NICE National Institute for Health and Care Excellence

Single Technology Appraisal

Nivolumab with platinum- and fluoropyrimidine-based chemotherapy for untreated HER2-negative advanced gastric, gastro-oesophageal junction or oesophageal adenocarcinoma [ID1465]

Committee Papers

ISBN: 978-1-4731-4939-7

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

SINGLE TECHNOLOGY APPRAISAL

Nivolumab with platinum- and fluoropyrimidine-based chemotherapy for untreated HER2-negative advanced gastric, gastro-oesophageal junction or oesophageal adenocarcinoma [ID1465]

Contents:

The following documents are made available to consultees and commentators:

1. **Comments on the Appraisal Consultation Document from Bristol-Myers Squibb**
   a. Appendix A – utility analysis
   b. Appendix B – survival report

2. **Comments on the Appraisal Consultation Document from experts:**
   a. David Chuter – patient expert, nominated by Guts UK

3. **Comments on the Appraisal Consultation Document received through the NICE website**

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

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

Response to the Appraisal Consultation Document

Nivolumab in combination with chemotherapy for untreated advanced gastric or gastro-oesophageal junction cancer

ID1465

Bristol-Myers Squibb Pharmaceuticals Ltd

November 2021

Contents

Executive summary

This document provides a response to the Appraisal Consultation Document (ACD) describing the use of nivolumab in combination with chemotherapy (NIVO+CHEMO) for patients with untreated advanced gastric or gastro-oesophageal junction cancer. In line with the ACD, this response outlines the additional clinical and economic evidence as requested by the Appraisal Committee, which can be used to support decision making.

Key points of the additional evidence presented in this response:

- Additional evidence provided from the CheckMate 649 [redacted] database lock for the CPS≥5 population demonstrate that results remain consistent with the results of the primary analysis presented in the company submission

- A partitioned survival model (PSM) was produced. The base case PSM results are presented alongside with the results of the initially chosen semi-Markov model that included the long-term remission state

The semi-Markov model was presented in the original submission as it can incorporate the impact of both time and duration of progression, and the associated likelihood of death. BMS initially ruled out a PSM as it is not able to capture the influence of time since progression on survival. However, BMS understands that the PSM is the preferred approach for the committee as OS data is used directly and, therefore, are presenting the PSM as the base case. The OS estimates from the PSM match the observed trial data closely. The resulting analysis from the PSM shows that NIVO+CHEMO can be considered cost-effective (i.e., an ICER below a £50,000 per QALY willingness-to-pay threshold). This validates the cost-effective ICER seen with the semi-Markov model presented in the initial submission. Additionally, although overestimating OS, the results from the semi-Markov model have improved using the most recent data and still produces a cost-effective ICER.

Further, clinical experts consulted for the ACD stated that long-term data supporting a cure assumption does not exist. However, one expert noted that about 4% of people may achieve long-term remission with chemotherapy alone and that it is expected to be double when given with nivolumab.[1] Data from the CheckMate 649 trial shows that 15.8% of the patients receiving NIVO+CHEMO are alive and have not progressed at 30 months, which is approximately twice as many as for patients receiving CHEMO only. The NHS England clinical lead noted that patients in long-term remission may relapse, although this tends to be uncommon.[1]

In summary, 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. BMS believes that the Committee recommendations do not consider all the relevant evidence. The updated evidence provides a better reflection of the clinical and cost-effectiveness conclusions and supports a sound and suitable basis to produce clear guidance to the NHS. It is anticipated that further evidence presented in response to the ACD will be considered by the Appraisal Committee and will further demonstrate that nivolumab is cost-effective and is associated with substantial clinical benefit in a population with short survival and limited treatment options. The adoption of nivolumab for this therapeutic indication within NHS England would represent a significant advance in the management of this life-threating condition.

1 CheckMate 649 [redacted] database lock

As requested in the Appraisal Consultation Document, additional evidence is provided from the CheckMate 649 [redacted] database lock (data cut-off: [redacted]), reporting a minimum follow-up of 24 months for the NIVO+CHEMO and CHEMO arms.[2]

Table 1 provides patient disposition at the time of the database lock.[2] The overall rates of discontinuation were comparable in the NIVO+CHEMO and CHEMO arms (98% in both arms), with disease progression noted as the primary reason for discontinuation (69% and 71%, respectively). The median duration of therapy for all treated patients was 6.8 (range: 0.1–45.0) months in the NIVO+CHEMO arm and 4.86 (range: 0.0–44.2) months in the CHEMO arm. In line with the design of CheckMate 649, all patients in the NIVO+CHEMO arm had received a [redacted]; the maximum number of nivolumab doses was [redacted] in the NIVO+XELOX arm and [redacted] in the NIVO+FOLFOX arm.

Table 1. CheckMate 649 [redacted] Patient disposition[2]

	NIVO+CHEMO N=782	CHEMO N=767
Minimum follow up, months	24.0	
Median duration of treatment, months (range)	6.8 (0.1–45.0)	4.9 (0–44.2)
Discontinued treatment, n (%)	766 (98)	749 (98)
Disease progression	538 (69)	546 (71)
AE related to treatment	65 (8)	43 (6)
Completed treatment	55 (7)	0
AE unrelated to treatment	47 (6)	35 (5)
Withdrawal of consent	20 (3)	42 (5)
Other	41 (5)	83 (11)

AE: adverse event; CHEMO: chemotherapy; NIVO: nivolumab.
Other reasons for discontinuation: patient request to discontinue treatment (n = 55), maximum clinical benefit (n = 41), poor/non-compliance (n = 5), lost to follow-up (n = 4), patient no longer met study criteria (n = 1), death (n = 1), and additional reasons (n = 17)

As shown in

Table 2, Figure 1 and Figure 2, NIVO+CHEMO provided statistically significant and clinically meaningful improvements in PFS per BICR and OS in all randomised patients with PD-L1 CPS ≥5. These results were supported by outcomes in all randomised patients, which also showed significant benefit for NIVO+CHEMO versus CHEMO (Figure 1 and Figure 2). Additionally, there were benefits in terms of objective response rate and duration of response in patients with PD-L1 CPS ≥5: rate of response remained higher and responses were more durable with NIVO+CHEMO versus CHEMO. Further, responses deepened in the NIVO+CHEMO arm relative to the 12 month follow up, with 5 additional patients achieving complete response in the subgroup of patients with PD-L1 CPS ≥5 receiving NIVO+CHEMO.

Table 2. CheckMate 649 [redacted] database lock clinical outcomes for patients with PD-L1 CPS ≥5[2]

	NIVO+CHEMO (N=473)	CHEMO (N=482)
Overall survival		
Median OS (95% CI), months	14.4 (13.1–16.2)	11.1 (10.0–12.1)
Stratified HR (95% CI)	0.70 (0.61–0.81)	
Survival at one year, %	57	46
Survival at two years, %	31	19
Progression-free survival per BICR assessment		
Median PFS (95% CI), months	8.1 (7.0–9.2)	6.1 (5.6–6.9)
Stratified HR (95% CI)	0.70 (0.60–0.81)	
PFS at one year, %	37	19
PFS at two years, %	23	11
Response per BICR assessment		
Assessed, N	378	390
ORR, % (95% CI)	60 (55–65)	45 (40–50)
CR, %	13	7
PR, %	47	38
SD, %	28	34
PD, %	7	11
Duration of response		
Number of responders, N	226	176
Median DOR, months (95% CI)	9.7 (8.2–12.4)	7.0 (5.6–7.9)

BICR: blinded independent central review; CHEMO: chemotherapy; CI: confidence interval; CR: complete response; DOR: duration of response; HR: hazard ratio; NIVO: nivolumab; ORR: objective response rate; OS: overall survival; PD: progressed disease; PFS: progression-free survival; PR: partial response; SD: stable disease.

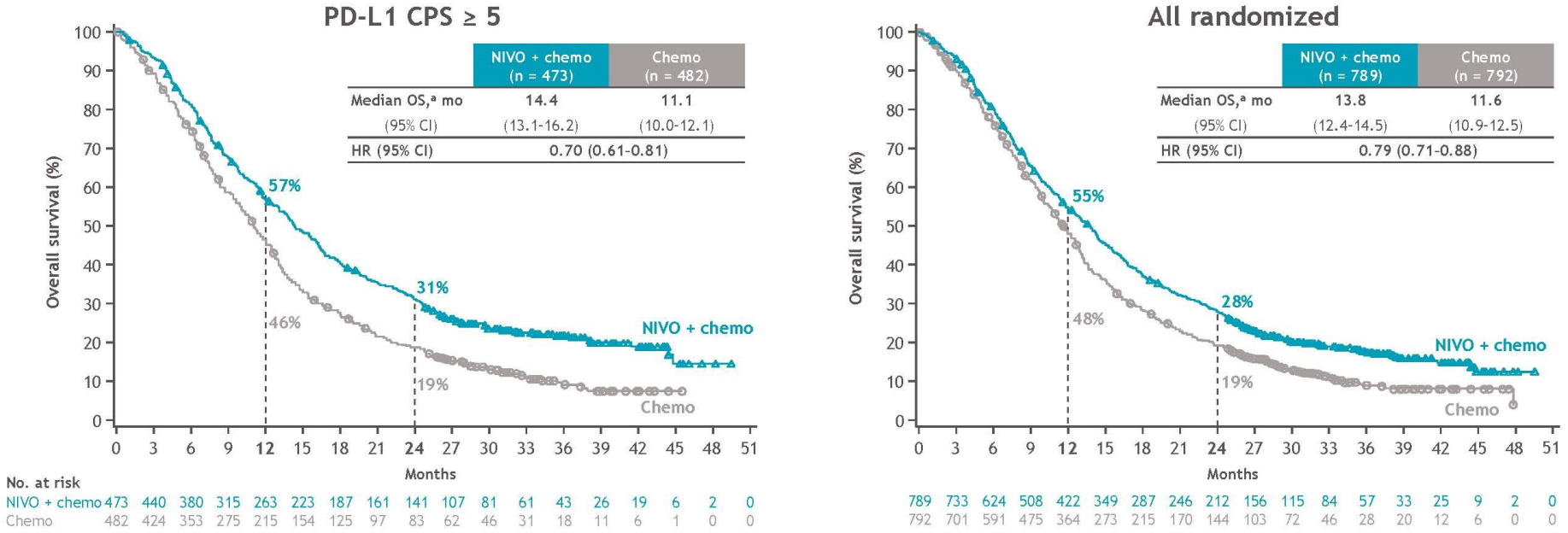

Figure 1. CheckMate 649 [redacted] database lock: Overall survival (OS) for all randomised patients and patients with PD-L1 CPS ≥5[2]

