinical practice?**

    a. **Is there a subgroup within ATTRACTION-3 which is considered to have similar characteristics to the UK patient population?**

As outlined in the response to Question 7, evidence indicates that OSCC risk factors are relatively comparable across ethnicities and there is less ethnicity-specific impact on prognosis than observed for oesophageal adenocarcinoma. Further, clinical experts agree that the biology of OSCC is comparable between Asian and Western patients. More generally, it should be noted that any differences in outcomes for gastro-oesophageal cancer are observed in Asian patients versus Western patients, as opposed to Japanese patients versus predominantly Chinese patients.

It should also be noted that the ATTRACTION-3 trial was powered to show differences in efficacy for nivolumab versus taxanes in the overall population, rather than specifically in the rest of world population. Although the improvement remains significant, reducing the patient numbers increases uncertainty. Hence, it is more appropriate to use the overall population in the absence of evidence that it is not relevant to the UK patient population.

Additionally, it may be noted that the statement “Efficacy data in the model was based on the hazard ratio in the intention to treat population from ATTRACTION-3” is considered to be factually inaccurate. The data populating the company cost-effectiveness model was derived from the intention to treat population but was not “based on the hazard ratio”, so this part of the sentence should be deleted for clarity.

Safety data for nivolumab, early deaths higher on nivolumab

9. **To what extent does the company’s rationale provide an explanation for the early deaths observed in the nivolumab arm?**

a. **Do clinical experts agree with the company's rationale for higher death rate over first 3 months in the nivolumab arm?**

b. **Clinical expert advice is sought on whether the initial higher death rate with nivolumab in ATTRACTION-3 is likely to be seen in NHS practice.**

c. **Is it possible to determine in advance which patients are likely to die before they can benefit from treatment with nivolumab (if so please explain)?**

Of primary interest, it should be noted that the statistics quoted in the Technical Engagement report are factually inaccurate and arise from a misunderstanding with the ERG's definition of the term "early deaths". The values quoted in the statement (i.e. [redacted]) refer to "early deaths", defined by the ERG as "[redacted]". As [redacted] received nivolumab >6 months (versus [redacted] for taxanes) and [redacted] received >12 months of treatment (versus [redacted] for taxanes), "early deaths" as defined by the ERG could in fact occur far after the first three months. This was highlighted at the factual accuracy check stage, where the company requested that this naming was changed (e.g. "on treatment deaths"); further, the initial high hazard period could more accurately be referred to as "deaths in the initial three months".

The correct statistics are quoted in the amended statement from the Technical Engagement report, below:

- However, the number of deaths in the first 3 months of treatment was higher in the nivolumab compared with the taxane arm ([redacted].

As noted in the response to Question 2, although there is initial crossover in the OS Kaplan-Meier data, median OS and PFS rates from 6 months to end of follow up show a beneficial impact for nivolumab versus taxanes. Landmark analyses provided in response to Question 2 (Figure 3) demonstrate that outcomes are significantly improved for nivolumab versus taxanes in those patients alive at three months.

This is a common pattern of response for immuno-oncology therapies, particularly those indications where survival is short and evidence is versus an active comparator. As outlined in Figure 4, OS in CheckMate 057 showed a similar initial crossover, followed by long-term survival improvement. Similarly, OS in CheckMate 141 (provided in Figure 6 and Figure 7) showed comparable outcomes to standard therapy in the first three months followed by significant benefit in the longer term. Further, this effect is observed across immuno-oncology therapies, wherein there is either limited benefit in the initial period (as observed in Figure 8 and Figure 9) or there is higher initial hazard (as observed in Figure 10). Hence, there is significant evidence for a class effect driving this observation.

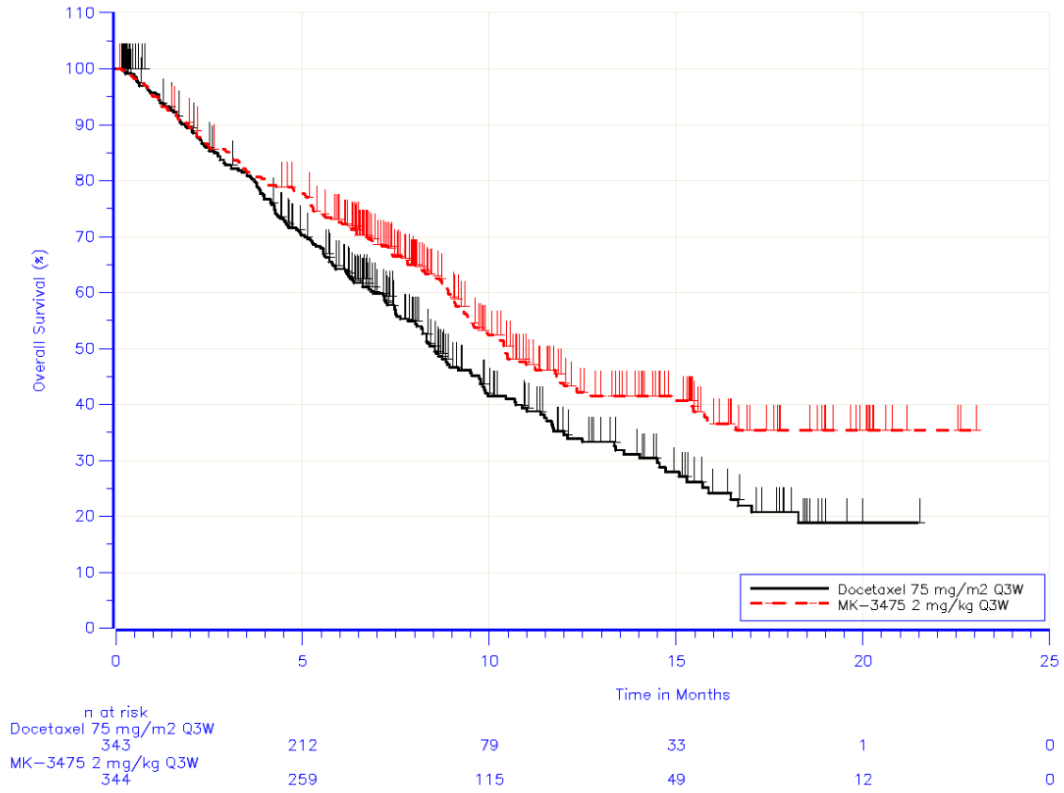


Figure 8. KEYNOTE-010 OS Kaplan-Meier outcomes (reproduced from TA428)[37]

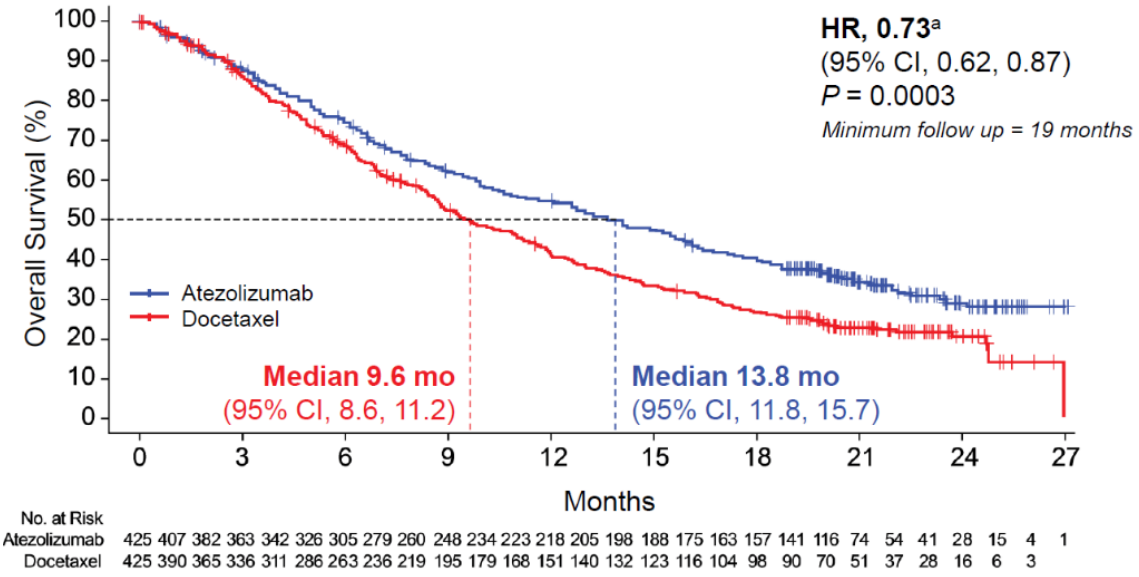
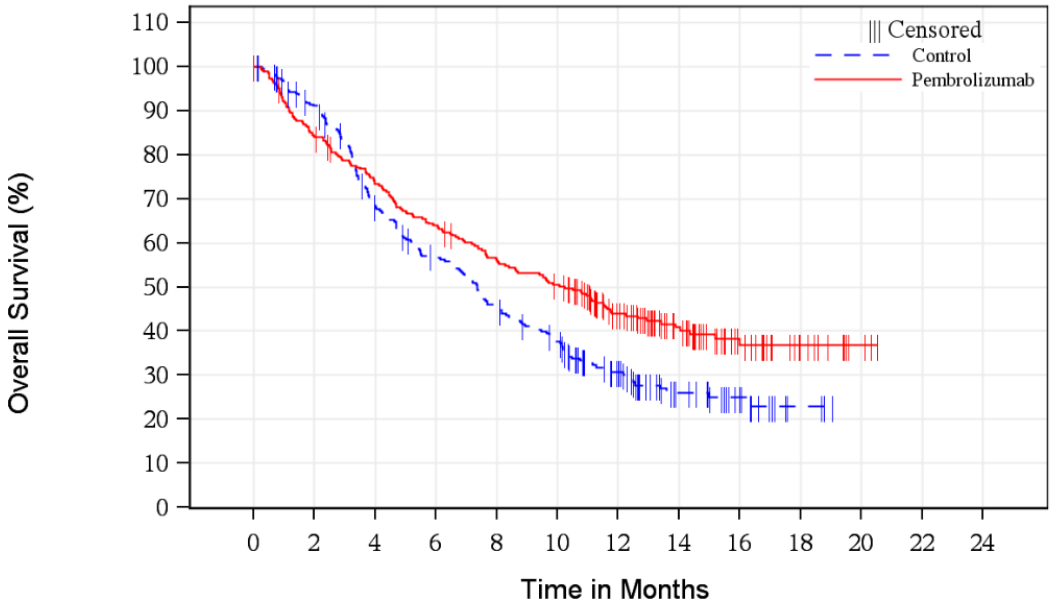


Figure 9. OAK OS Kaplan-Meier outcomes (reproduced from TA520)[38]



Number of subject at risk

Control	272	232	171	138	109	89	55	27	14	3	0	0	0
Pembrolizumab	270	226	194	169	147	131	87	54	27	13	4	0	0

Figure 10. KEYNOTE-045 OS Kaplan-Meier outcomes (reproduced from TA519)[39]

No adjustment to efficacy for any beneficial effects of third-line therapy

10. In clinical practice, is active third-line therapy administered following progression after the use of a taxane or nivolumab?

a. Clinical experts to clarify whether patients in NHS practice receive active treatment after progression on nivolumab or taxanes. What proportion of patients receive different 3rd line treatments?

b. Is nivolumab expected to completely replace taxane use for the proposed indication and population? If so, would taxane therapy then be offered to nivolumab patients post-progression?

Again, it is of primary interest that the data quoted in the Technical Engagement report contains a factual inaccuracy, as it is based on a mischaracterisation of the available data. In ATTRACTION-3, patients could continue study treatment beyond progression, based on set criteria outlined in the Clinical Study Report and in Section B.2.6.1.3.1 of the company submission. The data quoted in the Technical Engagement report refers to treatment post-progression [i.e. continued treatment with nivolumab following progression], which was received by [redacted] of the 210 nivolumab patients ([redacted] with a median of [redacted] treatments (range: [redacted] treatments) and a median post-discontinuation time on treatment of [redacted] days (95% CI: [redacted]). For the taxanes arm, [redacted] of the 209 patients ([redacted] received treatment post-progression [i.e. continued treatment with taxanes following progression]; all patients had one subsequent treatment and a median post-discontinuation time on treatment of one day. This treatment beyond progression is reflected in the economic model in both costs and benefits, for both treatment arms. As noted in the ERG report, subsequent therapy (i.e. not allocated study therapy) was received by 119 (57%) of 210 patients in the nivolumab group and 115 (55%) of 209 patients in the taxanes group.[17]

As noted in the submission and in the response to Question 5, there is some uncertainty in current therapy usage in OSCC, due in part to a lack of formal treatment guidelines for OSCC and heterogeneity between clinical experts. Market research conducted in 2019 indicated that the vast majority of patients in the third-line setting receive BSC (4% of patients receive irinotecan and 16% receive other therapies). In support of this, a retrospective review of gastro-oesophageal adenocarcinoma (i.e. not OSCC) conducted by the Royal Marsden hospital found that only 36% of patients receiving second-line treatment go on to receive a third-line treatment, and that the predominant treatment in these patients clinical trial agents (36% of patients receiving third-line active treatment).[18] It is likely that third-line treatment usage in OSCC may be lower, given the absence of treatment options. Further, it should be noted that there is limited evidence to support third-line treatment decision, so the beneficial impact may be uncertain.

Although usage of active treatment as a subsequent therapy in ATTRACTION-3 is higher than indicated by market research, it may be more comparable to that reported in the Royal Marsden hospital retrospective review.

It should be noted that nivolumab is less toxic and has improved survival, so it may be considered patient sparing, which is important for clinicians and patients. Hence, use of nivolumab in clinical practice may enable more patients to be suitable for third-line treatment with taxanes. This may be reflected in ATTRACTION-3 data, wherein taxane therapy was the most common subsequent therapy in the nivolumab arm (100 patients; 47.6%), whereas fewer patients in the taxane arm patients received subsequent taxane therapy (43 patients; 20.6%) and a larger number received more speculative subsequent treatments, including fluoropyrimidine-based chemotherapy (18.7%), immunotherapies (6.2%), platinum-based chemotherapy (10.5%) and other systemic chemotherapies (13.4%).

11. Should the efficacy of nivolumab be adjusted to account for beneficial effects of third-line therapy?

Given the uncertainty around the composition of third-line standard of care, this may not be possible to do in any robust manner. Further, as outlined in response to Question 10, subsequent treatments in ATTRACTION-3 may reflect usage in clinical practice following use of a patient sparing regimen, so that more patients may be able to receive taxanes in the third-line setting.