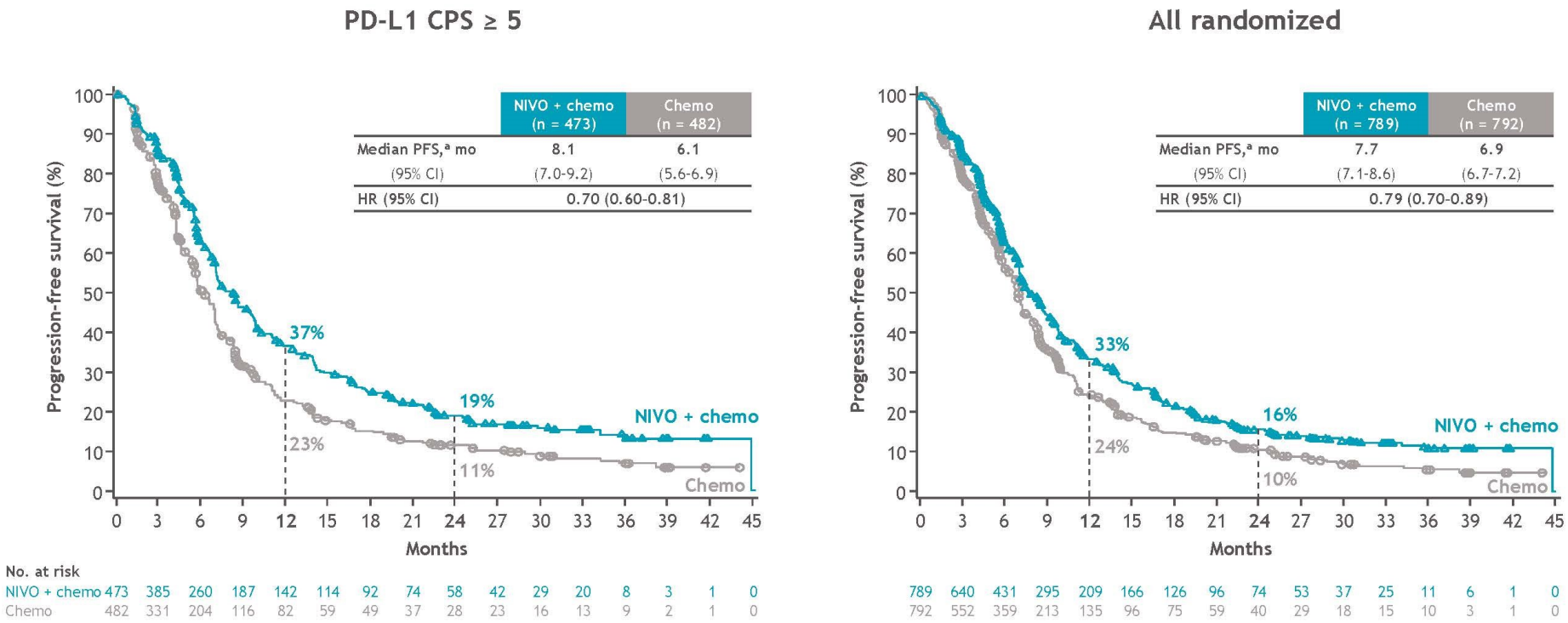

Figure 2. CheckMate 649 [redacted] database lock: Progression-free survival (PFS) for all randomised patients and patients with PD-L1 CPS ≥5[2]

As shown in Table 3, the safety profile for NIVO+CHEMO aligned with previous database locks, which can be considered manageable and reflective of the known safety profiles of the nivolumab and chemotherapy components. The incidence of TRAEs in patients with PD-L1 CPS ≥ 5 was consistent with all treated patients across both arms.

The most common grade 3-4 TRAEs for NIVO+CHEMO were neutropenia (15%), decreased neutrophil count (11%) and anaemia (6%), whereas for CHEMO this was neutropenia (11%-13%), decreased neutrophil count (9%-10%) and diarrhoea (3%-4%).[2]

Table 3. CheckMate 649 [redacted] database lock: adverse events for all randomised patients[2]

	NIVO+CHEMO (N=782)		CHEMO (N=767)	
	Any grade	**Grade 3–4**	**Any grade**	**Grade 3–4**
Treatment-related adverse events				
Any TRAE, n (%)	739 (95)	471 (60)	682 (89)	344 (45)
Serious TRAE, n (%)	175 (22)	133 (17)	94 (12)	77 (10)
TRAEs leading to discontinuation, n (%)	300 (38)	141 (18)	188 (25)	70 (9)
Treatment-related deaths, n (%)	16 (2)		4 (<1)	
Treatment-related adverse events with potential immunologic aetiology				
Endocrine, n (%)	109 (14)	6 (< 1)	3 (< 1)	0
Gastrointestinal, n (%)	266 (34)	43 (5)	208 (27)	25 (3)
Hepatic, n (%)	207 (26)	31 (4)	138 (18)	17 (2)
Pulmonary, n (%)	41 (5)	14 (2)	4 (< 1)	1 (< 1)
Renal, n (%)	29 (4)	7 (< 1)	9 (1)	2 (< 1)
Skin, n (%)	218 (28)	27 (3)	108 (14)	8 (1)

CHEMO: chemotherapy; NIVO: nivolumab; TRAE: treatment-related adverse event.

2 Has all the relevant evidence been taken into account?

Since publication of the Appraisal Consultation Document, additional evidence has been sought and further economic evaluations have been undertaken in order to address the Committee's requests. This includes:

- Data describing additional follow-up in CheckMate 649
- Economic evaluations applying the Committee's preferred assumptions and modelling methods, addressing the Committee's stated concerns
- Economic evaluations using data describing additional follow-up from CheckMate 649

In light of the updates to the evidence base, the Committee has not yet reviewed all relevant evidence; however, this will be remedied following receipt of the evidence contained in this report.

3 Are the summaries of clinical and resource savings reasonable interpretations of the evidence?

The summaries of clinical and resource savings are reasonable interpretations of the evidence. More clinical evidence from the most recent database lock is provided and presented in Section 1 as requested in the Appraisal Consultation Document.

4 Are the recommendations sound and a suitable basis for guidance to the NHS?

BMS does not believe that the recommendations can be considered sound and a suitable basis for guidance to the NHS. A thorough discussion of the Appraisal Committee recommendations and Appraisal Consultation Document has been provided in the response to address the uncertainties raised, with additional clinical and economic evidence that can be used to support decision-making. Thus, the recommendations made within the Appraisal Consultation Document should be reviewed in the light of the new evidence and analyses available.

5 Are there any aspects of the recommendations that need particular consideration to ensure we avoid unlawful discrimination against any group of people on the grounds of race, gender, disability, religion or belief, sexual

orientation, age, gender reassignment, pregnancy and maternity?

The committee recognised that patients and clinicians would welcome a new effective treatment for untreated HER2-negative advanced gastric, gastro-oesophageal junction or oesophageal adenocarcinoma as currently these patients have a poor prognosis which has a significant impact on their quality of life. Although diagnosis is often at an advanced stage and more common in men, patient experts reported an increasing number of younger people and women presenting with this disease.

6 Updated cost-effectiveness analysis

6.1 Cost-effectiveness analysis methods

6.1.1 Economic model approach

A semi-Markov model structure was adopted due to the requirement to incorporate the impact of both time and duration of progression on the likelihood of death. Initially, a partitioned survival approach was also considered, however, partitioned survival models effectively preclude explicit consideration of the influence of time since progression on survival and so were not considered further.

The structure of the model was able to capture all important aspects of gastric cancer and the expected benefits of NIVO+CHEMO, including delayed progression, improved survival and benefits to HRQoL. The model also includes the impact of introducing a long-term remission health state to capture the long plateau in the OS curve seen in both arms of the CheckMate 649 trial which can be indicative for a mixed population with a small “low-risk” fraction. The long-term remission health state in the semi-Markov model captures those patients still progression-free after a specified period of time and applies general population mortality rates instead of disease-specific mortality.