Further, if this adjustment is considered appropriate, it must also be applied to the taxane arm. It should be noted that 115 (55%) patients in the taxanes arm received subsequent therapy, which included subsequent taxanes (43 patients) and also immunotherapies (13 patients; nivolumab: 7 patients).

Subgroup analysis by taxane was not provided

12. Is a comparison of nivolumab with individual taxane or combined taxanes more relevant to NHS practice? Which taxane is most commonly used in the NHS?

a. Is it reasonable to assume a class effect for taxane therapy?

b. Could post-hoc analysis of the effectiveness of efficacy of nivolumab compared with either docetaxel or paclitaxel have been carried out?

c. Is a comparison of nivolumab with individual taxanes more relevant to NHS practice than comparisons with a combined taxane arm?

Effectiveness analysis of the efficacy of nivolumab compared with either docetaxel or paclitaxel was provided in the original company submission (page 37, Table 13). Further, cost-effectiveness analysis for nivolumab versus docetaxel or paclitaxel was provided on page 147, Table 79. Hence, it is not accurate to say that subgroup analysis by taxane was not provided.

During provision of the submission, it was considered more appropriate to use the combined taxane arm as a comparator. Treatment guidelines recommend taxane monotherapy (i.e. not specific taxanes) for the second-line treatment of OC.[33] Further, published clinical outcomes are comparable between docetaxel and paclitaxel in this setting. As the ATTRACTION-3 trial was powered to show differences in efficacy for nivolumab against the combined taxane arm, as opposed to docetaxel and paclitaxel separately, low patient numbers receiving individual treatments may impact on outcomes, particularly during later periods of follow-up. Hence, the base case analysis applied a combined taxane arm as a primary analysis and a comparison to individual taxanes as scenario analyses.

Within ATTRACTION-3, clinician preference was observed to be associated with study centre; therefore there is a risk of bias in post-hoc analysis per taxane, as detailed in the response to question 13. Further, any post-hoc analysis would produce models based on fewer patients than the overall taxane population, which, even in the absence of bias, would decrease the precision of estimates, particularly in extrapolation. Given that a difference in outcomes is not supported by either the data or clinical opinion, this analysis would inappropriately increase the uncertainty around the outcomes informing the decision problem.

In terms of current usage, there is an absence of evidence. Market research conducted in 2019 indicates that paclitaxel is the primary taxane for active second-line treatment for OSCC, reflecting 93.7% of active comparator usage. This is supported by retrospective review data from the Royal Marsden hospital; although it reflects gastro-oesophageal adenocarcinoma, as opposed to OSCC, paclitaxel has the highest usage of single-agent regimens.[18]

13. Are health outcomes expected to be different between 'docetaxel-preferred' and 'paclitaxel-preferred' populations?

a. Are there systematic differences in people who would be suitable for treatment with either docetaxel or paclitaxel? How should these separate populations be defined?

Published clinical outcomes are comparable between docetaxel and paclitaxel in this setting. Additionally, clinical expert opinion obtained during a clinical advisory board meeting suggested that there is no "standard of care" for treatment-experienced OSCC patients, so that treatment options were described as highly individualised, both for the patient and for the clinician.

The ERG suggest that there may be docetaxel-preferred and paclitaxel-preferred patients, based on clinical advice provided to the ERG, suggesting that docetaxel would be the preferred choice for most patients, owing to the fact that it is administered less frequently (i.e. once every three weeks instead of once per week). However, due to potential issues with tolerability, some patients may instead be treated with weekly paclitaxel which is considered to have a more favourable safety profile. However, it is acknowledged that paclitaxel and docetaxel may have similar efficacy. This is supported by company market research (Figure 11), demonstrating that key clinical outcomes are perceived to be equivalent between docetaxel and paclitaxel, with dosing schedule, tolerability and quality of life distinguishing the two therapies. This is in line with ATTRACTION-3, wherein key baseline prognostic factors and clinical outcomes are comparable between taxane groups.

In many cases, choice of taxanes may be related to clinician preference or local guidelines, as opposed to patient characteristics. Clinician preference was observed to be associated with study centre within ATTRACTION-3 and thus there is a risk of bias in post-hoc analysis per taxane. Of the 51 sites with 3 or more patients randomised, [redacted] uniquely preferred a single taxane, with [redacted] preferring docetaxel and [redacted] preferring paclitaxel. Thus, study centre was strongly associated with taxane preference. One of the aims of multicentre randomised controlled trials is to reduce the bias associated with unmeasured confounders at single sites, so to systematically decrease the heterogeneity of site-associated confounders in a subgroup analysis by subgrouping on variables highly associated with study centres is only justified if the homogeneity of interest is within an identifiable population – e.g. patients within a specified geographic region or with specific socioeconomic status. If the population is not identifiable, i.e. sites with strong clinician preference for a specific taxane cannot be identified a priori, then the results cannot be generalised and are at risk of bias.

In summary, there does not appear to be any difference between docetaxel-preferred and paclitaxel-preferred patients, and it is unclear if these patient groups exist consistently in clinical practice.

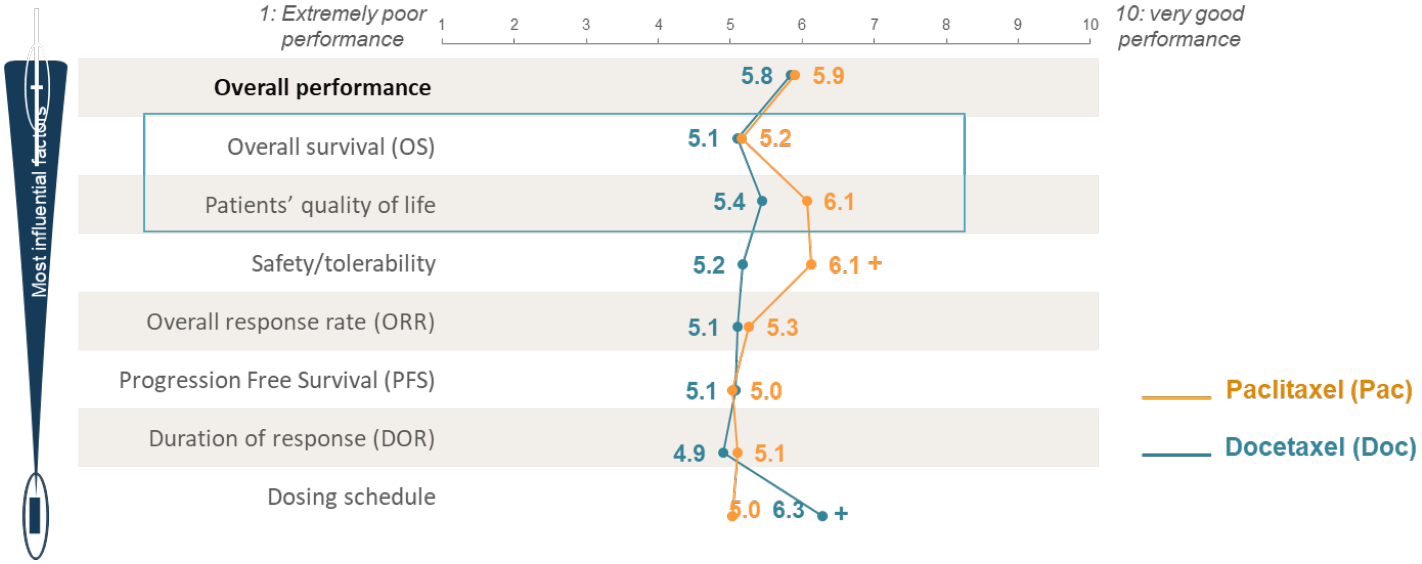


Figure 11. Market research: treatment perceptions for docetaxel versus paclitaxel

Difference between company and ERG ICER

14. Which modelling methods are most suitable for estimating overall survival and expected time on treatment?

Relevant survival modelling guidelines indicate that survival extrapolations should reflect the disease pathway and plausible biological explanation for treatment effect. This includes reviewing the overall impact across model inputs, rather than reviewing models in isolation. This is specifically of note in the ERG base case, where a case is made for each input in isolation, but the overall impact is to predict clinically implausible outcomes.

As noted by the ERG, the ERG base case applies more conservative extrapolations for nivolumab OS due to the concerns around the generalisability of the evidence. However, taxane OS is assumed to be more optimistic, to the extent that ATTRACTION-3 extrapolations predict longer mean OS for taxanes than for nivolumab, which can be considered implausible in the context of the observed data. Further, despite shorter OS assumptions, it is assumed that time on treatment is increased. As the Summary of Product Characteristics (SmPC) for nivolumab specifies that treatment should be administered for as long as there is clinical benefit, derivation of time on treatment should be

considered in the context of both PFS and OS.[40] Although it is likely that some patients will receive treatment beyond progression, it is not plausible that there will be extended post-progression treatment period in the absence of clear benefits such as improved quality of life. Further, it is implausible that extended post-progression treatment would be seen in the absence of clinical improvement, which would be reflected in OS. In this context, it is not appropriate to extend time on treatment, reduce OS (for nivolumab only) and assume no post-progression utility differential between treatment arms. Hence, survival modelling in this indication should be considered across model inputs, rather than in isolation, otherwise the results are clinically implausible.

The ERG quote NICE DSU TSD 14 guidance to say that *"where parametric models are fitted separately to individual treatment arms it is sensible to use the same 'type' of model, that is if a Weibull model is fitted to one treatment arm a Weibull should also be fitted to the other treatment arm"*.[41] The company agrees that it should be considered appropriate where models are fitted separately to individual treatment arms to use the same model "type" where the hazards presenting are the same and where the data directs. However, the company does not consider that these conditions are satisfied by the nivolumab and taxane arm in ATTRACTION-3. Bagust and Beale (2014)[42] describe recommendations for pragmatic modelling that were used to guide analysis as reported in the CS Section 3.3.2.1.1 page 96. This report recommends that "*The presumption should be against joint modeling of treatment arms unless modeling the trial arms independently reveals that functional forms and parameter estimates are closely aligned. Nonetheless, the appropriateness of each separate functional form needs careful justification, from both the available data and other sources (such as clinical experience and, if available, patient registries)".* These recommendations were adhered to in the company base case approach and seem to be at odds with the technique used in the ERG base case selection. Appendix M (Figure 48) shows the cumulative hazard profiles from ATTRACTION -3 are quite different. This is largely driven by very different mechanisms of action, as outlined in the response to Question 2, which would be expected to have different hazard profiles that could not be reflected by the same "type" of model. This is also apparent in the responses to clarification question B10 on pages 54-55, figures 46-49 where the SP Weibull (cut point at 5.75 months) for nivolumab can be seen to deviate greatly from the observed data, both over and underestimating. In contrast, the same model fit to the taxane arm adheres much more closely throughout.

Questions 19 and 20 provide a detailed comparison of the company base case analysis survival methods versus the ERG survival analysis methods.

15. Which estimates of treatment, administration and medical resource use costs are most reliable?

The NICE reference case states that analyses based on price reductions for the NHS will only be considered when the reduced prices are transparent and consistently available across the NHS, and if the period for which the specified price is available is guaranteed.

Costs available the electronic Market Information Tool (eMIT) capture volume-based discounts provided to the NHS. It is unclear if all NHS trusts will have access to these medicines at this price, particularly due to the large standard deviation. Further, there is no confirmation these prices will be available for a guaranteed period.

The clinician consultation cost, as stated in Table 70 of the company submission, is derived as a weighted average of costs for a consultant led and non-consultant led medical oncologist, using service codes 370. This yields a cost of £187.36.

The nerve block cost of £532.96 is sourced from the National Cost Collection 2018/19[43], *nerve block or destruction of nerve, for Pain Management*, total HRGs, currency code AB26Z.

As the majority of the components informing the hospitalisation cost had a length of stay exceeding one week, it was deemed inappropriate to use these costs, in order to avoid double counting. On this basis, hospitalisation costs were standardised using the length of stay in order to gain a daily cost appropriate for use in the model.

Model time horizon (40 years in company base-case)

16. What is the most appropriate time horizon for the economic model?

a. Clinical expert advice is required on the life expectancy of patients with unresectable, advanced oesophageal cancer treated with nivolumab or taxanes

b. What is the most relevant time horizon for the economic model?

The Guide to the methods of technology appraisal 2013[19] states that analysis should be “Long enough to reflect all important differences in costs or outcomes between the technologies being compared” and that “Analyses that limit the time horizon to periods shorter than the expected impact of treatment do not usually provide the best estimates of benefits and costs” Thus, in line with NICE requirements, a lifetime horizon was used.

While the company does not dispute that that the majority of patients have an estimated survival of less than 12 months, all patients should be modelled to until death in line with NICE guidance. Of note, even using conservative assumptions (as provided in the ERG base case), patients remain alive after 10 years in the nivolumab arm. As such, assuming a time horizon of 10 years is insufficient when modelling in line with NICE guidance.

Alternative extrapolations for overall survival

17. Is a fully-parametric or semi-parametric model most appropriate to predict the long term effectiveness of nivolumab and taxane therapy?

Relevant survival modelling guidelines indicate that survival extrapolations should reflect the disease pathway and plausible biological explanation for treatment effect. This includes reviewing the overall impact across model inputs, rather than reviewing models in isolation. Parametric models were deemed inappropriate for modelling the extrapolated OS based on poor visual inspection and the inability to adequately capture the hazard associated with each treatment. The company also wishes to note that the ERG also chose to apply a semi-parametric model to predict the benefit of nivolumab and taxane therapy.

18. If a semi-parametric model is preferred, is 2.99 months an appropriate point to start the extrapolation?

a. If a semi-parametric model is preferred, is 2.99 or 5.75 months an appropriate point to start the extrapolation?

Upon assessment for cut points, 2.99 months was deemed most appropriate based on the rationale that 2.99 months exceeded the mean and median time to response in all treatment arms, meaning that the majority of responses will already have occurred and will be captured directly from observed data. As well as this, placing the cut point at 2.99 months allows for as much data as possible to inform the extrapolated region. Importantly, the curve chosen by the ERG to best represent the taxane OS was generalised gamma, citing its flexibility as one of the reasons. However, where cut points were placed at 5.75 months, the generalised gamma curve and the exponential curve are almost indistinguishable from each other on the plots as they are almost completely overlaid. This would suggest that while the generalised gamma model has increased flexibility and the ability to fit better to the data presented than an exponential, it is in fact taking on the exact shape of the exponential model. Therefore, there seems to be no particular reason to disregard that the exponential model fits just as well or to move the cut point further from the initially proposed base case.