The company acknowledges the ERG’s and the Committee critique around the choice of economic model used. As such, the assumptions made in the semi-Markov model were deemed to lead to overestimating how long people live after treatment and it was noted that the model’s survival estimates were not supported by clinical trial evidence. It is argued that the company’s model survival estimates were higher than the overall survival seen in the trial. Because the company’s model did not correspond with the CheckMate 649 data, the ERG stated that the model long-term survival estimates and cost-effectiveness results lack reliability. The committee agreed the model lacked face validity because the modelled survival estimates did not match the survival estimates over the time that data was available.
The ERG suggested that a 3-state partitioned survival model (PSM) could use the survival data from CheckMate 649 directly.

Therefore, while the company supports the initial choice of semi-Markov model, it has produced a PSM as requested in the Appraisal Consultation Document, for the purpose of outcomes comparison and is presenting those results as the base case alongside with the results of the semi-Markov model including the long-term remission state.

Model structure

The original company base case used a semi-Markov model with 4 health states. All patients entered the model in the pre-progression state and remained there until death, disease progression or until they moved into the long-term remission health state (Figure 3). The PSM developed by the company in response to the ACD contains 3 health states, and again patients enter in the progression-free survival state and remain there until death or

progression. Health state occupancy in the PSM is determined directly by PFS and OS curves (Figure 4).

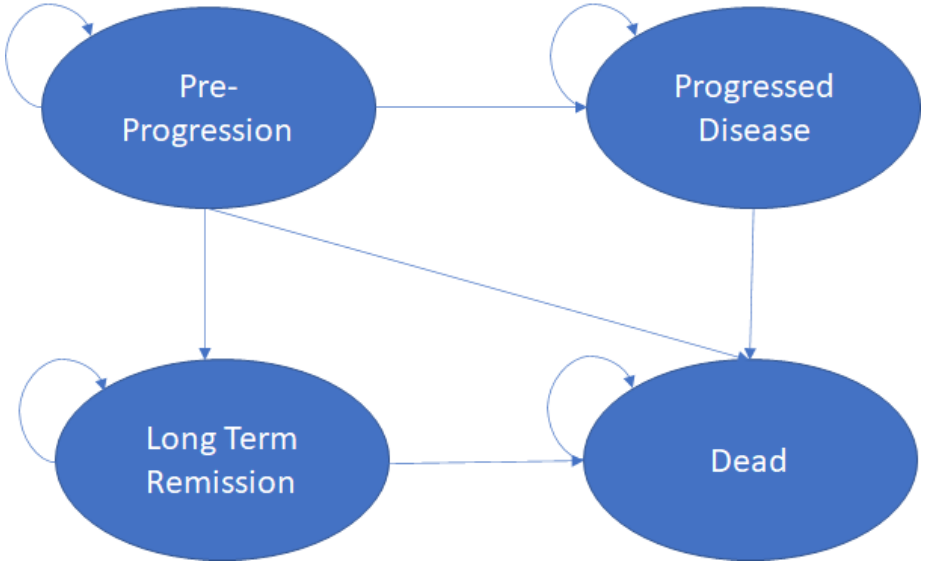

Figure 3. Base case Markov model with 4 health states

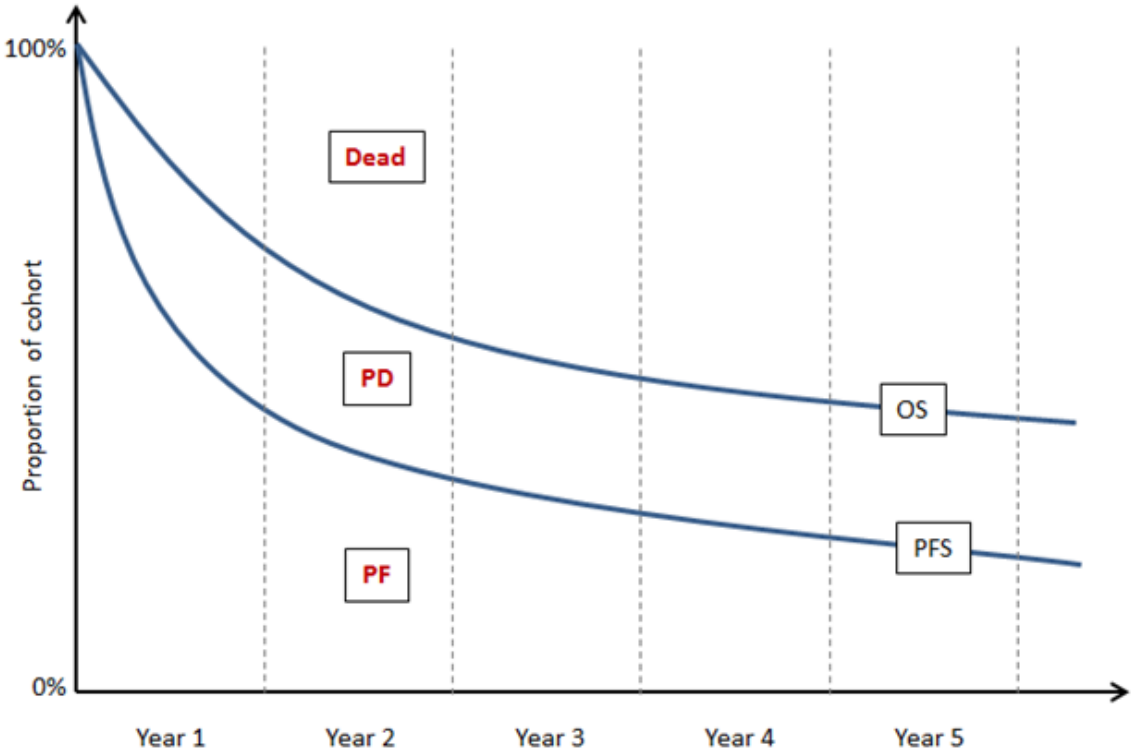

Figure 4. Health state occupancy in PSM.

PF = progression-free, PD = progressed disease, PFS = progression-free survival, OS = overall survival

Both models were designed to capture the relevant benefits of treatment with NIVO+CHEMO and represent improvements in PFS, OS and health utility as observed in CheckMate 649. As described in the original submission, both models use a fortnightly cycle length without half-cycle corrections. Half-cycle corrections are not required in economic models where the cycle length is rather short and where treatment costs are applied at specific intervals (every two to three weeks).

The clinical inputs informing progression through the model include PFS (in both models), OS (for PSM only), the likelihood of death upon progression and overall survival post-progression (OSPP); both of which are in the Markov model only (Table 4). Time on treatment (ToT) survival curves were used to determine the duration of treatment, in addition to the treatment stopping rule applied in CheckMate 649 and reflected in the SmPC.[3] In both models, these were derived from individual patient data (IPD) from CheckMate 649. Clinical outcomes are further detailed below in Sections 6.1.4.2 for OS, 6.1.4.3 for PFS, and 6.1.4.4 for time on treatment.

Table 4. Clinical model inputs

Clinical input	Partitioned survival model	Semi-Markov model
PFS	✓	✓
OS	✓	
Death upon progression		✓
OSPP		✓

OS: overall survival; OSPP: overall survival post progression; PFS: progression free survival

The Markov model also contains a transition from PFS to long term remission. This remained unchanged from the company submission (assuming 100% of patients in PFS at 30 months are assumed to move to long term remission). This is further described in Section 6.1.4.1.

Additional therapy effects, including subsequent therapies, discontinuation and adverse events, are applied in the PSM in the same way as the original Markov model. For subsequent therapies, as a simplifying assumption, it is assumed that all patients receive single agent taxane as subsequent therapy in the base case. BSC is defined as 50% of patients receiving paclitaxel and the other 50% receiving docetaxel. Discontinuation is informed by a time on treatment curve, with the timing of discontinuation impacting treatment costs and resource use. Additionally, a two-year stopping rule is present for NIVO + CHEMO. Treatment-related AEs from the safety population were applied within the PSM and Markov model as described in the original company submission, using a one-off cost and disutility in the first cycle only.

6.1.2 Modelled patient population

6.1.2.1 Patients with PD-L1 CPS ≥5

As requested in the Appraisal Consultation Document and based on the Committee for Medicinal Products for Human Use (CHMP) modified licensed indication, the economic evaluation considers the use of [redacted].

6.1.2.2 Baseline demographics

Following the Technical Engagement process, the updated baseline patient parameters are presented in Table 5.

Table 5. Baseline patient parameters

Parameter	Mean	SE	Source
Base case analysis			
Baseline age, years	64.15	12.02	ERG age CRUK
Proportion of cohort male, %	69.5	30.8	Baseline overall ITT population
SE: standard error.			

6.1.3 Comparators

The Appraisal Consultation Document states that XELOX is the most relevant comparator for this appraisal. Hence, the company base case analysis presented in this response focuses on NIVO+XELOX versus XELOX.

6.1.4 Clinical outcomes

As requested in the Appraisal Consultation Document, the most recent database lock ([redacted]) has been used to inform clinical effectiveness for patients with PD-L1 CPS ≥5.[2]

The clinical results are presented in Section 1, with NIVO+CHEMO showing statistically significant and clinically meaningful improvements in PFS per BICR and OS in all randomised patients with PD-L1 CPS ≥5.

Despite poor prognosis for the average patient, a small proportion of patients with locally advanced or metastatic GC demonstrate improved outcomes versus the overall cohort. This effect is demonstrated in several studies of patients receiving standard chemotherapy or symptom control.[4-7] Further, it has been demonstrated that this proportion of patients is increased in patients receiving nivolumab.[4,5]

CheckMate 649 patients receiving NIVO+CHEMO demonstrated the same reduction in long-term hazard observed in patients receiving standard chemotherapy, with no death events observed following 30 months. However, importantly, the proportion of patients with prolonged survival is increased in the NIVO+CHEMO arm: OS at one year was 57.0% (versus 46.4% for CHEMO), 31.0% at two years (versus 18.6% for CHEMO) and [redacted] at three years (versus [redacted] for CHEMO).[2] These patients with prolonged survival indicate that NIVO+CHEMO increases the proportion of patients entering long-term remission, which can be considered a vital potential benefit for NIVO+CHEMO therapy.