19. Does the overall survival predicted in the company ([redacted] at 10 years) or ERG model (0.2% at 10 years) best represent the likely overall survival in NHS practice?

Per the design of ATTRACTION-3, it has been predicted that the population in receipt of nivolumab will consist predominantly of patients with a conventional treatment response profile, with a small fraction (5% per the statistical analysis plan) exhibiting long term response not comparable on the timescale over which patients receiving taxanes were expected to survive. This was justified by the precedent of nivolumab

vs investigator's choice in CA209141 (metastatic platinum-refractory squamous cell carcinoma of the head and neck) and a phase 2 study of nivolumab in oesophageal cancer refractory or intolerant to standard therapy (ONO-4538-07).

Under this condition of a relatively small fraction experiencing long term response, accurate detection of this fraction whilst a larger group of non-long term survival patients remain alive (OS at month 24 was 19.1% in the nivolumab arm) is difficult. However, the existence of this subgroup remains suggested by the survival profiles of patients receiving nivolumab in other indications and trials and remains consistent with the collected data of ATTRACTION-3. The resultant reducing marginal hazard/increasing conditional survival in long term follow-up expected by such a group is consistent with the company model.

Table 4. Summary of predicted overall survival

Year	Survival			
	Company base case		ERG preferred base case	
	Nivolumab	Taxane	Nivolumab	Taxane
Year 1	45.61%	36.57%	46.07%	35.40%
Year 2	21.27%	11.06%	20.70%	12.20%
Year 3	12.33%	3.34%	10.22%	4.42%
Year 5	5.84%	0.30%	2.93%	0.63%
Year 10	1.92%	0.00%	0.20%	0.01%

Exploratory analysis of utility values

20. Are the differences in utility between the nivolumab and taxane arm in the company model clinically plausible?

There are substantial clinical benefits for nivolumab over taxanes in previously treated OSCC that may be driving differences in utility. In particular, patients in the nivolumab arm have improved OS. As utility in oncology is typically a function of time to death, improved OS rates are a key component in postponing quality of life decrements.[44-49] Significantly, observed ATTRACTION-3 data demonstrates that there is a large post-progression survival benefit compared with taxanes, supporting the impact of nivolumab on quality of life. As the ERG accepts that there is an extension of survival in post-progression for nivolumab (present in the ERGs base case) and does not dispute the profile of decreasing utility with proximity to death (within 18 months of death), the assumption of equal utility in the progressed disease state can be considered illogical.

This can be considered particularly illogical, given that the ERG assumes that time on treatment is extended into post-progression but patients receive limited post-progression utility benefit.

Further, the utility differences between nivolumab and taxanes reflect the safety profile of nivolumab compared with chemotherapy; 65.6% of patients in the nivolumab arm reported a drug-related AE (grade 3-5: 18.2%) versus 95.2% for patients receiving paclitaxel or docetaxel (grade 3-5: 64.0%).

It should be noted that outcomes in the taxane arm validate very well to outcomes in previous NICE HTAs. Although there is limited evidence in the oesophageal cancer setting, the utility associated with the pre-progression state for the taxane arm was [redacted], which can be considered comparable with the published literature for gastric cancer (0.737[50]). Similarly, the post-progression utility in the taxane arm was [redacted], which is only slightly below published values (0.587[50]). Where there is such close validation for the taxane arm to published literature values, there is limited evidence to support a lack of immunotherapy-specific impact on utility values.

Further, the utility values observed during ATTRACTION-3 are broadly equivalent to utility values observed from other immunotherapy indications,[51-56] indicating that this utility gain may be due to the novel mechanism of action for immunotherapies. These treatment-specific utilities are frequently queried during NICE appraisal, despite being consistently observed across indications, immunotherapies and studies. These data demonstrate that progression has limited impact on utility in patients receiving immunotherapies, whereas time to death is more impactful.[44, 46, 47, 57] Hence, this evidence should support the impact of immunotherapies on utilities.

Additionally, it should be noted that quality of life outcomes during ATTRACTION-3 remained relatively stable in the nivolumab arm, as determined by EQ-5D and EQ-VAS; however, patients receiving taxanes frequently reported worsened quality of life outcomes during the trial period. It would be inappropriate to not reflect that difference in quality of life outcomes. Thus, the quality of life data derived from patients during ATTRACTION-3 reflects the expected benefits of nivolumab over taxanes, including the potential for immune system stimulation following progression.

21. Would a treatment independent approach (e.g. using the same values for pre and post progression across the treatment arms) be more appropriate?

This is considered inappropriate, for the reasons provided in response to Question 20.

Alternative extrapolations of time on treatment

22. Are fully-parametric or semi-parametric methods most appropriate for estimating time on treatment?

a. It is not clear why a fully-parametric model was used by the company after using semi-parametric for OS and PFS. What was the rationale for this?

b. Is the long term extrapolation based on fully-parametric (company) or semi-parametric (ERG) methods most appropriate for estimating time on treatment?

A fully parametric model was applied in the company base case submission for ease of review, adaptation and assessment of scenarios. Given the ERG preference for the semi-parametric model, the company would agree but advise that an earlier cut point would be more informative, as this would optimally balance modelling the heterogeneity of the population while providing as much data as possible to inform long-term extrapolations.

23. How long are people likely to remain on treatment with nivolumab in NHS practice?

Outside of adverse events, time on treatment is highly dependent on benefit to the patients. Patients who do not respond and progress rapidly will have short time on treatment. By contrast, those patients who respond and have good outcomes will receive longer term treatment

It should be noted that time on treatment curves are nearly closed, with median time on treatment of [redacted] months in the nivolumab arm and [redacted] of patients receiving treatment by two years. Hence, mean time on treatment should reflect this short time on treatment.

24. Is a stopping rule appropriate? If so, what stopping rule(s) are most relevant for NHS practice?

The SmPC for nivolumab specifies that treatment should be continued as long as clinical benefit is observed or until treatment is no longer tolerated by the patient.[40] In terms of immunotherapies, this means that treatment may be discontinued in patients with limited clinical benefit. However, it also refers to patients in whom maximum clinical benefit has been reached. Although no formal stopping rule was applied during ATTRACTION-3, clinicians and patients are aware that a stopping rule at two years is frequently applied for immunotherapies, and nivolumab

specifically. Hence, it is plausible that clinicians may informally apply this stopping rule in clinical practice, where patients have reached maximum clinical benefit.

During the undertaking of TA483[58] and TA484[59], the NICE Appraisal Committee noted that a 2-year stopping rule was not included in the pivotal trial or described in the SmPC and so queried whether clinicians would follow a stopping rule, especially if the patient was still benefitting from the treatment. When discussing the stopping rule, the committee noted comments on the second ACD that a two-year stopping rule is acceptable to both patients and clinicians and would be implementable. Further, the committee commented on the uncertainty of treatment effects following cessation of treatment but considered it biologically plausible for effects to continue, which may be up to three years, based on the available clinical evidence.[58]

Have the costs of comparator treatment been appropriately estimated?

25. Which source of cost estimates for medical technologies is most reflective of those paid by NHS trusts?

a. Company to clarify why was MIMS used as a tool to source treatment costs, given that eMIT prices are more reflective of those paid by NHS trusts?

The NICE reference case states that analyses based on price reductions for the NHS will only be considered when the reduced prices are transparent and consistently available across the NHS, and if the period for which the specified price is available is guaranteed.

Costs available the electronic Market Information Tool (eMIT) capture volume-based discounts provided to the NHS. It is unclear if all NHS trusts will have access to these medicines at this price, particularly due to the large standard deviation. Further, there is no confirmation these prices will be available for a guaranteed period. Based on this, the company deemed prices sourced from MIMs to be a more reliable and widely available source for treatment costs in UK clinical practice.

26. Is it appropriate to assume a 50:50 market share of taxanes?

a. Clinical experts to advise whether in clinical practice docetaxel is preferred over paclitaxel.

As outlined in Question 12, market research in 2019 indicated that paclitaxel represented 93.7% of treatment for OSCC. Table 5 shows the impact of different taxane combinations on cost-effectiveness outcomes, with all ICERs falling below the £50,000/QALY threshold. Whilst the company submission assumed a simple 50:50 market share of docetaxel and paclitaxel for the application of chemotherapy costs in the taxane

arm, assuming the costs are distributed as market research suggests with paclitaxel representing 93.7% of treatment options, the ICER falls to £43,668/QALY.

Table 5. Impact of different market shares of taxanes on ICER

Taxane combination	ICER (£/QALY)
100% paclitaxel	£43,405
6.3% docetaxel, 93.7% paclitaxel (representative of UK clinical practice)	£43,668
50% docetaxel, 50% paclitaxel (company base case)	£45,491
100% docetaxel	£47,578

Administration and medical resource use costs

27. Have the most appropriate sources been used to calculate administration and medical resource use costs? If so, which estimates of cost are most reliable?

a. Are the medical resource costs calculated by the company or ERG most relevant to NHS practice?

The clinician consultation cost, as stated in Table 70 of the company submission, is derived as a weighted average of costs for a consultant led and non-consultant led medical oncologist, using service codes 370. This yields a cost of £187.36.

The nerve block cost of £532.96 is sourced from the National Cost Collection 2018/19, *nerve block or destruction of nerve, for Pain Management*, total HRGs, currency code AB26Z.

For the cost of hospitalisation used in the health state cost derivation, the majority of the components had a length of stay exceeding one week, therefore it was deemed inappropriate to use a weighted average of these costs in their current form in order to avoid double counting. On this basis, the weighted average of the daily hospitalisation costs, calculated by dividing the costs by length of stay, were deemed more appropriate for use in the model.

End of life criteria

28. Is the estimate of extension to life robust? Is an extension to life of at least 3 months expected to be realised in NHS practice?

a. Is the company or ERG method of survival extrapolation most appropriate?

b. Could further follow-up data from ATTRACTION-3 be made available to support survival analysis?

c. Is there a subgroup within ATTRACTION-3 that would best reflect the health outcomes typically seen in NHS practice?

As noted by the ERG, ATTRACTION-3[17] demonstrated an added 2.5 months of median OS benefit for nivolumab versus taxanes. Median OS improvement is unlikely to provide an accurate reflection of the mean benefit accrued by an average patient, both in general (survival time distributions being generally skewed) and in the case of ATTRACTION-3[17] specifically due to the OS curves crossing at approximately four months. However, both BMS and the ERG predict mean OS benefit exceeding 3 months, as the submission base case analysis yields an estimated 7.8 months of additional mean OS for patients treated with nivolumab, while the ERG's base case analysis predicts that nivolumab is associated with an estimated 4.0 months of added mean OS.

The ERG notes that there is uncertainty as a result of limited follow-up. However, nivolumab provides an additional 2.5 months of median OS benefit in an indication where standard of care has median OS of 8.38 months. In addition to the crossover in the Kaplan-Meier, this indicates that mean OS benefit is likely to be significantly extended beyond 3 months. Further, when inappropriate extrapolations (defined as those with implausibly long mean survival or those with extrapolations that extend beyond the confidence intervals of the observed data) are excluded, predicted mean OS benefit is above 3 months in almost all scenario analyses exploring alternative survival extrapolations.

The ERG notes the uncertainty stemming from ATTRACTION-3 generalisability issues, which may impact the expected extension to survival attributable to nivolumab. To explore this aspect, an SLR was undertaken to identify studies reporting patient characteristics and treatment outcomes of oesophageal cancer patients in Asian and Western countries, specifically adult patients with advanced, metastatic or recurrent unresectable OSCC who were refractory, intolerant or resistant to first line therapy. Full results are provided in the response to Question 7, but the results indicate that outcomes in OSCC may be more comparable between populations, with discrepancies in the overall population driven by poorer outcomes in the oesophageal adenocarcinoma population, which is more common in Western populations. This in line with clinical expert opinion, which suggests that OSCC evidence is generalisable because risk factors are relatively comparable across ethnicities and the biology of OSCC is comparable between Asian and Western patients.

ATTRACTION-3 is an ongoing study and additional data will be provided when possible in order to support long-term survival estimates. Further, EAMS data will be provided to NICE when possible and can be used to validate generalisability of the data.

3 References

1. Borcoman E, Nandikolla A, Long G, Goel S, Le Tourneau C. Patterns of Response and Progression to Immunotherapy. American Society of Clinical Oncology Educational Book. 2018(38):169-78.
2. Chiou VL, Burotto M. Pseudoprogression and immune-related response in solid tumors. Journal of Clinical Oncology. 2015;33(31):3541-3.
3. Ono Pharmaceutical Co. Ltd. ONO-4538 Phase III Study. A Multicenter, Randomized, Open-label Study in Patients with esophageal Cancer refractory or intolerant to Combination Therapy with Fluoropyrimidine and Platinum-based Drugs. Protocol (ONO-4538-24). 2016.
4. Vokes EE, Ready N, Felip E, Horn L, Burgio MA, Antonia SJ, et al. Nivolumab versus docetaxel in previously treated advanced non-small-cell lung cancer (CheckMate 017 and CheckMate 057): 3-year update and outcomes in patients with liver metastases. Ann Oncol. 2018;29(4):959-65.
5. Gettinger S. WCLC 2019: Pooled Analysis of CheckMate 017 and 057: 5-Year Outcomes With Nivolumab vs Docetaxel in Previously Treated NSCLC 2019 [Available from: https://ascopost.com/news/september-2019/5-year-outcomes-with-nivolumab-vs-docetaxel-in-previously-treated-nsclc/.
6. Ferris RL, Blumenschein G, Fayette J, Guigay J, Colevas AD, Licitra L, et al. Nivolumab for Recurrent Squamous-Cell Carcinoma of the Head and Neck. New England Journal of Medicine. 2016;375(19):1856-67.
7. Ferris RL, Blumenschein G, Jr., Fayette J, Guigay J, Colevas AD, Licitra L, et al. Nivolumab vs investigator's choice in recurrent or metastatic squamous cell carcinoma of the head and neck: 2-year long-term survival update of CheckMate 141 with analyses by tumor PD-L1 expression. Oral Oncol. 2018;81:45-51.
8. Bristol-Myers Squibb. CheckMate 141 Data on File (15th October 2019).
9. Yen C-J, Kiyota N, Hanai N, Takahashi S, Yokota T, Iwae S, et al. Two-year follow-up of a randomized phase III clinical trial of nivolumab vs. the investigator's choice of therapy in the Asian population for recurrent or metastatic squamous cell carcinoma of the head and neck (CheckMate 141). Head & Neck. 2020;42(10):2852-62.
10. Herbst RS, Baas P, Kim D-W, Felip E, Pérez-Gracia JL, Han J-Y, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. The Lancet. 2016;387(10027):1540-50.
11. Rittmeyer A, Barlesi F, Waterkamp D, Park K, Ciardiello F, von Pawel J, et al. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. Lancet (London, England). 2017;389(10066):255-65.
12. Bellmunt J, de Wit R, Vaughn DJ, Fradet Y, Lee J-L, Fong L, et al. Pembrolizumab as Second-Line Therapy for Advanced Urothelial Carcinoma. New England Journal of Medicine. 2017;376(11):1015-26.