Table 6 compares the previously predicted OS outcomes to observed OS outcomes from CheckMate 649 at 12, 24, 36 and 60 months, for the company's base case presented in the NICE submission, the ERG base case, and the company's base case during technical engagement. Additionally, the OS rates for the new base case from the partitioned survival model (PSM) are presented as well as for the semi-Markov model which includes long-term remission.

It should be noted that the previous database lock (July 2020) was used for the company submission, the Evidence Review Group (ERG) response and for the technical engagement. Additionally, starting from the ERG base case, the age of the patients in the economic model was increased leading to a higher all-cause mortality and hence, reduced OS.

Table 6. Comparison of previously predicted overall survival outcomes versus observed outcomes from CheckMate 649 [redacted] database lock for PD-L1 CPS ≥5

OS Rates (%)		12 months	24 months	36 months	60 months
NIVO+ CHEMO	Observed during CheckMate 649[2]	57.0	31.0	[redacted]	[redacted]
	Company submission base case	[redacted]	[redacted]	[redacted]	[redacted]
	ERG base case	[redacted]	[redacted]	[redacted]	[redacted]
	Company technical engagement base case	[redacted]	[redacted]	[redacted]	[redacted]
	Company ACD response base case: PSM	[redacted]	[redacted]	[redacted]	[redacted]
	Company ACD response: long-term remission	[redacted]	[redacted]	[redacted]	[redacted]
CHEMO	Observed during CheckMate 649[2]	46.4	18.6	[redacted]	[redacted]
	Company submission base case	[redacted]	[redacted]	[redacted]	[redacted]
	ERG base case	[redacted]	[redacted]	[redacted]	[redacted]
	Company technical engagement base case	[redacted]	[redacted]	[redacted]	[redacted]
	Company ACD response base case: PSM	[redacted]	[redacted]	[redacted]	[redacted]
	Company ACD response: long-term remission	[redacted]	[redacted]	[redacted]	[redacted]

ACD: Appraisal Consultation Document; CHEMO: chemotherapy; ERG: Evidence Review Group; NIVO: nivolumab; OS: overall survival; PSM: partitioned survival model.
[redacted]

Since the PSM uses survival data directly from the clinical trial, it is expected that PSM predictions for OS match the observed CheckMate 649 better than the semi-Markov model predictions.

The PSM slightly underestimates OS for NIVO+CHEMO at month 12 and month 36, compared to CheckMate 649 values, with the latter (36 months) due to the extrapolated tail of the OS data. Overall, the OS estimates at 36 months are broadly consistent across the different base cases in the NIVO+CHEMO arm. In contrast, the OS predictions for CHEMO in the PSM base case at 24 and 36 months, respectively, are above the observed values in CheckMate 649. These CHEMO PSM estimates are lower than the OS predictions in other models, but still closer to the observed OS when compared to the other base cases. No observed data is available yet for 60 months, but when compared to other clinical trials such as the Royal Marsden study or Chau et al. which show an OS of 4.1% (1L treatment of OC/GEJC/GC) and 4.0% (for GC) at 5 years, respectively,[6,8] the prediction of [redacted]% for CHEMO in PSM base case is equivalent. Overall, the PSM provides a closer match to observed OS data, but overestimates OS for chemotherapy, and underestimates OS for nivolumab, and as such is a conservative base case.

6.1.4.1 Evidence to support long-term remission

Treatment of solid tumours with immunotherapies (IO) in advanced and metastatic settings has resulted in outcomes that are challenging to model with traditional techniques. In particular, the response characteristics of IO therapies differ from traditional systemic oncology treatments; notably, the delayed onset of treatment effects and the potential for long-term remission.[9]

These features of the response to IO therapy were considered when developing models to extrapolate the survival outcomes of CheckMate 649, as part of the model selection algorithm outlined by the NICE Decision Support Unit (DSU)[3] and in line with guidance relating to the difficulties in modelling survival outcomes in immuno-oncology, published recently by the DSU.[10]

Mixture/mixture-cure models were considered potentially appropriate for the heterogeneity of survival time in the observed data and were examined for appropriateness. Mixture-cure models assume that one subpopulation is at no risk of a specific event. In the relative survival context, it is the excess hazard of event due to disease that is absent in the “cured” subpopulation. All nominated statistical models were evaluated as potential fits for the “uncured” population, with the exception of the generalised gamma.

As an alternative to assuming zero excess hazard in the “cured” population, a secondary distribution was specified in addition to the background survival distribution. To limit the number of potential models, this was limited in all cases to a Weibull model, whilst the first excess hazard was fitted in turn by each of the nominated statistical models except the generalised gamma. In the majority of cases, however, these models converged to implicit/statistical cure models where the excess hazard in one sub-population was negligible.

For PFS, the mixed-cure models assuming that a proportion of the cohort are within a long-term responsive fraction from randomisation demonstrated similar features to the parametric relative survival models for the NIVO+CHEMO arm. A long-term response fraction was not always identified, particularly for the non-monotonic hazard functions. With the exception of the exponential and Gompertz models, fit to the trial data was acceptable. For the CHEMO arm, the same conclusions are drawn, i.e. that all fits with the exception of the exponential and

Gompertz models are acceptable, but the long-term response fraction is dependent upon the distribution. These models did not introduce an exploration of uncertainty in a controlled way and so were not used.

When considering OS, mixed-cure models assuming a proportion of the cohort are within a long-term surviving fraction from randomisation demonstrated better goodness of fit than the standard statistical models. For NIVO+CHEMO, the rate of long-term survival varied depending upon the hazard function of the non-long-term survival fraction. For the CHEMO arm, fits to the trial data were not as well centred, and for the lognormal and log-logistic distributions demonstrated negligible long-term survival fraction. Distributions with a non-negligible cure fraction had expected survival in excess of two years, which may be unexpectedly high from a clinical perspective.

After assessing the appropriateness of alternative models, semi-parametric models were chosen for survival outcomes for both arms due to the high rate of change of hazard over the initial period after commencing treatment, followed by a more consistent period. Given the clinical expectation that a small number of responsive patients may enter long-term remission after treatment with nivolumab, the long-term survival benefit produced by this semi-parametric modelling approach may be conservative.

The long-term remission health state included in the semi-Markov model captures those patients still progression-free after a specified period of time and applies general population mortality rates instead of disease-specific mortality. This assumption is based on three key aspects: that long-term remission is plausible in the advanced gastric cancer population; that patients in the long-term remission health state have a mortality hazard similar to the general population and that those patients reach long-term remission at about 30 months.

The evidence to support the plausibility of long-term remission is primarily derived from the published literature and supporting evidence from CheckMate 649. Several primarily UK-based studies have demonstrated that a small proportion of gastric cancer patients receiving chemotherapy may survive for a number of years.[6-8,11] These patients have a high initial hazard followed by low hazard with limited events hereafter, which indicates the potential for a prolonged survival and/or long-term remission state in a small number of patients.

Aligned with this evidence, CheckMate 649 reported short median OS (11.1 months,

Table 2) for patients receiving chemotherapy (XELOX/FOLFOX).[12] However, as outlined in Figure 1, a small proportion of patients have prolonged survival, evidenced by very low hazard during the long-term follow-up. Patients receiving standard chemotherapy demonstrated 46% OS at one year and 19% surviving at two years.[12] The observed Kaplan-Meier data indicate that a proportion of patients may enter long-term remission in clinical practice.

Patients receiving NIVO+CHEMO demonstrated extended median OS benefit (14.4 months versus 11.1 months,

Table 2). However, importantly, the proportion of patients with prolonged survival is increased in the NIVO+CHEMO arm: OS at one year was 57.0% (versus 46% for CHEMO), and 31% at two years (versus 19% for CHEMO).[12] These patients with prolonged survival indicate that NIVO+CHEMO increases the proportion of patients entering long-term remission, which can be considered a vital potential benefit for NIVO+CHEMO therapy.

Evidence to support specific outcomes for patients in long-term remission is sparse. However, supporting evidence for the assumption that these patients have a mortality hazard similar to the general population is provided by the published literature, where few deaths are observed during long-term follow up, independent of the treatment received.[6-8,11]

The assumption that patients reach long-term remission at 30 months is primarily informed by CheckMate 649, as this study has large patient numbers and patient-level data is available, so it is possible to assess the precise hard profile and identify the hazard turning point.

External validation of these assumptions was sought. In a pre-specified analysis using the US Flatiron database,[13] the outcomes of patients with advanced gastric, gastro-intestinal junction or oesophageal cancers with HER2 status negative or unknown who initiated first line systemic therapy were sought.[13] Table 7 details the attrition of patients due to these selection criteria from the database.