13. Clinical Trials.gov. Study of Pembrolizumab (MK-3475) Versus Investigator's Choice Standard Therapy for Participants With Advanced Esophageal/Esophagogastric Junction Carcinoma That Progressed After First-Line Therapy (MK-3475-181/KEYNOTE-181) 2020 [Available from: https://clinicaltrials.gov/ct2/show/results/NCT02564263?view=results.
14. Kim SB, Doi T, Kato K, Chen J, Shah M, Adenis A, et al. 124O - KEYNOTE-181: Pembrolizumab vs chemotherapy in patients (pts) with advanced/metastatic adenocarcinoma (AC) or squamous cell carcinoma (SCC) of the esophagus as second-line (2L) therapy. Annals of Oncology. 2019;30:ix42-ix3.
15. Cohen EEW, Soulières D, Le Tourneau C, Dinis J, Licitra L, Ahn M-J, et al. Pembrolizumab versus methotrexate, docetaxel, or cetuximab for recurrent or metastatic head-and-neck squamous cell carcinoma (KEYNOTE-040): a randomised, open-label, phase 3 study. The Lancet. 2019;393(10167):156-67.
16. Soulieres D, Cohen E, Tourneau CL, Dinis J, Licitra L, Ahn M-J, et al. Abstract CT115: Updated survival results of the KEYNOTE-040 study of pembrolizumab vs standard-of-care chemotherapy for recurrent or metastatic head and neck squamous cell carcinoma. Cancer Research. 2018;78(13 Supplement):CT115.
17. Kato K, Cho BC, Takahashi M, Okada M, Lin CY, Chin K, et al. Nivolumab versus chemotherapy in patients with advanced oesophageal squamous cell carcinoma refractory or intolerant to previous chemotherapy (ATTRACTION-3): a multicentre, randomised, open-label, phase 3 trial. Lancet Oncol. 2019;20(11):1506-17.
18. Davidson M, Cafferkey C, Goode EF, Kouvelakis K, Hughes D, Reguera P, et al. Survival in Advanced Esophagogastric Adenocarcinoma Improves With Use of Multiple Lines of Therapy: Results From an Analysis of More Than 500 Patients. Clinical Colorectal Cancer. 2018;17(3):223-30.
19. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013 2013 [Available from: https://www.nice.org.uk/article/pmg9.
20. Moriwaki T, Kajiwara T, Matsumoto T, Suzuki H, Hiroshima Y, Matsuda K, et al. Survival analysis of platinum-refractory patients with advanced esophageal cancer treated with docetaxel or best supportive care alone: a retrospective study. Dis Esophagus. 2014;27(8):737-43.
21. Cheng S, Qureshi M, Pullenayegum E, Haynes A, Chan KKW. Do patients with reduced or excellent performance status derive the same clinical benefit from novel systemic cancer therapies? A systematic review and meta-analysis. ESMO Open. 2017;2(4):e000225.
22. Arnold M, Soerjomataram I, Ferlay J, Forman D. Global incidence of oesophageal cancer by histological subtype in 2012. Gut. 2015;64(3):381-7.
23. Lu CL, Lang HC, Luo JC, Liu CC, Lin HC, Chang FY, et al. Increasing trend of the incidence of esophageal squamous cell carcinoma, but not adenocarcinoma, in Taiwan. Cancer Causes Control. 2010;21(2):269-74.
24. Zhang H-Z, Jin G-F, Shen H-B. Epidemiologic differences in esophageal cancer between Asian and Western populations. Chin J Cancer. 2012;31(6):281-6.
25. Pakzad R, Mohammadian-Hafshejani A, Khosravi B, Soltani S, Pakzad I, Mohammadian M, et al. The incidence and mortality of esophageal cancer and their relationship to development in Asia. Ann Transl Med. 2016;4(2):29.
26. World Health Organization International Agency for Research on Cancer. Cancer Today: Cancer Fact Sheets 2018 [Available from: https://gco.iarc.fr/today/fact-sheets-cancers.

27. Bang YJ, Van Cutsem E, Feyereislova A, Chung HC, Shen L, Sawaki A, et al. Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. Lancet (London, England). 2010;376(9742):687-97.
28. Kang YK, Kang WK, Shin DB, Chen J, Xiong J, Wang J, et al. Capecitabine/cisplatin versus 5-fluorouracil/cisplatin as first-line therapy in patients with advanced gastric cancer: a randomised phase III noninferiority trial. Annals of Oncology. 2009;20(4):666-73.
29. National Institute for Health and Care Excellence. Capecitabine for the Treatment of Advanced Gastric Cancer 2009 [Available from: https://www.nice.org.uk/guidance/ta191/documents/gastric-cancer-advanced-capecitabine-manufacturer-submission-roche2.
30. National Institute for Health and Care Excellence. Trastuzumab for the treatment of HER2 positive metastatic adenocarcinoma of the stomach or gastro-oesophageal junction (mGC) 2010 [Available from: https://www.nice.org.uk/guidance/ta208/resources/gastric-cancer-advanced-her2-positive-trastuzumab-herceptin-roche-pharmaceuticals2.
31. National Institute for Health and Care Excellence. Ramucirumab for treating advanced gastric cancer or gastro–oesophageal junction adenocarcinoma previously treated with chemotherapy (TA378). 2016.
32. Wilke H, Muro K, Van Cutsem E, Oh SC, Bodoky G, Shimada Y, et al. Ramucirumab plus paclitaxel versus placebo plus paclitaxel in patients with previously treated advanced gastric or gastro-oesophageal junction adenocarcinoma (RAINBOW): a double-blind, randomised phase 3 trial. Lancet Oncol. 2014;15(11):1224-35.
33. Lordick F, Mariette C, Haustermans K, Obermannová R, Arnold D. Oesophageal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of Oncology. 2016;27(suppl 5):v50-v7.
34. National Comprehensive Cancer Network. Esophageal and Esophagogastric Junction Cancers. NCCN Clinical Practice Guidelines in Oncology. Version 2.2018. 2018 11 September 2018. Available from: https://www.nccn.org/professionals/physician_gls/f_guidelines.asp.
35. Ford HER, Marshall A, Bridgewater JA, Janowitz T, Coxon FY, Wadsley J, et al. Docetaxel versus active symptom control for refractory oesophagogastric adenocarcinoma (COUGAR-02): an open-label, phase 3 randomised controlled trial. The Lancet Oncology. 2014;15(1):78-86.
36. Wu Y-L, Lu S, Cheng Y, Zhou C, Wang J, Mok T, et al. Nivolumab Versus Docetaxel in a Predominantly Chinese Patient Population With Previously Treated Advanced NSCLC: CheckMate 078 Randomized Phase III Clinical Trial. Journal of Thoracic Oncology. 2019;14(5):867-75.
37. National Institute for Health and Care Excellence. Pembrolizumab for treating PD-L1-positive non-small-cell lung cancer after platinumbased chemotherapy [ID840] 2016 [Available from: https://www.nice.org.uk/guidance/ta428/documents/committee-papers.
38. National Institute for Health and Care Excellence. Atezolizumab for treating non-small-cell lung cancer after platinum-based chemotherapy [ID970] 2017 [Available from: https://www.nice.org.uk/guidance/ta520/documents/committee-papers.
39. National Institute for Health and Care Excellence. Pembrolizumab for previously treated advanced or metastatic urothelial cancer 2017 [Available from: https://www.nice.org.uk/guidance/ta519/documents/committee-papers.
40. Bristol-Myers Squibb Pharmaceutical Limited. Summary of Product Characteristics. OPDIVO 10 mg/mL concentrate for solution for infusion 2017 [Available from: https://www.medicines.org.uk/emc/medicine/30476.

41. Latimer N. NICE DSU TECHNICAL SUPPORT DOCUMENT 14: SURVIVAL ANALYSIS FOR ECONOMIC EVALUATIONS ALONGSIDE CLINICAL TRIALS - EXTRAPOLATION WITH PATIENT-LEVEL DATA 2013 [Available from: http://nicedsu.org.uk/wp-content/uploads/2016/03/NICE-DSU-TSD-Survival-analysis.updated-March-2013.v2.pdf.
42. Bagust A, Beale S. Survival Analysis and Extrapolation Modeling of Time-to-Event Clinical Trial Data for Economic Evaluation An Alternative Approach. Medical Decision Making. 2014;34(3):343-51.
43. NHS Improvement. 2018/19 National Cost Collection data 2020 [Available from: https://improvement.nhs.uk/resources/national-cost-collection/#ncc1819.
44. National Institute for Health and Care Excellence. Pembrolizumab for treating advanced melanoma after disease progression with ipilimumab - Technology Appraisal [TA357] 2017 [Available from: https://www.nice.org.uk/guidance/ta357.
45. National Institute for Health and Care Excellence. Pembrolizumab for advanced melanoma not previously treated with ipilimumab-Technology Appraisal [TA366] 2015 [Available from: https://www.nice.org.uk/Guidance/TA366.
46. National Institute for Health and Care Excellence. Atezolizumab with carboplatin and etoposide for untreated extensive-stage small-cell lung cancer - Technology Appraisal [TA638] 2020 [Available from: https://www.nice.org.uk/guidance/ta638.
47. National Institute for Health and Care Excellence. Pembrolizumab for untreated PD-L1-positive metastatic non-small-cell lung cancer - Technology appraisal [TA531] 2018 [Available from: https://www.nice.org.uk/guidance/ta531.
48. National Institute for Health and Care Excellence. Pembrolizumab with pemetrexed and platinum chemotherapy for untreated, metastatic, non-squamous non-small-cell lung cancer - Technology Appraisal [TA557] 2019 [Available from: https://www.nice.org.uk/guidance/ta557.
49. Hatswell AJ, Pennington B, Pericleous L, Rowen D, Lebmeier M, Lee D. Patient-reported utilities in advanced or metastatic melanoma, including analysis of utilities by time to death. Health Qual Life Outcomes. 2014;12:140.
50. Grothey A, Van Cutsem E, Sobrero A, Siena S, Falcone A, Ychou M, et al. Regorafenib monotherapy for previously treated metastatic colorectal cancer (CORRECT): an international, multicentre, randomised, placebo-controlled, phase 3 trial. Lancet (London, England). 2013;381(9863):303-12.
51. National Institute for Health and Care Excellence. Nivolumab for previously treated locally advanced or metastatic non-squamous non-small-cell lung cancer [ID900].In development [GID-TAG524].2017 1 September 2017. Available from: https://www.nice.org.uk/guidance/indevelopment/gid-tag524.
52. National Institute for Health and Care Excellence. Nivolumab for previously treated locally advanced or metastatic squamous non-small cell lung cancer [ID811]. Single Technology Appraisal: Evidence Review Group Report. In development [GID-TAG506].2015 14 September 2017. Available from: https://www.nice.org.uk/guidance/gid-tag506/documents/committee-papers
53. National Institute for Health and Care Excellence. Technology appraisal guidance [TA462]. Nivolumab for treating relapsed or refractory classical Hodgkin lymphoma. 2017 14 September 2017. Available from: https://www.nice.org.uk/guidance/ta462.
54. National Institute for Health and Care Excellence. Technology appraisal guidance [TA384]. Nivolumab for treating advanced (unresectable or metastatic) melanoma.2016 14 September 2017. Available from: https://www.nice.org.uk/guidance/ta384.
55. National Institute for Health and Care Excellence. Technology appraisal guidance [TA400]. Nivolumab in combination with ipilimumab for treating advanced melanoma. 2016 14 September 2017. Available from: https://www.nice.org.uk/guidance/ta400.

56. National Institute for Health and Care Excellence. Technology appraisal guidance [TA417]. Nivolumab for previously treated advanced renal cell carcinoma. 2016 14 September 2017. Available from: https://www.nice.org.uk/guidance/ta417.
57. National Institute for Health and Care Excellence. Nivolumab with ipilimumab for untreated advanced renal cell carcinoma - Technology Appraisal [TA581] 2019 [Available from: https://www.nice.org.uk/guidance/ta581.
58. National Institute for Health and Care Excellence. Nivolumab for previously treated squamous non-small-cell lung cancer. Technology appraisal guidance [TA483].2017 14 November 2017. Available from: https://www.nice.org.uk/guidance/ta483.
59. National Institute for Health and Care Excellence. Nivolumab for previously treated non-squamous non-small-cell lung cancer. Technology appraisal guidance [TA484].2017 14 November 2017. Available from: https://www.nice.org.uk/guidance/ta484.

NICE National Institute for Health and Care Excellence

Technical engagement response form

Nivolumab for previously treated unresectable advanced oesophageal cancer [ID1249]

As a stakeholder you have been invited to comment on the technical report for this appraisal. The technical report and stakeholders responses are used by the appraisal committee to help it make decisions at the appraisal committee meeting. Usually, only unresolved or uncertain key issues will be discussed at the meeting.

We need your comments and feedback on the questions below. You do not have to answer every question. The text boxes will expand as you type. Please read the notes about completing this form. We cannot accept forms that are not filled in correctly. Your comments will be summarised and used by the technical team to amend or update the scientific judgement and rationale in the technical report.

Deadline for comments **17 September 2020**

Thank you for your time.

Please log in to your NICE Docs account to upload your completed form, as a Word document (not a PDF).

Notes on completing this form

- Please see the technical report which summarises the background and submitted evidence. This will provide context and describe the questions below in greater detail.
- Please do not embed documents (such as PDFs or tables) because this may lead to the information being mislaid or make the response unreadable. Please type information directly into the form.
- Do not include medical information about yourself or another person that could identify you or the other person.
- Do not use abbreviations.
- Do not include attachments such as journal articles, letters or leaflets. For copyright reasons, we will have to return forms that have attachments without reading them. You can resubmit your form without attachments, but it must be sent by the deadline.
- If you provide journal articles to support your comments, you must have copyright clearance for these articles.
- Combine all comments from your organisation (if applicable) into 1 response. We cannot accept more than 1 set of comments from each organisation.
- Please underline all confidential information, and separately highlight information that is submitted under 'commercial in confidence' in turquoise, all information submitted under 'academic in confidence' in yellow, and all information submitted under 'depersonalised data' in pink. If confidential

information is submitted, please also send a second version of your comments with that information replaced with the following text: 'academic/commercial in confidence information removed'. See the Guide to the processes of technology appraisal (sections 3.1.23 to 3.1.29) for more information.