Table 7. Patients selected from Flatiron database

Patient Selection Criteria	Flatiron	
	Patient Count	%
Inclusion Criteria		
Patients diagnosed with advanced GC, GEJC, or EAC between January 1, 2011 and June 30, 2021. The initial advanced diagnosis date is defined as index date.	6,595	100
Patients with confirmed adenocarcinoma (only apply to GC/GEJC patients)	6,416	97
Patients with 1L Chemo treatment initiating between 1/1/2011 and 11/30/2018	2,723	41
Exclusion Criteria		
Patients with clinical trial study medications any time during the study period	2,642	40
Patients with other primary cancers prior to the index date	2,618	40
Patients with IO on any line of treatment	2,397	36
HER2 positive patients and patients treated with trastuzumab	2,188	33
Patients progression-free at 30 months	74	
Patients alive at 30 months	158	
HER2 Negative	1,541	
Tested Unknown/Untested	647	
Total sample size	2,188	

The demographics of this cohort were compared to those in the CPS ≥ 5 subgroup of CheckMate 649, as presented in Table 8. The mean age was slightly higher, though in line with the submitted model population. CheckMate 649 included a higher proportion of Asian patients (24.7% vs 5%), had fewer oesophageal adenocarcinomas (12.4% vs 31%) and had a higher proportion of patients who were at stage IV at diagnosis (81.3% vs 75%). Among the 1,173 patients who had a valid ECOG assessment recorded at initiation of first line therapy, 236 (20.1%) were assessed as ECOG PS ≥ 2.

Table 8. Demographic and baseline characteristics of selected Flatiron population and CheckMate 649 CPS ≥ 5 population

Baseline characteristics	Flatiron, HER2 negative, unknown or untested, at initiation of 1st line[13]	CM 649 CPS ≥ 5[14] N = 955

		N = 2,188	
Age	Mean (sd)	64 (11.1)	60.9 (NR)
	< 65 (%)	1,022 (47)	552 (57.8)
Sex	Male (%)	1,640 (75)	680 (71.2)
	Female (%)	548 (25)	275 (28.8)
Race	White (%)	1,364 (62)	655 (68.6)
	Black or African American (%)	154 (7)	9 (0.9)
	Asian (%)	101 (5)	236 (24.7)
	Other (%)	320 (15)	35 (3.7)
	Missing (%)	249 (11)	N/A
Initial diagnosis	Gastric / Gastroesophageal junction cancer (%)	1,520 (69)	837 (87.6)
	Oesophageal adenocarcinoma (%)	668 (31)	118 (12.4)
Disease stage at initial diagnosis	Stage I (%)	43 (2)	8 (0.8)
	Stage II (%)	127 (6)	38 (4.0)
	Stage III (%)	245 (11)	131 (13.7)
	Stage IV (%)	1,648 (75)	776 (81.3)
	Unknown/not documented (%)	125 (6)	2 (0.2)

OS from initiation of first-line systemic therapy and PFS from initiation of first-line systemic therapy were defined with both censored at the data cut-off point. Median follow-up was 75 months, with minimum 32 and maximum 127 months. There was a significant difference per logrank test in the OS dependent upon HER2 status being known negative or unknown (Figure

5); however, both subgroups demonstrated a flattening in the tail upon conventional therapies, and the survival curves were indistinguishable from 2 years.

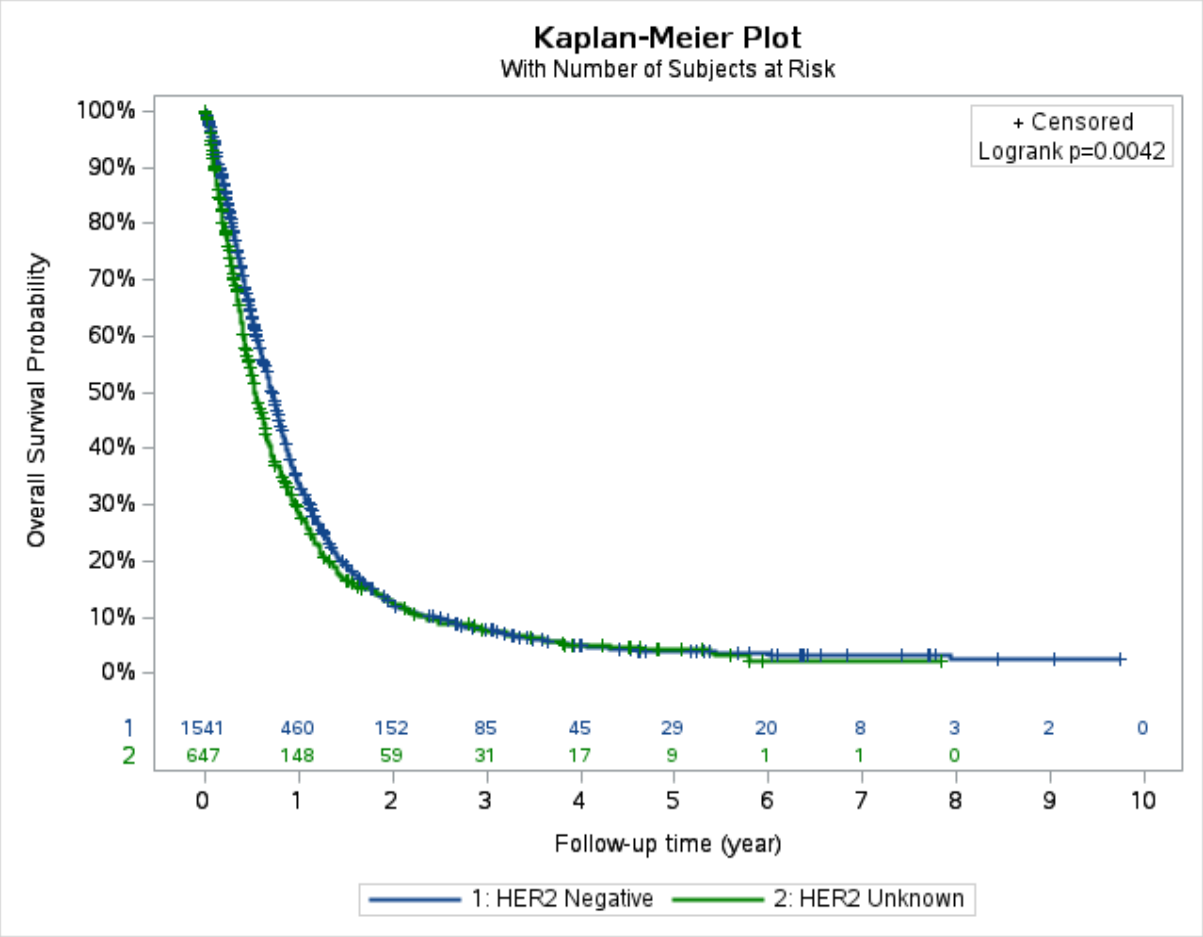

Figure 5. Overall Survival from initiation of first-line therapy by HER 2 status, Flatiron population

Examining the cohort who had survived in PFS to 30 months, OS decreased to 3.5 years after this landmark. By this point, OS had decreased to 48% (7.8% standard error), and no further events were observed to the end of potential follow-up at approximately 7 years after landmark (Figure 6). This suggests that, under current standard of care, approximately 50% of patients being progression-free at 2.5 years may experience long-term remission.

It is necessary to caution that the number of patients informing the analysis from 3.5 to 7 years is very low. It should also be noted that these patients are among the earliest treated in the extracted database; for 6 years of follow-up (2.5 + 3.5 years) patients must have initiated first line therapy prior to July 31, 2015 (the data cut off for the US Flatiron database was June 30, 2021).

Despite these caveats, the existence of a small proportion of long-term responders is consistent with these data. Identification of this fraction under current standard of care may require more than 2.5 years of follow-up. Treatment with nivolumab plus chemotherapy may

promote long-term response, and this may be partially responsible for an increase in overall survival. It is not necessary that the point of identification of the LTR fraction is the same in patients treated with nivolumab due to the novel mechanism of action of this therapy.

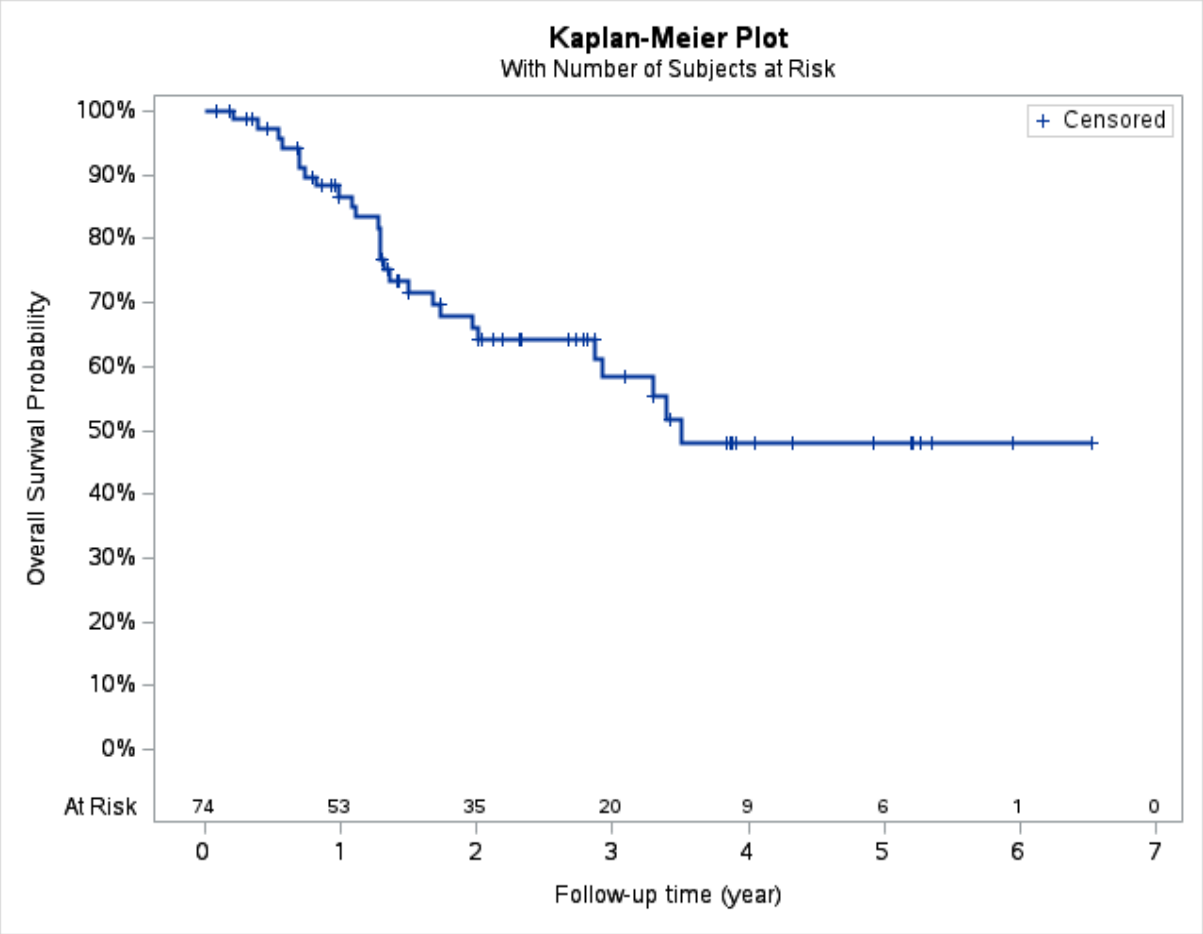

Figure 6. Overall survival after 30 months for patients who had not progressed by 30 months, HER2 negative or unknown, Flatiron population

Therefore, an LTR fraction may exist under existing chemotherapies and this fraction experiences extended overall survival with mortality rates consistent with the general population. This fraction may be a substantial fraction of those progression-free at 2.5 years and may differ for patients treated with nivolumab + chemotherapy.

In summary, there is uncertainty regarding the portion of patients experiencing LTR due to low patient numbers and limited evidence from clinical trials and therapies. The clinical expert consulted for the ACD stated that approximately 4% of people could be expected to achieve LTR with chemotherapy which is expected to at least double with additional treatment with nivolumab. The NHS England clinical lead noted that patients in LTR can relapse but that this is uncommon.[1]

6.1.4.2 Extrapolation of overall survival

Within the partitioned survival model, OS is modelled directly. Models were informed by patient-level data from CheckMate 649. Model selection was undertaken following guidance

provided by the NICE Decision Support Unit (DSU) in technical support documents (TSD) 14 and 21.[15,16] Details on survival modelling are provided in Appendix B.

In brief, the following considerations were made:

1. Due to delay in separation of OS curves and the novel mechanism of action of nivolumab, a scaling relationship (proportional hazards, accelerated failure time) was not imposed upon the model and both arms were modelled independently

2. Despite not imposing a scaling relationship upon the models, it was assumed that the hazard functions of both arms would be sufficiently similar as to be described by the same distribution type, with parameters allowed to vary. Therefore, if a long-tailed log-logistic distribution was chosen for one arm, it would follow that the log-logistic distribution would be chosen for the other arm.

3. Due to the expectation of a long-term decreasing hazard function and the expected need to modify survival curves in the tail to account for the general population rate of mortality in the aging cohort, a relative survival approach was undertaken where the additional hazard due to disease predicted by the model is added to matched general population mortality hazard to form the overall hazard of mortality. Accounting for this at the analytical stage prevented double-counting of this hazard when forming model predictions.

4. An attempt was made to model the disease-specific hazard of mortality by conventional parametric models of the types specified by TSD 14, i.e. exponential, Weibull, Gompertz, lognormal, log-logistic and generalised gamma.[16] These models were insufficiently flexible to represent the changes in hazard demonstrated over trial follow-up within CheckMate 649.

5. A delay to the start of the parametric extrapolation was imposed, in line with the delay used for the model of PFS. These “semi-parametric” models demonstrated much better fit to the tail of the data, and the Gompertz model was chosen as the superlative model by AIC (Δ AIC ~ 2) in both arms.

6. Mixture-cure relative survival models were fitted to the data and demonstrated better goodness-of-fit than equivalent simple parametric relative survival models. However, due to the relative immaturity of the OS data in comparison to the PFS data used to inform the assumption of long-term remission in the semi-Markov model, these models were not used in the base case for the PSM.

The selected models for the NIVO+CHEMO and CHEMO arms are shown in Figure 7 and Figure 8, respectively.

Nivo + Chemo CPS ≥ 5%, semiparametric relative survival cut at 6.44 months

Figure 7. OS, NIVO+CHEMO CPS≥5%, semi-parametric (piecewise) relative survival models. Gompertz chosen for base case in PSM

Chemo CPS ≥ 5%, semiparametric relative survival cut at 6.44 months

Figure 8. OS, CHEMO CPS≥5%, semi-parametric (piecewise) relative survival models. Gompertz chosen for base case in PSM

Semi-Markov Model

Overall survival in the SMM is informed by a combination of:

1. transitions to death from the pre-progression state
2. transitions to death from the pre-progression (long-term responder) state
3. transition to death from the post-progression state

These transition rates are informed by:

1. Transitions out of pre-progression (informed by PFS model) x probability of PFS event being death (informed by logistic regression)
2. Equal to life-table mortality multiplied by a mortality ratio (assumed in base case to be 1)
3. Post-progression survival (PPS)

The PFS models are described in Section 6.1.4.3.

The ratio of death to progression events informing the PFS curve, dependent upon time from randomisation, was informed by a logistic regression model upon patient-level data from CheckMate 649 as described in Appendix B. This model was fitted to PFS event data as assessed by BICR, without censoring for subsequent treatment, from the CPS ≥ 5% cohort. Independent models were fitted per arm, with time of event as the only dependent variable. Various combinations of transformation of time were attempted, but that consisting of an intercept term, linear time and the natural logarithm of time was superlative by AIC and chosen. This model had the expected characteristic that long-term, the probability of a PFS event being death increased and could represent the growing influence of non-disease-specific causes of death in extrapolation, as shown in Figure 9.

Figure 9. Model predictions - probability that PFS event is death

Thin lines – logistic model predictions. Dashed lines – smoothing generalised additive models

Models of PPS were informed by patient-level data from CheckMate 649 in the CPS ≥ 5% cohort, from time of progression per BICR. Model selection was undertaken following guidance provided in TSD 14 and 21,[15,16] as detailed in Appendix B. In brief:

1. Due to divergence of PPS curves on the Weibull plot, a scaling relationship (proportional hazards, accelerated failure time) was not assumed.
2. Despite not imposing a scaling relationship upon the models, it was assumed that the hazard functions of both arms would be sufficiently similar as to be described by the same distribution type, with parameters allowed to vary, as with OS for the PSM.

3. Due to the expectation of a long-term decreasing hazard function and the expected need to modify survival curves in the tail to account for the general population rate of mortality in the aging cohort, a relative survival approach was undertaken where the additional hazard due to disease predicted by the model is added to matched general population mortality hazard to form the overall hazard of mortality. Accounting for this at the analytical stage prevented double-counting of this hazard when forming model predictions. Fitting this relative survival model necessitated calculating the age at progression of each patient in order to calculate their age at death.

4. An attempt was made to model the disease-specific hazard of mortality by conventional parametric models of the types specified by TSD 14, i.e. exponential, Weibull, Gompertz, lognormal, log-logistic and generalised gamma.[16] These models were sufficient to fit this endpoint, and the log-logistic model was chosen due to its low BIC and best approximation the tail of the cumulative hazard function.

The selected models for the NIVO+CHEMO and CHEMO arms are shown in Figure 10 and Figure 11, respectively.

Nivo+Chemo CPS ≥ 5% Progressed (BICR) (2), parametric relative survival

Figure 10. PPS, NIVO+CHEMO CPS≥5%, parametric relative survival models. Log-logistic chosen for base case in SMM

Figure 11. PPS, CHEMO CPS≥5%, parametric relative survival models. Log-logistic chosen for base case in SMM

6.1.4.3 Extrapolation of progression-free survival

Models of progression-free survival were required for both the SMM and PSM, and the same model could be applied in each. Models were informed by patient-level data from CheckMate 649, using the CPS ≥ 5% subgroup and the PFS per BICR outcome. Model selection was undertaken following guidance provided by TSD 14 and 21.[15,16] Details on survival modelling are provided in Appendix B.

In brief, the following considerations were made:

1. Due to delay in separation of PFS curves and the novel mechanism of action of nivolumab, a scaling relationship (proportional hazards, accelerated failure time) was not imposed upon the model and both arms were modelled independently

2. Despite not imposing a scaling relationship upon the models, it was assumed that the hazard functions of both arms would be sufficiently similar as to be described by the same distribution type, with parameters allowed to vary.

3. Due to the expectation of a long-term decreasing hazard function and the expected need to modify survival curves in the tail to account for the general population rate of mortality in the aging cohort, a relative survival approach was undertaken where the additional hazard due to disease predicted by the model is added to matched general population mortality hazard to form the overall hazard of mortality. Accounting for this at the analytical stage prevented double-counting of this hazard when forming model predictions.