We reserve the right to summarise and edit comments received during engagement, 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 engagement 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.

About you

Your name	Elizabeth Smyth
Organisation name – stakeholder or respondent (if you are responding as an individual rather than a registered stakeholder please leave blank)	Royal College of Physicians
Disclosure Please disclose any past or current, direct or indirect links to, or funding from, the tobacco industry.	None

Questions for engagement

Issue 1: Clinical effectiveness of nivolumab	
1. Is the relative treatment effect of nivolumab compared with taxanes likely to be constant over time?	No, the benefit from nivolumab is likely to increase over time. The trajectory of benefit for chemotherapy versus immunotherapy is different. The benefit from chemotherapy is immediate but is short lived, as evidenced by immediate six week improvement in overall survival for chemotherapy in second line oesophageal cancer. The benefit from immunotherapy takes longer to appear, but is more sustained when present.
2. How does nivolumab provide overall survival benefit without improvements in progression-free survival or overall response rate	Response rate to chemotherapy vs. immunotherapy have different meanings. With chemotherapy, response is an immediate reduction in the number of rapidly dividing cells caused by cytotoxic cell death. This will provide a rapid reduction in tumour burden. However, the response to cytotoxic chemotherapy is short lived. In contrast, response to immunotherapy relies on activation of the immune system. Initially, pseudo progression may occur as immune cells infiltrate the tumour. The presence of response on imaging may take longer to detect with immunotherapy than with cytotoxic chemotherapy. It is for this reason that separate imaging guidelines (irRECIST) have been developed for immunotherapy. Progression free survival is not an accurate or adequate metric to measure the efficacy of immunotherapy (https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2685626) The benefit is in the long term and alternative measures for benefit have been suggested such as landmark survival to best assess this. Similar differences in PFS and OS in practice changing trials were also observed in second line lung cancer trials for NICE approved therapies (https://www.nejm.org/doi/full/10.1056/nejmoa1507643).

Issue 2: Is best supportive care a relevant comparator?	
3. Is nivolumab expected to completely replace taxane use for the given population?	There will always be patients for whom nivolumab is unsuitable, for example those with autoimmune diseases or transplants. However nivolumab would be expected to replace paclitaxel in the majority of cases.
4. In current NHS practice, what treatment would nivolumab-eligible patients be assigned to?	Paclitaxel, docetaxel or irinotecan.
5. What percentage of patients currently receive docetaxel, paclitaxel and irinotecan in NHS practice?	Approximately 30-40% of patients receive second line chemotherapy for oesophageal cancer. Of these, the majority would be treated with paclitaxel and docetaxel and irinotecan are less well tolerated.
6. Given the limitations of the indirect treatment comparison, is it appropriate to compare nivolumab with best supportive care in people who are not eligible for further chemotherapy?	Yes, nivolumab could be considered in patients who are not eligible for further chemotherapy.
Issue 3: Generalisability of ATTRACTION-3 results	
7. Given the characteristics of the study population, are the results from the ATTRACTION-3 trial generalisable in a UK decision-making context?	The causes of squamous cancer of the oesophagus are universal, most frequently alcohol, tobacco and hot beverages. There is no evidence in The Cancer Genome Atlas (TCGA) that squamous oesophaeal cancer from different regions of the world have a different biology https://www.nature.com/articles/nature20805. Therefore, the underlying mutational profile and sensitivity to immunotherapy should be similar in both populations. The age profile of patients in the study is common to all studies, to state that this is not generalisable would be to say that no trial is ever generalisable. A statement is made that different treatments to UK patients are recommended in the ESMO-JSMO guidelines (eg. nedaplatin) and that this could mean that UK patients might not benefit from the nivolumab. As an author of the ESMO-JSMO guidelines I can confirm that nedaplatin is a cisplatin analogue which directly comparable to cisplatin and oxaliplatin

	(drugs used in the UK), and that cisplatin is most commonly used in Asia. The treatment for squamous cancer of the oesophagus shows very little variation globally therefore I disagree that differences in the treatment pathway invalidate these results for UK patients. A statement is made that "Participants in ATTRACTION-3 are also fitter and more able to tolerate treatment with nivolumab and taxanes than people in UK practice". The indication for nivolumab would be as per the trial – PS 0-1 good performance status patients. There is no reason to believe that a UK PS 0-1 patient treated with nivolumab should not benefit from the drug. If more UK patients are PS2, then they are not eligible. It is important not to conflate general population fitness with those eligible for treatment. A statement is made that Japanese patients live longer than UK patients and therefore the results of the trial cannot be generalised to UK patients. It is generally accepted in all trials of gastric cancer and oesophageal cancer (and also lung cancer) that Japanese patients live longer which is why recruitment is frequently capped in trials. However, this does not mean that no survival benefit is derived for non-Asian patients.
8. Would it be more appropriate to use efficacy data from the rest of world population compared with the intention to treat population to estimate the clinical and cost effectiveness in clinical practice?	No, the Japanese and "ROW" population are essentially identical as they are closely located East Asian countries that share treatment approaches.
Issue 4: Safety data for nivolumab, early deaths higher on nivolumab	
9. To what extent does the company's rationale provide an explanation for the early deaths observed in the nivolumab arm?	I agree with the company's assertion. The early death rate is due to slower responses to nivolumab than chemotherapy. Unfortunately, when patients with advanced oesophageal cancer progress on second line treatment, they are very likely to die in a short period. So, if a treatment does not have an immediate effect, there will be a proportion of patients who progress quickly, and sadly pass away. Regarding whether this will be seen in NHS practice, it is possible that it will not. We now understand that because of the mode of action of immunotherapy that treatment of patients with very advanced disease, for example large volume metastases, is unlikely to be helpful. Therefore

	there may be a pre-selection of patients who are more likely to benefit based on experience, whereas in a clinical trial everyone gets treated if they meet inclusion criteria.
Issue 5: No adjustment to efficacy for any beneficial effects of third-line therapy	
10. In clinical practice, is active third-line therapy administered following progression after the use of a taxane or nivolumab?	It depends on the patient. Generally in the UK, the proportion of patients who receive treatment for third line is likely to be ~15%. If a patient is fit after second line treatment, there is no reason not to provide treatment as long as that is considered appropriate and what the patient wants.
11. Should the efficacy of nivolumab be adjusted to account for beneficial effects of third-line therapy?	No, comparable numbers of patients treated with chemotherapy and nivolumab in ATTRACTION-3 received third line treatment. It would not have been ethical to deny patients another effective treatment after a trial, and data is not routinely collected for post-trial treatment. This would add massively to the costs of clinical trials and there are too many confounders for this to be reliable.
Issue 6: Subgroup analysis by taxane was not provided	
12. Is a comparison of nivolumab with individual taxane or combined taxanes more relevant to NHS practice? Which taxane is most commonly used in the NHS?	Paclitaxel and docetaxel are both used. My personal preference is paclitaxel as it is associated with fewer side effects, but some prefer docetaxel as visits are less frequent. Taxanes can be considered equivalent in efficacy.
13. Are health outcomes expected to be different between 'docetaxel-preferred' and 'paclitaxel-preferred' populations?	No reason to expect this, if different taxanes are chosen it is more likely to be due to institutional preference or personal choice of patient/oncologist.
Issue 7: Difference between company and ERG ICER	
14. Which modelling methods are most suitable for estimating overall survival and expected time on treatment?	General replies to statements • Patients should not be were aware of their treatment allocation at screening visit.

	• Pre-progression utility could be higher than UK baseline population if trial patient is generally fitter than UK population? • It makes sense that pre-progression utility would decline with chemotherapy due to chemotherapy related side effects. • I disagree with a treatment independent approach, clearly treatment can effect utility
15. Which estimates of treatment, administration and medical resource use costs are most reliable?	These could be taken from a selection of CCG or Trusts and averaged
Issue 8: Model time horizon (40 years in company base-case)	
16. What is the most appropriate time horizon for the economic model?	No comment – I lack expertise in this area.
Issue 9: Alternative extrapolations for overall survival	
17. Is a fully-parametric or semi-parametric model most appropriate to predict the long term effectiveness of nivolumab and taxane therapy?	Semi-parametric to reflect the two phase model of outcome: in the early stage an advantage to chemotherapy in response rate and PFS followed by late stage: advantage to nivolumab.
18. If a semi-parametric model is preferred, is 2.99 months an appropriate point to start the extrapolation?	Yes, the point at which 50% of patients have progressed in both arms.
19. Does the overall survival predicted in the company ([redacted] at 10 years) or ERG model (0.2% at 10 years) best represent the likely overall survival in NHS practice?	I believe that even the company model is conservative. We see a small proportion of patients who are extremely sensitive to immunotherapy and enjoy long term survival or essentially cure. It could be 5% or greater.
Issue 10: Exploratory analysis of utility values	

20. Are the differences in utility between the nivolumab and taxane arm in the company model clinically plausible?	Yes, as per my answer above. Chemotherapy has devastating side effects and even low grade toxicity can lead to fatigue and decreased QoL. In contrast, the vast majority of patients on immunotherapy feel quite normal.
21. Would a treatment independent approach (e.g. using the same values for pre and post progression across the treatment arms) be more appropriate?	No, as per my answer above.
Issue 12: Alternative extrapolations of time on treatment	
22. Are fully-parametric or semi-parametric methods most appropriate for estimating time on treatment?	Semi-parametric.
23. How long are people likely to remain on treatment with nivolumab in NHS practice?	Until definitive radiological progression or toxicity. The median would be based on the trial.
24. Is a stopping rule appropriate? If so, what stopping rule(s) are most relevant for NHS practice?	Some immunotherapy drugs are recommended to stop at 2 years. The argument is that the patient will continue to respond without the drug. We do see relapses after this so I would argue that for the minority of patients who are still on study then I would continue. There is a difference in this case where perhaps in melanoma a large number of patients would be on treatment but in oesophageal squamous cancer this number would be small.
Issue 13: Have the costs of comparator treatment been appropriately estimated?	
25. Which source of cost estimates for medical technologies is most reflective of those paid by NHS trusts?	I lack expertise in this area, but if eMIT takes an average from a large number of Trusts, then it seems a reasonable choice.
26. Is it appropriate to assume a 50:50 market share of taxanes?	It may be, the SACT dataset might be helpful in this regard. I disagree that the preference is docetaxel.

Issue 14: Administration and medical resource use costs	
27. Have the most appropriate sources been used to calculate administration and medical resource use costs? If so, which estimates of cost are most reliable?	I am not an expert in this area, but the ERG calculated cost per hospitalisation appears to me at the upper end of what would be considered possible. Using a previous NICE TA seems reasonable as a source of costs, however these should be updated.
Issue 15: End of life criteria	
28. Is the estimate of extension to life robust? Is an extension to life of at least 3 months expected to be realised in NHS practice?	Yes, by selecting the correct patients for treatment nivolumab is likely to lead to meaningful improvement in quality of life, without the attendant toxicity of chemotherapy.

Nivolumab for unresectable, advanced oesophageal cancer when standard chemotherapy has failed [ID1249]

A Single Technology Appraisal

ERG Review of Company's Response to Technical Engagement

Produced by

Peninsula Technology Assessment Group (PenTAG)
University of Exeter Medical School
South Cloisters
St Luke's Campus
Heavitree Road
Exeter
EX1 2LU

Authors

Mr Ash Bullement, Associate[1] and Analyst[2]
Dr Linda Long, Research Fellow[1]
Dr Kevin Deighton, Associate[1] and Analyst[2]
Ms Naomi Shaw, Information Specialist[1]
Dr Stephen Falk, Consultant Clinical Oncologist[3]
Dr Nicole Dorey, Consultant Clinical Oncologist[4]
Ms Louise Crathorne, Senior Research Fellow[1]
Prof G.J. Melendez-Torres, Professor[1]
Dr Maxwell S. Barnish, Research Fellow[1]

[1] Peninsula Technology Assessment Group (PenTAG), University of Exeter Medical School, Exeter
[2] Delta Hat Limited, Nottingham
[3] Bristol Haematology and Oncology Centre, Bristol
[4] Royal Devon and Exeter NHS Foundation Trust, Exeter

Correspondence to

Dr Maxwell S. Barnish
3.09f South Cloisters, St Luke's Campus, Heavitree Road, Exeter, EX1 2LU; m.s.barnish@exeter.ac.uk

Source of funding	This report was commissioned by the NIHR Systematic Reviews Programme as project number 18/54/01.
Declared competing interests of the authors	Dr Dorey declares educational support from Boehringer-Ingelheim; Roche; Lilly and Astra-Zeneca.
Rider on responsibility for document	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 TE response is linked to ERG report	Bullement A, Long L, Deighton K, Shaw N, Falk S, Dorey N, Crathorne L, Melendez-Torres GJ, Barnish MS. Nivolumab for unresectable, advanced oesophageal cancer when standard chemotherapy has failed [ID1249]: A Single Technology Appraisal. Peninsula Technology Assessment Group (PenTAG), 2020.

1. INTRODUCTION

This document provides the Evidence Review Group's (ERG's) critique of the company's response to the technical engagement report produced by the National Institute for Health and Care Excellence (NICE) for the appraisal of Nivolumab for unresectable, advanced oesophageal cancer when standard chemotherapy has failed (ID1249). Each of the issues outlined in the technical report are discussed in further detail in Section 2. The company has not provided a revised PAS, presented any additional data or made any changes to the economic model as part of its response to technical engagement. Changes to the ERG's base case are presented in Section 3.

2. ERG REVIEW OF KEY ISSUES

Issue 1: Clinical effectiveness of nivolumab

1. Is the relative treatment effect of nivolumab compared with taxanes likely to be constant over time?

The Evidence Review Group (ERG) considers that there is uncertainty over whether the relative treatment effect of nivolumab compared with taxanes is likely to be constant over time. The ERG highlights that the effect of nivolumab could be quantified in several different ways, one of which could be the hazard ratio (HR). The company highlights within its submission that the proportional hazards assumption was violated, and so it may be inferred that a constant treatment effect expressed as a HR is unlikely to hold over time.

The ERG notes the response from the Royal College of Physicians (RCP) representative Dr Elizabeth Smyth that the benefit from nivolumab is likely to increase over time, as a result of a different trajectory of benefit for immunotherapies compared to chemotherapies (in immunotherapies the benefit typically takes longer to emerge but is more sustained). The ERG considers this to be plausible.