4. An attempt was made to model the disease-specific hazard of mortality by conventional parametric models of the types specified by TSD 14, i.e. exponential, Weibull, gompertz, lognormal, log-logistic and generalised gamma.[16] These models were insufficiently flexible to represent the changes in hazard demonstrated over trial follow-up within CheckMate 649 (as with OS), and were particularly unable to represent the protocol-driven "steps" in the Kaplan-Meier estimator that represented standard assessment times, which were more pronounced in the initial study period.

5. A delay to the start of the parametric extrapolation was imposed, with 6.44 months (representing the end of the high-frequency assessment period at trial start + 3 weeks) chosen as an appropriate delay to ensure modelling of a continuous hazard function. Models fitted after this point were sufficiently flexible to represent the survival data (with the exception of the exponential model), and the lognormal model was chosen for base case due to its low AIC (Δ AIC ~ 2).

6. Mixture-cure relative survival models were fitted to the data and demonstrated better goodness-of-fit than equivalent simple parametric relative survival models. Due to the variation in LTR fraction dependent upon model selection, the delayed identification of LTR approach used in the economic model was preferred, but these models supported the use of an LTR fraction.

The selected models for the NIVO+CHEMO and CHEMO arms are shown in Figure 7 and Figure 8, respectively.

Nivo + Chemo CPS ≥ 5%, semiparametric relative survival cut at 6.44 months

[redacted]

Figure 12. PFS (BICR), NIVO+CHEMO CPS≥5%, parametric relative survival models. Log-normal chosen for base case in SMM

[redacted]Figure 13. PFS (BICR), CHEMO CPS≥5%, parametric relative survival models. Log-normal chosen for base case in SMM

6.1.4.4 Time on treatment

Models of progression-free survival were required for both the SMM and PSM, and the same model could be applied in each. Models were informed by patient-level data from CheckMate 649, using the CPS ≥ 5% subgroup and the time to treatment discontinuation (TTD) outcome.

Model selection was undertaken following guidance provided by TSD 14 and 21.[15,16] Details on survival modelling are provided in Appendix B.

These data were almost complete, and for NIVO+CHEMO had a marked discontinuity at 24 months due to treatment protocol within CheckMate 649. As such, direct representation of the survival function via the Kaplan-Meier estimator was preferred, although parametric relative survival models were also fitted, as per the considerations of PFS, as a scenario.

[redacted]

Figure 14 TTD, CHEMO CPS≥5%, Kaplan-Meier estimator.

Due to delays, not all doses predicted by the area under the TTD curve were received within the trial. As this represents a real-world modification of the expected number of doses dispensed, a modification to the treatment costs was calculated based upon the ratio of the expected and dispensed number of doses within the trial. The relative dose intensity for each treatment component was used to scale the mean cost of the combined treatment, via the ratios in Table 9.

Table 9. Treatment modifier values

Treatment:	Component	Treatment modifier value
FOLFOX	5-fluorouracil	[redacted]
	leucovorin	[redacted]
	oxaliplatin	[redacted]
	5-fluorouracil continuous	[redacted]
XELOX	oxaliplatin	[redacted]
	capecitabine	[redacted]
NIVO+FOLFOX	nivolumab	[redacted]
	5-fluorouracil	[redacted]
	leucovorin	[redacted]
	oxaliplatin	[redacted]
	5-fluorouracil continuous	[redacted]
NIVO+XELOX	nivolumab	[redacted]
	oxaliplatin	[redacted]
	capecitabine	[redacted]

6.1.5 Health-related quality of life

The health utility of patients is dependent upon their disease state and so consequently, during each cycle, patients are assigned the health utility value equivalent to their current disease

state based on the results from the CheckMate 649 trial. In both models, utility values were updated to incorporate analysis from the [redacted] database lock (Table 10).

Table 10. Health state utility values informed by CheckMate 649 [redacted] database lock

	Mean	SE
Pre-progression	[redacted]	[redacted]
Progressed	[redacted]	[redacted]
Mean data sourced from CheckMate 649, SE 20% of mean SE = standard error		

6.1.6 Cost and healthcare resource use

6.1.6.1 Updated Patient Access Scheme

The agreed Patient Access Scheme (PAS) for nivolumab has been updated from [redacted]% to [redacted]% impacting on vial costs as follows:

- Nivolumab costs without PAS[17]
    - £2,633.00 per 240 mg (24 mL) vial;
    - £1,097.00 per 100 mg (10 mL) vial;
    - £439.00 per 40 mg (4 mL) vial.
- Nivolumab costs with PAS
    - [redacted] per 240 mg (24 mL) vial;
    - [redacted] per 100 mg (10 mL) vial;
    - [redacted] per 40 mg (4 mL) vial.

This updated PAS has been applied within this response.

6.1.6.2 Cost of PD-L1 testing

Aligned with the Appraisal Consultation Document and the licensed indication for nivolumab, the updated analysis includes the cost of PD-L1 testing.

The assay used in the clinical trial was Dako IHC-28-8 pharmDX. Agilent Dako provided the costs on request with 50 tests at a price of £4,104.00, leading to £82.08 per test.

6.2 Updated economic analysis outcomes

6.2.1 Partitioned survival model analysis

6.2.1.1 Base case analysis results

Total discounted costs associated with NIVO+XELOX (with PAS), accrued over the modelled time horizon, were predicted to be [redacted] for NIVO+XELOX. By comparison, total discounted costs associated with comparators were notably lower. Incremental discounted costs for NIVO+XELOX were predicted to be [redacted] (versus XELOX), under base case assumptions. The incremental discounted QALYs for NIVO+XELOX were predicted to be [redacted] (versus XELOX. The resulting ICER estimates for NIVO+XELOX were £45,383 per QALY (NIVO+XELOX versus XELOX).

The results of the base-case analysis are summarised in Table 11.

Table 11. NIVO+XELOX base-case results ([redacted] PAS)

	NIVO+XELOX	XELOX
Patient level survival (undiscounted)		
Median ToT (years)*	[redacted]	0.383
Mean ToT (years)*	[redacted]	0.554
Median PFS (years)	[redacted]	0.537
Mean PFS (years)	[redacted]	1.109
Median OS (years)	[redacted]	0.920
Mean OS (years)	[redacted]	1.531
Patient-level progression		
Time in pre-progression (years)	[redacted]	1.109
Time in long term remission (years)	[redacted]	0.000
Time in post-progression (years)	[redacted]	0.422
Costs (with PAS)		
HS costs	[redacted]	£9,256
Treatment costs	[redacted]	£3,545
AE costs for initial therapy	[redacted]	£429
Discontinuation costs	[redacted]	£43
Death costs	[redacted]	£5,256
Total costs	[redacted]	£18,529
Health benefits		
HS QALYs	[redacted]	1.114
Age-dependent utility	[redacted]	0.000
Adverse event utility	[redacted]	-0.001
Time-to-death utility	[redacted]	-0.068
Total QALYs	[redacted]	1.045
Total LYs (undiscounted)	[redacted]	1.531
Incremental results		
Incremental total costs		[redacted]
Incremental QALYs		[redacted]
Incremental LYs (undiscounted)		[redacted]
Cost/QALY		**£45,383**

AE: adverse event; HS: health state; LY: life year; OS: overall survival; PFS: progression free survival QALY: quality-adjusted life year; ToT: Time on Treatment.

6.2.1.2 Deterministic sensitivity analysis

Results of the deterministic sensitivity analysis are presented in Figure 15. These figures demonstrate the impact of specific parameters on ICER estimates. The factors with the greatest impact on the ICER were discounting, baseline age of patients and health state utilities.

For the majority of scenarios, the ICER for NIVO+CHEMO versus XELOX stayed near the £50,000 per QALY willingness-to-pay threshold; scenarios where the ICER exceeded the

£50,000 threshold included increasing the benefits discounting, as well as increasing the baseline age of patients and the health state utility utilities.