2. How does nivolumab provide overall survival benefit without improvements in progression-free survival or overall response rate

The ERG notes that the evidence for a significant benefit of nivolumab compared to taxanes in the ATTRACTION-3 trial is more convincing when overall survival (OS) is considered than when progression-free survival (PFS) or overall response rate (ORR) are considered. The presence of response on imaging may be delayed for immunotherapies, due to pseudoprogression and/or differences in the trajectory of benefit. PFS as an outcome is not especially probative for efficacy in the context of immunotherapies,[1] especially acknowledging that response in ATTRACTION-3 was not measured according to the iRECIST criteria.

Issue 2: Is best supportive care a relevant comparator?

3. Is nivolumab expected to be completely replace taxane use for the given population?

The ERG considers that nivolumab, if approved, would replace taxane use in this setting for the majority of patients. However, the ERG agrees with the RCP representative Dr Elizabeth Smyth that there will always be some patients; for example, those with autoimmune diseases or transplants, for whom nivolumab will be unsuitable. Therefore, the ERG does not expect that nivolumab would completely replace taxane use for the given population.

4. In current NHS practice, what treatment would nivolumab-eligible patients be assigned to?

Clinical advice to the ERG has stated that in current NHS practice, nivolumab-eligible patients would typically be assigned to a taxane, either docetaxel or paclitaxel. The balance of each of these treatments in the patient population is addressed in response to Question 5 below.

However, there may be some patients in NHS practice considered eligible for nivolumab but not considered eligible for taxanes. Therefore, some patients deemed eligible to receive nivolumab would currently receive BSC in practice. The ERG highlights, however, that this group is not represented in the ATTRACTION-3 population (given that all patients were required to be taxane eligible). This point is discussed further in the ERG's response to Question 6.

5. What percentage of patients currently receive docetaxel, paclitaxel and irinotecan in NHS practice?

5a. Is irinotecan used in clinical practice for patients who would be eligible for nivolumab?

5b. What proportion of people currently receive docetaxel or paclitaxel in NHS practice?

Clinical advice to the ERG has stated that irinotecan is not typically prescribed for this indication in current NHS practice. The ERG has been advised by its clinical experts that docetaxel would be used as the preferred treatment option for most NHS patients, due to the greater convenience and resource use savings resulting from less frequent administration (i.e. once every three weeks instead of weekly). The ERG was, however, advised that paclitaxel may be preferred for a minority of patients due to a more favourable safety profile. The ERG acknowledges that the clinical advice it received that docetaxel was the preferred treatment option for most patients does not accord with the company's market research or the response provided by Dr. Elizabeth Smyth (representative of the RCP), which showed a preference for paclitaxel. The ERG therefore cannot rule out the possibility of a more general preference for paclitaxel, or a 50:50 split, at an England- and Wales-wide level. These differences may partly result from regional variation, with the ERG's advice reflecting practice in two major centres in the South West. The ERG therefore does not consider it has the information necessary to state accurately what the split between docetaxel and paclitaxel would be at a national level.

6. Given the limitations of the indirect treatment comparison, is it appropriate to compare nivolumab with best supportive care in people who are not eligible for further chemotherapy?

6a. Is best supportive care a relevant comparator? Can the efficacy of nivolumab in ATTRACTION-3 be generalised to people for whom best supportive care is suitable? Is the ITC robust given the heterogeneity between studies?

6b. Are the results of the ITC generalisable to NHS practice?

The ERG does not disagree that BSC may be a relevant comparator to the extent that it is always a treatment option available to clinicians. However, it remains uncertain that the efficacy of nivolumab can be generalised to people for whom BSC is suitable given likely differences in patient populations.

In their response, the company reiterate the methodological decisions taken in order to estimate an indirect treatment comparison (ITC). These methodological decisions, while unclear in their original presentation and as documented in the original ERG report, were in the main more reasonable than not. However, the rigour of these decisions does not alter the fundamental issues with the ITC; namely, a lack of robustness in included studies, serious issues relating to transitivity, and incommensurability of outcome estimates. Indeed, as the ERG noted in its original response, the only aspect of the ITC that is used to inform decision-making relies on a single study, for one outcome, drawing on non-randomised evidence. Thus, the ERG maintains that the ITC is not robust, and thus its generalisability to NHS practice is on the one hand unlikely (based on the treatment context for the singular study), and on the other hand unknowable (based on the sparsity of evidence).

Issue 3: Generalisability of ATTRACTION-3 results

7. Given the characteristics of the study population, are the results from the ATTRACTION-3 trial generalisable in a UK decision-making context?

7a. Clinical expert clarification on whether the rest of the world efficacy data is more relevant to the expected effectiveness of nivolumab in the NHS practice

7b. Who is likely to receive nivolumab in clinical practice? Will it be restricted to people with a good performance status (ECOG 0-1)? Is the efficacy of nivolumab in NHS practice (people with a worse performance status) likely to be worse than in the ATTRACTION-3?

With regard to the rest of the world (ROW) data, this is addressed below in response to Question 8. Clinical advice to the ERG indicated that nivolumab would likely only be considered in patients with a good performance status (PS) (Eastern Cooperative Oncology Group [ECOG]

PS 0-1). Patients with poor PS would be less likely to be able to tolerate treatment with nivolumab, and likely to experience worse efficacy if treated.

8. Would it be more appropriate to use efficacy data from the rest of world population compared with the intention to treat population to estimate the clinical and cost effectiveness in clinical practice?

8a. Is there a subgroup within ATTRACTION-3 which is considered to have similar characteristics to the UK patient population?

The ERG acknowledges the limitation to generalisability to a UK clinical practice context posed by the very high proportion of Asian patients (96%) in the pivotal ATTRACTION-3 trial. Additionally, around two thirds of all patients in this trial are Japanese. The ERG acknowledges the potential benefit of using the ROW (i.e. non-Japanese) population, through mitigating against any Japan-specific effects. Nevertheless, the benefit of this approach is reduced by the fact that the population would remain almost entirely Asian. As its name suggests, the Pan-Asian version of the ESMO guidelines apply across Asia, and this may introduce systematic differences in treatment pathways compared to a European, or specifically UK, context. In particular, there are treatments for the current indication, such as nedaplatin, that are recommended by the Pan-Asian ESMO guidelines but not in UK clinical practice. Therefore, the ERG considers that there are substantial limitations in the generalisability of treatment pathways in ATTRACTION-3 to a UK context, and that using the ROW subgroup is unlikely to completely resolve this issue, although may have a small benefit.

Issue 4: Safety data for nivolumab, early deaths higher on nivolumab

9. To what extent does the company's rationale provide an explanation for the early deaths observed in the nivolumab arm?

9a. Do clinical experts agree with the company's rationale for higher death rate over first 3 months in the nivolumab arm?

9b. Clinical expert advice is sought on whether the initial higher death rate with nivolumab in ATTRACTION-3 is likely to be seen in NHS practice.

9c. Is it possible to determine in advance which patients are likely to die before they can benefit from treatment with nivolumab (if so please explain)?

In the company submission, the company refer to Table 29 for data describing deaths that occurred between the start date of the first administration of the investigational product and either the date 28 days after the end of the treatment period or the start date of post-study treatment after the end of the treatment period (whichever was earlier). In the ERG report, the ERG have used the term 'early deaths' to describe these values [redacted] in the nivolumab arm

and [redacted] in the control arm). However, the ERG's main concern regarding these deaths is with 'deaths in the first three months' vs 'on treatment deaths'. The ERG identified that deaths in the first three months were considerably higher on nivolumab than on comparator taxanes, contrary to the overall pattern of a superior safety profile for nivolumab. The ERG sought further clarification from the company regarding the reason for the increase in 'deaths in the first three months' vs 'on treatment deaths'. The company responded (clarification question A13), linking the increase in 'deaths in the first three months' vs 'on treatment deaths' to the mechanism of action of nivolumab. The RCP representative Dr Elizabeth Smyth supports the company's assertion that this is a class effect related to immunotherapies due to a different trajectory of benefit for immunotherapies compared to chemotherapies, including slower initial benefit.

Issue 5: No adjustment to efficacy for any beneficial effects of third-line therapy

10. In clinical practice, is active third-line therapy administered following progression after the use of a taxane or nivolumab?

10a. Clinical experts to clarify whether patients in NHS practice receive active treatment after progression on nivolumab or taxanes. What proportion of patients receive different 3rd line treatments?

10b. Is nivolumab expected to completely replace taxane use for the proposed indication and population? If so, would taxane therapy then be offered to nivolumab patients post-progression?

Clinical advice to the ERG indicated that nivolumab would be generally expected to replace taxane use for this indication, although there would be some patients; for example, those with autoimmune diseases or transplants, for whom nivolumab will be unsuitable. The ERG was advised that nivolumab would not likely be used post-progression, although taxanes may be used if the patient remains fit enough, and best supportive care/ radiotherapy are also options in this context. The ERG does not have access to information regarding what proportion of patients would be expected to receive each available third-line treatment.

11. Should the efficacy of nivolumab be adjusted to account for beneficial effects of third-line therapy?

The ERG agrees with the company with respect to the uncertainty surrounding the composition of third-line treatment in current NHS practice, and that consequently it is difficult to understand how costs and outcomes may be robustly adjusted to reflect the differences between the ATTRACTION-3 and NHS patient populations. However, it is the ERG's view that an

assumption of zero impact of third-line treatments on both the costs and effects likely leads to an inaccurate estimate of the 'true' incremental cost-effectiveness ratio (ICER).

The ERG would consider it useful to explore scenario analyses to understand the directional effect on the ICER were additional costs to be added to both treatment arms to reflect third-line therapy (which would act in favour of the cost-effectiveness of nivolumab, relative to the current base-case analysis). Equivalently, scenarios wherein outcomes were reduced to 'remove' the impact of third-line treatment on outcomes would also be useful (which would act against the cost-effectiveness of nivolumab, relative to the current base-case analysis).

In spite of this, the ERG accepts that any adjustment to outcomes to 'remove' the beneficial effects of third-line treatment would be highly uncertain, and would emphasize that any scenarios produced in relation to this would unavoidably need to be considered with caution.

Issue 6: Subgroup analysis by taxane was not provided

12. Is a comparison of nivolumab with individual taxane or combined taxanes more relevant to NHS practice? Which taxane is most commonly used in the NHS?

12a. Is it reasonable to assume a class effect for taxane therapy?

12b. Could post-hoc analysis of the effectiveness of efficacy of nivolumab compared with either docetaxel or paclitaxel have been carried out?

12c. Is a comparison of nivolumab with individual taxanes more relevant to NHS practice than comparisons with a combined taxane arm?

The company's response to this question states that it is not accurate to say that subgroup analysis by taxane was not provided. The ERG highlights that this statement corresponds to the following section of the ERG's report:

"The ERG also notes that allocation of specific taxanes … was determined prior to allocation of nivolumab or taxanes. This means that it would be theoretically possible to consider a comparison of patients considered suitable for treatment with docetaxel or paclitaxel, which would enable an assessment of how similar outcomes were for nivolumab-treated patients that were deemed suitable for each taxane. These analyses were not provided."

The ERG understands the rationale behind the specification of a 'taxanes' comparator arm for ATTRACTION-3, in recognition of treatment guidelines, as well as clinician and patient preference. Given that the choice of taxane is guided at least in part by patient fitness (i.e. tolerability concerns), the ERG highlighted the importance of considering a potential difference

in outcomes between taxane choice. In the response provided by the RCP representative Dr Elizabeth Smyth, it is noted that differences in taxane outcomes are likely due to institutional preference or personal choice of the patient/oncologist.

Considering the limitations of statistical analyses of non-pre-specified subgroup analyses based on small patient numbers, and the lack of biological plausibility for a specific difference in outcomes by taxane, the ERG considers it reasonable to assume a class effect for taxane therapy. However, in light of the company's response to this question (and Question 26 – that paclitaxel may be the predominant choice of taxane in NHS practice), the ERG highlights that an analysis based on only paclitaxel-eligible patients may nevertheless be of interest to the Committee.

13. Are health outcomes expected to be different between 'docetaxel-preferred' and 'paclitaxel-preferred' populations?

13a. Are there systematic differences in people who would be suitable for treatment with either docetaxel or paclitaxel? How should these separate populations be defined?

Please see the ERG's response to question 12.

Issue 7: Difference between company and ERG ICER

14. Which modelling methods are most suitable for estimating overall survival and expected time on treatment?

In the company's response to the technical report, it is stated:

> *"Relevant survival modelling guidelines indicate that survival extrapolations should reflect the disease pathway and plausible biological explanation for treatment effect. This includes reviewing the overall impact across model inputs, rather than reviewing models in isolation. This is specifically of note in the ERG base case, where a case is made for each input in isolation, but the overall impact is to predict clinically implausible outcomes. As noted by the ERG, the ERG base case applies more conservative extrapolations for nivolumab OS due to the concerns around the generalisability of the evidence. However, taxane OS is assumed to be more optimistic, to the extent that ATTRACTION-3 extrapolations predict longer mean OS for taxanes than for nivolumab, which can be considered implausible in the context of the observed data. Further, despite shorter OS assumptions, it is assumed that time on treatment is increased."* Company response to technical engagement question 14, page 39.

The company is correct to highlight that the ERG's base-case analysis includes the specification of an OS curve for nivolumab which is more conservative (compared with the company's base-case analysis), and an OS curve for taxanes which is more optimistic (compared with the company's base-case analysis). However, the ERG disagrees that its base-case analysis predicts clinically implausible outcomes.

The company's assertion here is that the estimated mean OS for taxanes is greater than that of nivolumab. This is incorrect – the ERG's base-case analysis estimates mean total life-years for nivolumab of [redacted] versus [redacted] for taxanes. The company previously raised the same point in the FAC (Issue 2) which the ERG responded to highlighting these values. It is still unclear to the ERG why the company believes the ERG's extrapolations predict longer mean OS for taxanes than for nivolumab.

Later in the company's response, it is stated:

> *"Although it is likely that some patients will receive treatment beyond progression, it is not plausible that there will be extended post-progression treatment period in the absence of clear benefits such as improved quality of life. Further, it is implausible that extended post-progression treatment would be seen in the absence of clinical improvement, which would be reflected in OS. In this context, it is not appropriate to extend time on treatment, reduce OS (for nivolumab only) and assume no post-progression utility differential between treatment arms. Hence, survival modelling in this indication should be considered across model inputs, rather than in isolation, otherwise the results are clinically implausible."*

With respect to the estimated time on treatment (ToT), the company is correct to highlight that the ERG's preferred analysis includes the specification of a curve for nivolumab which leads to an increase in the estimated ToT (compared with the company's base-case analysis). However, the ERG highlights that its choice of preferred models should not be considered plausible or implausible versus the company's choice of preferred models. Rather, both the ERG's and company's preferred models should be considered against the evidence available to assess their suitability (e.g. visual fit to the Kaplan-Meier curve, goodness of fit scores, clinical plausibility etc., per NICE DSU TSD 14).[2]

The issue raised above in the company's response is around the ERG's preferred base-case analysis leading to an increased ToT, decreased OS, and removal of a post-progression utility

benefit for nivolumab *relative to* the company's base-case analysis. The ERG does not consider this sufficient justification for considering the ERG's analysis clinically implausible.

The company also highlights the following in its response:

> *"The ERG quote NICE DSU TSD 14 guidance to say that "where parametric models are fitted separately to individual treatment arms it is sensible to use the same 'type' of model, that is if a Weibull model is fitted to one treatment arm a Weibull should also be fitted to the other treatment arm". The company agrees that it should be considered appropriate where models are fitted separately to individual treatment arms to use the same model "type" where the hazards presenting are the same and where the data directs. However, the company does not consider that these conditions are satisfied by the nivolumab and taxane arm in ATTRACTION-3."*

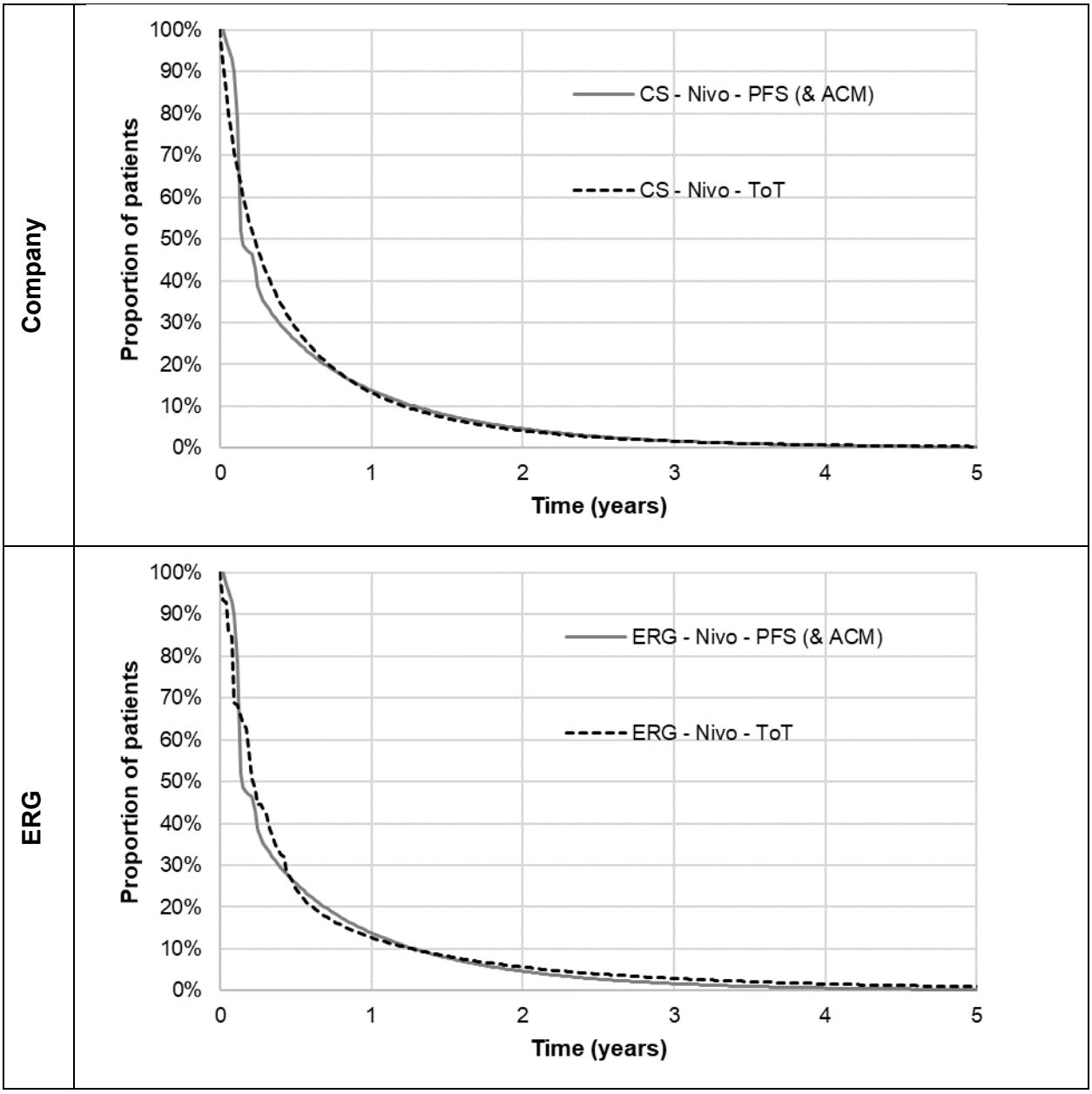
The ERG re-iterates that the remainder of this excerpt from its report is generally in accordance with the company's comment above; namely:

> *"… While this principle is subject to debate within the context of two treatments with very different mechanistic properties, the ERG calculated ICER values for nivolumab versus taxanes using the same method of OS extrapolation (where deemed clinically plausible for both treatment arms in the CS, based on CS Figures 35 and 36)."*

The ERG notes the final sentence of this full quote, which notes that the models considered in this scenario were deemed *"clinically plausible"* by the company within its submission. In spite of this, the ERG acknowledges that a specific combination of models may provide very different (and potentially unrealistic) estimates of incremental survival benefit, and therefore as stated in its report, the most appropriate estimation of OS is highly uncertain and subject to debate.

With respect to the company's point concerning treatment beyond progression, it is important to acknowledge the difference in PFS and ToT curves in the company's and the ERG's base-case analyses, which are provided in Figure 1. From this plot, it can be seen that a small proportion of patients are estimated to be progressed and still on treatment from approximately 1.4 years onwards, though this is a relatively small proportion of patients (the largest difference across the model time horizon occurs at approximately 2.75 years, where PFS = 2.2%, ToT = 3.5%, and therefore the difference between PFS and ToT = 1.3%).

Figure 1: Comparison of company's and ERG's base-case extrapolations of PFS and ToT



Key: ACM, all-cause mortality; ERG, Evidence Review Group; PFS, progression-free survival; ToT, time-on-treatment.

As further exploration, the ERG considered an analysis wherein the ERG's preferred ToT curve is capped by the PFS curve at approximately 1.4 years onwards (i.e. where the PFS and ToT curves cross in the ERG's base-case analysis, shown in Figure 1). This caused the ERG's original base-case ICER to reduce from £125,984 to £115,956, reflective of a reduction in the incremental costs from +£27,845 to +£25,629. However, as highlighted in the company's response to this question, treatment beyond progression can occur (and indeed, was permitted

in the ATTRACTION-3 trial). Therefore, this scenario is provided purely to illustrate the impact on the ICER were the ToT curve capped by the PFS curve in the longer term.

It is also important to acknowledge that within the confines of a partitioned survival analysis (PartSA) model structure, PFS and ToT are independent. Therefore, the 'true' mix of patients still on treatment by progression status is not quantified.

The company also comments on the recommendations provided by Bagust and Beale[3] with respect to the choice of survival model:

> *"These recommendations were adhered to in the company base case approach and seem to be at odds with the technique used in the ERG base case selection."*

For context, the company's and ERG's preferred models for OS are presented in Table 1. In the quote above from the company's response, the company is referring to its choice of a log-logistic versus exponential model for the nivolumab and taxanes arms (respectively), as compared to the ERG's choice of a generalized gamma model for both arms. The ERG considers its selection of the generalized gamma model to not be "at odds" with the recommendations of Bagust and Beale, given that this model includes several other models as special cases (including the exponential, Weibull, and lognormal). To clarify, the ERG's preferred model is not a jointly-fitted model (i.e. a single generalized gamma model with a covariate for treatment assignment), but is instead two separately-fitted generalized gamma models.

Table 1: Comparison of company versus ERG preferred models for overall survival

	Company	**ERG**
Nivolumab	Semi-parametric Kaplan-Meier to 2.99 months with parametric extrapolation using log-logistic distribution	Semi-parametric Kaplan-Meier to 5.75 months with parametric extrapolation using generalized gamma distribution
Taxanes	Semi-parametric Kaplan-Meier to 2.99 months with parametric extrapolation using exponential distribution	Semi-parametric Kaplan-Meier to 5.75 months with parametric extrapolation using generalized gamma distribution

In light of the above, the ERG still considers its base-case analysis to be clinically-plausible, but acknowledges that the most suitable choice of model for each outcome (OS or ToT) is uncertain. The ERG also appreciates that the generalized gamma and exponential extrapolations may be similar, though does not consider this to be of particular concern within

the estimation of the ICER (i.e. the fact that the curves are similar means that the choice of model should not have a marked effect on the ICER).

The ERG is in agreement with the company that within the context of the models fitted and provided by the company, the semi-parametric approach is suitable to inform decision making. The choice of cut-off point, and the selection of the parametric model to inform the latter part of the curve, are both subject to debate, and the ERG's preferred base-case settings remain unchanged.

15. Which estimates of treatment, administration and medical resource use costs are most reliable?

The ERG considers drug costs taken from eMIT to be standard in company submissions to NICE, and therefore does not consider the use of costs from MIMS or the BNF to be suitable to inform the economic model where eMIT costs are available. The ERG also re-iterates a point made in its report that the BNF (freely available via the NICE website), provides an alternative source of list prices which could have been used to inform the model.

Table 2 illustrates the difference in costs from each source (BNF, eMIT, and MIMS). From this table, it can be seen that for paclitaxel, the lowest costs from the BNF and MIMS costs are identical; however, costs from eMIT are markedly lower. For docetaxel, it can be seen that lower costs are available from the BNF versus those from MIMS, and even lower costs are available from eMIT.

Table 2: Comparison of taxane costs from BNF, eMIT, and MIMS

	Paclitaxel costs	Docetaxel costs
BNF[a]	6 mg/mL concentration for solution for infusion in vial • 5 mL: £66.85 to £120.85 • 16.7 mL: £200.35 to £374.00 • 25 mL: £300.52 to £561.00 • **50 mL: £601.03 to £1,122.00**	10 mg/mL concentration for solution for infusion in vial • 2 mL; £162.75 • 8 mL: £534.75 • 16 mL: £1,069.50 20 mg/mL concentration for solution for infusion in vial • 1 mL: £15.00 to £204.20 • 4 mL: £21.43 to £1,206.08 • **8 mL: £51.00 to £710.26**
eMIT[b]	6 mg/mL concentration for solution for infusion in vial • 5 mL: £4.69 • 16.7 mL: £23.06 • 25 mL: £18.88 • **50 mL: £39.32**	20 mg/mL concentration for solution for infusion in vial • 1 mL: £4.61 • 4 mL: £12.50 • **8 mL: £20.96**
MIMS[c]	6 mg/mL concentration for solution for infusion in vial • 5 mL: £66.85 • 16.7 mL: £200.35 • 25 mL: £300.52 • **50 mL: £601.03**	10 mg/mL concentration for solution for infusion in vial • 2 mL; £162.75 • 8 mL: £534.75 • 16 mL: £1,069.50 20 mg/mL concentration for solution for infusion in vial • 1 mL: £145.80 • 4 mL: £479.06 • 7 mL: £900.00 • **8 mL: £958.11** For solution in infusion in vial • 20mg, 1 mL: £153.47 • 80mg, 4 mL: £504.27 • 140mg, 7 mL: £720.10 • 160mg, 8 mL: £1,008.54

Key: BNF, British National Formulary; eMIT, electronic market information tool; MIMS, Monthly Index of Medical Specialties

Notes: The largest vial size available for each product across all three sources is **underlined and in bold print**. [a] As per the BNF website, 21 September 2020; [b] As per eMIT last updated 4 March 2020, accessed 21 September 2020; [c] Taken from CS, Table 61.

For administration, the ERG continues to prefer its base-case analysis assumptions, wherein treatment administration is assumed to take place in a day case setting, and chair time with taxanes is expected to be longer.

For monitoring, the company submission uses a cost of £187.36, based on a weighted average of a consultant led and non-consultant led medical oncologist appointment, using service code 370. The ERG report noted that the service code 370 was not stated in the company submission, but the currency codes of WF01A-WF02D were provided. From this, the ERG estimated the cost to be £196.33. However, upon further inspection, the ERG understands this cost to be reflective of *all* outpatient costs (including procedures). Therefore, the ERG considers the original value of £187.36 (excluding outpatient procedures) to be the most suitable, and has updated its preferred base-case analysis accordingly (presented in Section 3).

For nerve block, the ERG agrees that a cost of £532.96 is used in the model and is suitable. However, the ERG's concern was with regards to the proportion applied. This meant that the stated end cost was £26.62, whereas the ERG calculated this as £2.66 (£532.96 [cost of treatment] x 0.005 [proportion of patients requiring treatment]) based on information provided in Table 64 of the CS. While of relatively little concern with respect to the ICER, the ERG is still unclear if this calculation includes an error.

For hospitalization, the company's base-case analysis includes a "standardized" cost to obtain a cost per day to apply within the model. However, based on the clinician survey provided in the CS, it appears as though clinicians were asked how often patients would be hospitalized (e.g. once every three months, versus monthly, bi-weekly etc.) but without the concept of how long they would be in hospital for.

In the company's response to the technical report, it is stated that a daily length of stay was applied to avoid double counting results (given that the model cycle length was one week, and [based on the ERG's understanding] patients could theoretically be in hospital for more than one week). In light of this, the ERG considers the company's approach likely to have substantially under-estimated hospitalization costs, as it appears as though all hospitalizations are assumed to be one day in length – an assumption made to avoid any hospitalizations that exceed the model cycle length of seven days.

While the unadjusted value may lead to some patients technically accruing costs after death, assuming all patients have a length of stay of one day is also incorrect. For example, patients

with a length of stay greater than one day but less than one week will accrue the costs as if they had a length of stay of only one day. It is the ERG's view that the latter approach (assuming a one-day length of stay for all hospitalizations) is "more incorrect" than not adjusting for the length of stay (accepting that this approach is challenging to reconcile with the specification of a weekly model cycle length). The ERG therefore prefers its use of the full hospitalization cost (without adjusting for length of stay).

Issue 8: Model time horizon (40 years in company base-case)

16. What is the most appropriate time horizon for the economic model?

16a. Clinical expert advice is required on the life expectancy of patients with unresectable, advanced oesophageal cancer treated with nivolumab or taxanes

16b. What is the most relevant time horizon for the economic model?

The ERG agrees that a lifetime horizon should be used within the model, and that based on the extrapolations, a small proportion of patients are estimated to survive longer than 10 years. However, it was the opinion of clinical advisers to the ERG that close to all NHS patients receiving either nivolumab or taxanes are expected to have died within 10 years of treatment initiation (see ERG's response to FAC Issue 2 for more information).

The combination of the time horizon (which is capped at a maximum of 40 years in the company's model), and the choice of survival extrapolation can lead to a substantial proportion of patients still being alive for more than 10 years. Based on the information presented in the CS, the ERG's report, and in response to the technical engagement report; this may or may not be considered clinically plausible. Therefore, the ERG considers that if all patients (on both treatment arms) are expected to have died by 10 years, a 10-year time horizon should be sufficiently reflective of the lifetime of this patient population. However, if a small proportion of patients are expected to survive beyond 10 years, a longer time horizon is warranted.

A range of time horizons may be important to consider in decision making to understand the impact of the extrapolated tail of the survival curves on the ICER.

Issue 9: Alternative extrapolations for overall survival

17. Is a fully-parametric or semi-parametric model most appropriate to predict the long term effectiveness of nivolumab and taxane therapy?

As described in response to Question 14, the ERG considers a semi-parametric approach to be the most appropriate to predict the long-term effectiveness of nivolumab and taxane therapy.

However, the ERG caveats this statement by noting that there may be other modelling approaches not presented that may be equivalently (or perhaps even more) plausible than the semi-parametric approaches presented.

18. If a semi-parametric model is preferred, is 2.99 months an appropriate point to start the extrapolation?

18a. If a semi-parametric model is preferred, is 2.99 or 5.75 months an appropriate point to start the extrapolation?

The ERG recognises that the specification of a later cut-point (i.e. 5.75 months instead of 2.99 months), means that extrapolations are based on relatively fewer data points, and may therefore be subject to additional uncertainty. Nevertheless, the models fitted with a 5.75 month cut-point were still considered to provide more realistic extrapolations, and avoided the potential issues in selecting a cut-point close to where the Kaplan-Meier curves cross.

19. Does the overall survival predicted in the company (at 10 years) or ERG model (0.2% at 10 years) best represent the likely overall survival in NHS practice?

The ERG understands that a small proportion of patients may be expected to achieve long-term survival benefits (perhaps in the region of 5% as per the ATTRACTION-3 statistical analysis plan, stated in the company's response to this question). However, based on advice provided to the ERG, nearly all patients are expected to have died by 10 years; and long-term data in an oesophageal squamous cell carcinoma (OSCC) population are extremely limited.

The company highlights a range of other data sources demonstrating the efficacy of nivolumab in other populations. The ERG understands that these data sources provide potentially helpful information to understand the likely longer-term outcomes associated with nivolumab in an OSCC population. Given that with current care the majority of patients will not survive beyond one year (ERG report, Section 2.1), the fact that the ERG's preferred base-case extrapolation estimates five- and 10-year OS for nivolumab to be and (respectively), could be considered optimistic (versus the expected outcomes in NHS practice).

Accordingly, the ERG prefers extrapolations based on its preferred modelling assumptions (which estimates 10-year OS with nivolumab to be). However, alternative survival extrapolations may be of relevance to decision making (including the company's base-case extrapolation which estimates of nivolumab patients to still be alive at 10 years). 10-year survival with nivolumab in an NHS patient population is unknown, and will remain so for the foreseeable.

Issue 10: Exploratory analysis of utility values

20. Are the differences in utility between the nivolumab and taxane arm in the company model clinically plausible?

As stated in the ERG's report, the ERG has several concerns with the utility values derived from the ATTRACTION-3 trial data:

- ATTRACTION-3 was an open label study, meaning that patients were aware of their treatment assignment. There is therefore a possibility that patients in the active intervention arm (i.e. nivolumab) would, all other things equal, be more likely to report a higher utility versus those on the control arm (i.e. taxanes). There is, however, conflicting literature concerning bias in patient-reported outcome measures in open-label trials, particularly those conducted in cancer populations.[4-6]
- Mean baseline utility (taken at screening) for the taxanes arm was significantly lower than of the nivolumab arm ([redacted]) (CS Appendix N, Section 4.1). Therefore, differences between the arms for either health state may be plausibly explained (at least in part) by the difference in baseline utility (which could be related to the previous point concerning the open-label design of ATTRACTION-3)
- The median utility value for nivolumab-treated patients was [redacted] based on CS Figure 12. The ERG considers that a median value of [redacted], is unrealistic within the context of a patient population generally aged >65 years with an advanced cancer that has not responded to a previous line of chemotherapy.

The ERG accepts that there is some evidence in support of an improved utility for patients treated with a cancer immunotherapy versus chemotherapy, as acknowledged within the ERG's report (Section 4.2.5). However, the extent of this benefit, and whether or not it can be reasonably assumed to persist after disease progression, is unclear.

In the ERG's base-case analysis, an assumed difference in utility was applied for the progression-free health state. However, it was assumed that this difference would not apply after disease progression. The ERG re-emphasizes the point made within its report concerning the utility values applied in its preferred base-case analysis, presented below:

"… determination of the most appropriate utility values for use within the model is subject to debate. The approach taken to inform the ERG's base-case analysis may be considered in

some respects conservative (with respect to the assumed lack of difference between arms beyond progression) and in others, optimistic (given that the difference between arms in the PF state is unchanged, and is expected in part to be related to the differences seen in utility at screening, as well as the open label design of ATTRACTION-3).” – ERG report, Section 6.3.5

21. Would a treatment independent approach (e.g. using the same values for pre and post progression across the treatment arms) be more appropriate?

The ERG understands that treatment-specific utilities are frequently queried as part of other NICE assessments of cancer immunotherapies. However, as highlighted in response to Question 20, there are several specific considerations within the context of the ATTRACTION-3 trial, and the nature in which HRQoL data were collected, that warrant careful interpretation of the values produced. Consequently, an in-depth assessment of the suitability of treatment-specific utility values to inform the economic analysis was considered warranted.

For the pre-progression state, the ERG expects there may be a difference in the utility value between treatment arms, reflecting the difference in safety profiles/ mechanistic properties of the two treatments. However, whether this is to the extent suggested by the company's analysis or not is less clear (and could be due to the open-label trial design, as noted in response to Question 20 above). In the ERG's base-case analysis, the company's mixed-effects regression analysis values were applied for the progression-free health state (i.e. assuming a difference in utility by treatment arm), but were noted to be subject to palpable uncertainty.

However, it is the ERG's view that treatment-independent utility values for the post-progression health state should be factored into the base-case analysis. Utility data beyond progression are extremely limited, and the majority of patients discontinue treatment at, or prior to, disease progression. Some beneficial effects of nivolumab may be experienced after progression, though the ERG does not consider it likely that this benefit would (a) be to the extent suggested by the company's utility analysis (a difference in utility [redacted]), or (b) apply indefinitely.

Issue 11: Alternative extrapolations of time on treatment

22. Are fully-parametric or semi-parametric methods most appropriate for estimating time on treatment?

a. It is not clear why a fully-parametric model was used by the company after using semi-parametric for OS and PFS. What was the rationale for this?

b. Is the long term extrapolation based on fully-parametric (company) or semi-parametric (ERG) methods most appropriate for estimating time on treatment?

The ERG considers its base-case analysis (using a semi-parametric model, with a cut-off point at 5.75 months) to be the most suitable. This approach is consistent with the approaches taken for OS and PFS. Given that duration of treatment is expected to follow a reasonably-similar shape to the PFS curve, it is the ERG's view that a consistent modelling approach should be undertaken unless there is a clear justification for why an alternative approach should be taken (which the ERG does not consider warranted by the evidence presented). Further discussion concerning the choice of cut-off point is provided in the ERG's response to Question 14.

23. How long are people likely to remain on treatment with nivolumab in NHS practice?

The ERG has no further comments concerning the anticipated duration of treatment with nivolumab in NHS practice outside of those raised in its report.

24. Is a stopping rule appropriate? If so, what stopping rule(s) are most relevant for NHS practice?

Per the ERG's response to Question 23, the ERG has no further comments concerning the anticipated duration of treatment with nivolumab in NHS practice outside of those raised in its report.

Issue 12: Have the costs of comparator treatment been appropriately estimated?

25. Which source of cost estimates for medical technologies is most reflective of those paid by NHS trusts?

25a. Company to clarify why was MIMS used as a tool to source treatment costs, given that eMIT prices are more reflective of those paid by NHS trusts?

Please see the ERG's response to Question 15 for the ERG's view on costs to inform the economic model.

26. Is it appropriate to assume a 50:50 market share of taxanes?

26a. Clinical experts to advise whether in clinical practice docetaxel is preferred over paclitaxel.

Clinical advice provided to the ERG suggested that docetaxel was the preferred treatment option for most patients, owing to the fact that it is administered less frequently (i.e. once every three weeks instead of once per week). However, patients may instead be treated with paclitaxel which is considered to have a more favourable safety profile.

Based on the market research provided by the company, and the response provided by the RCP representative Dr Elizabeth Smyth, it may instead be the case that the more general preference is for paclitaxel, or that a 50:50 split may be suitable. The ERG also highlights that in ATTRACTION-3, n=65 of the 'taxanes' patients were treated with docetaxel, versus n=144 with paclitaxel (equivalent to a ratio of approximately 31:69).

As shown in the company's response to this question, assuming all other model parameters are held at the same value, a larger proportion of patients treated with paclitaxel (instead of docetaxel) leads to a reduction in the ICER. This is because paclitaxel is associated with a higher administration cost per treatment cycle (owing to it being administrated six times out of every six-week treatment cycle, versus being administered once every three weeks).

Due to the conflicting nature of taxane use in ATTRACTION-3, based on market research, and clinical advice; the ERG prefers the use of a 50:50 split to inform its base-case analysis, but expects that a range of scenarios (including 100% use of either taxane) may be helpful for decision making.

Issue 13: Administration and medical resource use costs

27. Have the most appropriate sources been used to calculate administration and medical resource use costs? If so, which estimates of cost are most reliable?

27a. Are the medical resource costs calculated by the company or ERG most relevant to NHS practice?

Please see the ERG's response to Question 15.

Issue 14: End of life criteria

28. Is the estimate of extension to life robust? Is an extension to life of at least 3 months expected to be realised in NHS practice?

28a. Is the company or ERG method of survival extrapolation most appropriate?

28b. Could further follow-up data from ATTRACTION-3 be made available to support survival analysis?

28c. Is there a subgroup within ATTRACTION-3 that would best reflect the health outcomes typically seen in NHS practice?

The ERG considers its base-case extrapolations of survival to be appropriate for decision making. However, the choice of the most appropriate extrapolation is (at least to some extent) subjective, and associated with uncertainty. Alternative extrapolations may therefore be useful to consider within the context of decision making.

Based on the company's economic model, an estimated survival benefit of at least three months was obtained in the ERG's preferred base-case analysis, as well as the company's base-case analysis. However, it would be remiss of the ERG to not highlight that estimates of survival benefit from the model are predominantly based on data from the ATTRACTION-3 trial, for which several generalisability issues were highlighted in the ERG's report.

Further follow-up data concerning longer-term OS and ToT may help resolve uncertainty inherent in the economic analysis presented. However, other uncertainties (such as utility values, and the difference in outcomes by taxane/ taxane choice) would not be resolved by further collection of data from ATTRACTION-3. External data collection may resolve some of these issues.

The ERG previously highlighted that differences between the Japanese and ROW populations may warrant further investigation. However, based on information provided by the company and Dr Elizabeth Smyth (RCP representative), the ROW subgroup may not be considered "more reflective" of the NHS population, versus the full intention-to-treat population.

3. REVISED ERG BASE-CASE ANALYSIS

In light of the company's response to Question 15, the ERG has made the following change to its preferred base-case analysis:

- The unit cost for an outpatient consultation is changed to reflect the value provided in the CS, of £187.36. This replaces the value of £196.33 used in the ERG's previous base-case analysis.

All other model settings and assumptions remain unchanged from those stated in the ERG's report. The impact of this change on the cost-effectiveness results is relatively small, decreasing the ICER from £125,984 to £125,886, as shown in Table 3.

Table 3: Comparison of original and updated ERG base-case results

Arm	Total			Incremental			ICER (£/QALY)
	Costs (£)	LYs	QALYs	Costs (£)	LYs	QALYs	
Original ERG base-case (deterministic)							
Taxane	[redacted]	[redacted]	[redacted]				
Nivolumab	[redacted]	[redacted]	[redacted]	27,845	0.302	0.221	125,984
Updated ERG base-case (deterministic)							
Taxane	[redacted]	[redacted]	[redacted]				
Nivolumab	[redacted]	[redacted]	[redacted]	27,824	0.302	0.221	125,886

Key: ICER, incremental cost-effectiveness ratio; LY, life year; QALY, quality adjusted life year.

4. REFERENCES

1. Gyawali B, Hey SP, Kesselheim AS. A Comparison of Response Patterns for Progression-Free Survival and Overall Survival Following Treatment for Cancer With PD-1 Inhibitors: A Meta-analysis of Correlation and Differences in Effect Sizes. JAMA Netw Open. 2018;1(2):e180416.

2. Latimer N. NICE DSU Technical Support Document 14: survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data 2011. Available from: http://nicedsu.org.uk/.

3. Bagust A, Beale S. Survival analysis and extrapolation modeling of time-to-event clinical trial data for economic evaluation: an alternative approach. Med Decis Making. 2014;34(3):343-51.

4. Anota A, Poset A, Lefevre C, Lemasson H, Cotte F-E, Guerzider S, et al. 169P Impact of open-label design on patient-reported outcomes (PROs) data in randomized clinical trials of immuno-oncology (IO) agents in patients with advanced or metastatic cancer: A 10-year systematic literature review (SLR). Annals of Oncology. 2019;30:xi58-xi61.

5. Gnanasakthy A, Barrett A, Evans E, D'Alessio D, Romano C. A Review of Patient-Reported Outcomes Labeling for Oncology Drugs Approved by the FDA and the EMA (2012-2016). Value in Health. 2019;22(2):203-9.

6. Roydhouse JK, Fiero MH, Kluetz PG. Investigating Potential Bias in Patient-Reported Outcomes in Open-label Cancer Trials. JAMA Oncology. 2019;5(4):457-8.