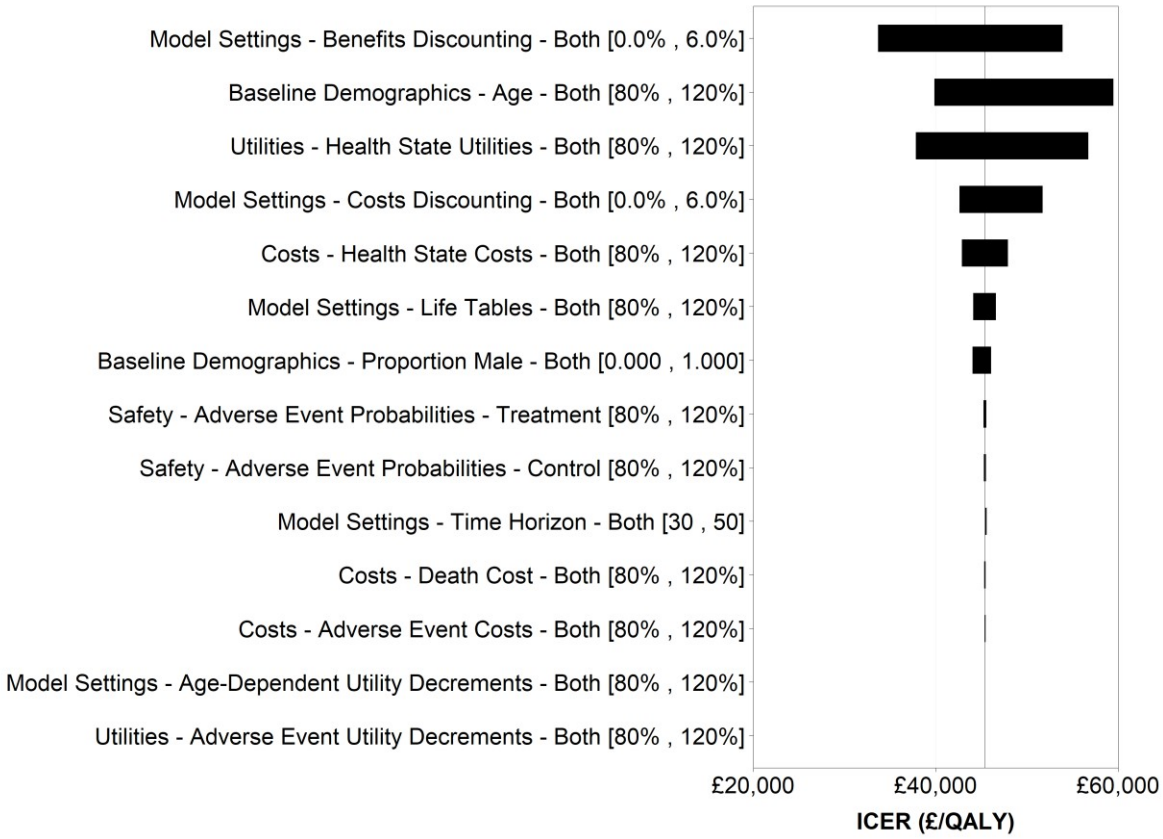

Figure 15. Deterministic sensitivity analysis for nivolumab + XELOX versus XELOX: impact on ICER

6.2.1.3 Probabilistic sensitivity analysis

The ICER scatterplot for the base case analysis, arising from 1,000 simulations of the model with all parameters sampled is presented in Figure 16 while the cost-effectiveness acceptability curve (CEAC) is presented in Figure 17.

Figure 16. ICER scatterplot: Nivolumab + XELOX versus XELOX

Figure 17. Cost-effectiveness acceptability curve: Nivolumab + XELOX versus XELOX

Based on this analysis, the probability that nivolumab + XELOX is cost-effective versus XELOX is estimated to be [redacted] at a willingness-to-pay threshold of £50,000 per QALY. The base case results are presented in Table 12.

Table 12. Base case results (probabilistic): Nivolumab + XELOX versus XELOX

Technology	Total costs (£)	Total life years	Total QALYs	Inc. costs (£)	Inc. life years	Inc. QALYs	ICER (£/QALY)
Nivolumab + XELOX	[redacted]	[redacted]	[redacted]	-	-	-	
XELOX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£47,873
ICER: incremental cost-effectiveness ratio; QALY: quality-adjusted life year							

6.2.2 Semi-Markov model with long-term remission population

The semi-Markov model assumes all patients who remain progression-free at 30 months move to a long-term remission state, where no further healthcare resource use or treatment costs are applied, and risk of death is equivalent to general population.

Table 13 shows the total discounted costs associated with NIVO+XELOX (with PAS), accrued over the modelled time horizon, were predicted to be [redacted] for NIVO+XELOX. By comparison, total discounted costs associated with comparators were notably lower. Incremental discounted costs for NIVO+XELOX were predicted to be [redacted] (versus XELOX), under base case assumptions. The incremental discounted QALYs for NIVO+XELOX were predicted to be [redacted] (versus XELOX. The resulting ICER estimates for NIVO+XELOX were £34,110 per QALY (NIVO+XELOX versus XELOX).

Table 13. NIVO+XELOX vs XELOX – semi-Markov model with long-term remission

Technology	Total costs (£)	Total life years	Total QALYs	Inc. costs (£)	Inc. life years	Inc. QALYs	ICER (£/QALY)
Nivolumab + XELOX	[redacted]	[redacted]	[redacted]	-	-	-	-
XELOX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£34,110

6.2.3 Scenario analysis

6.2.3.1 Alternative survival extrapolations

As the primary model is the PSM, scenarios will be presented for this model using the data-cut off at 6.44 months. Base case choices for PFS and OS extrapolations in the PSM were determined based on the best available statistical fit to the data. Alternative survival extrapolations were explored in scenario analyses. The impact of differing PFS and OS extrapolations were evaluated in turn, first by maintaining the base case PFS extrapolation and varying OS extrapolation, and then by maintaining the base case OS extrapolation and varying PFS extrapolation. In each scenario, the same extrapolation distributions were used for both treatment and control arms. The results of these different scenarios are presented in the tables below.

Most of the scenario analyses showed that NIVO+XELOX is cost-effective or close to being cost-effective.

A. PFS lognormal

a. OS log-logistic

Using log-logistic to extrapolate overall survival led to a decrease in incremental costs (£[redacted] in the base case to £[redacted]) for NIVO+CHEMO vs XELOX and a decrease in incremental QALYs as well, from [redacted] in the base case to [redacted]. The ICER increased compared to the base case, from £45,383 per QALY to £56,672 per QALY.

Table 14. Log-logistic distribution for OS

Technology	Total costs (£)	Total life years	Total QALYs	Inc. costs (£)	Inc. life years	Inc. QALYs	ICER (£/QALY)
Nivolumab + XELOX	[redacted]	[redacted]	[redacted]	-	-	-	-
XELOX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£56,672

b. OS Weibull

Using Weibull to extrapolate overall survival led to a decrease in incremental costs (£[redacted] in the base case to £[redacted]) for NIVO+CHEMO vs XELOX and a decrease in incremental QALYs as well, from [redacted] in the base case to [redacted]. The ICER increased compared to the base case, from £45,383 per QALY to £71,457 per QALY.

Table 15. Weibull distribution for OS

Technology	Total costs (£)	Total life years	Total QALYs	Inc. costs (£)	Inc. life years	Inc. QALYs	ICER (£/QALY)
Nivolumab + XELOX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	-
XELOX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£71,457

c. OS Lognormal

Using lognormal to extrapolate overall survival led to a decrease in incremental costs (£[redacted] in the base case to £[redacted]) for NIVO+CHEMO vs XELOX and a decrease in incremental QALYs as well, from [redacted] in the base case to [redacted]. The ICER increased compared to the base case, from £45,383 per QALY to £53,803 per QALY.

Table 16. Lognormal distribution for OS

Technology	Total costs (£)	Total life years	Total QALYs	Inc. costs (£)	Inc. life years	Inc. QALYs	ICER (£/QALY)
Nivolumab + XELOX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	-
XELOX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£53,803

d. OS Exponential

Using exponential to extrapolate overall survival led to a decrease in incremental costs (£[redacted] in the base case to £[redacted]) for NIVO+CHEMO vs XELOX and a decrease in incremental QALYs, from [redacted] in the base case to [redacted]. The ICER increased compared to the base case, from £45,383 per QALY to £82,249 per QALY.

Table 17. Exponential distribution for OS

Technology	Total costs (£)	Total life years	Total QALYs	Inc. costs (£)	Inc. life years	Inc. QALYs	ICER (£/QALY)
Nivolumab + XELOX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	-
XELOX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£82,249

e. OS Generalised gamma

Using generalised gamma to extrapolate overall survival led to a decrease in incremental costs (£[redacted] in the base case to £[redacted]) for NIVO+CHEMO vs XELOX and a decrease in incremental QALYs as well, from [redacted] in the base case to [redacted]. The ICER increased compared to the base case, from £45,383 per QALY to £62,717 per QALY.

Table 18. Generalised gamma distribution for OS

Technology	Total costs (£)	Total life years	Total QALYs	Inc. costs (£)	Inc. life years	Inc. QALYs	ICER (£/QALY)
Nivolumab + XELOX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	-
XELOX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£62,717

B. OS Gompertz

a. PFS log-logistic

Using log-logistic for extrapolating progression-free survival had little impact on incremental costs (a decrease from £[redacted] in the base case to £[redacted]) for NIVO+CHEMO vs XELOX. The incremental QALYs increased slightly from [redacted] in the base case to [redacted]. The ICER decreased compared to the base case, from £45,383 per QALY to £44,576 per QALY.

Table 19. Log-logistic distribution for PFS

Technology	Total costs (£)	Total life years	Total QALYs	Inc. costs (£)	Inc. life years	Inc. QALYs	ICER (£/QALY)
Nivolumab + XELOX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	-
XELOX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£44,576

b. PFS Weibull

Using Weibull to extrapolate progression-free survival had little impact on incremental costs (an increase from £[redacted] in the base case to £[redacted]) for NIVO+CHEMO vs XELOX. The incremental QALYs decreased slightly from [redacted] in the base case to [redacted]. The ICER increased compared to the base case, from £45,383 per QALY to £48,878 per QALY.

Table 20. Weibull distribution for PFS

Technology	Total costs (£)	Total life years	Total QALYs	Inc. costs (£)	Inc. life years	Inc. QALYs	ICER (£/QALY)
Nivolumab + XELOX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	-
XELOX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£48,878

c. PFS Gompertz

Using Gompertz to extrapolate progression-free survival led to a decrease in incremental costs (£[redacted] in the base case to £[redacted]) for NIVO+CHEMO vs XELOX and an increase in incremental QALYs from [redacted] in the base case to [redacted]. The ICER decreased compared to the base case, from £45,383 per QALY to £36,629 per QALY.

Table 21. Gompertz distribution for PFS

Technology	Total costs (£)	Total life years	Total QALYs	Inc. costs (£)	Inc. life years	Inc. QALYs	ICER (£/QALY)
Nivolumab + XELOX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	-
XELOX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£36,629

d. PFS Exponential

Using Exponential to extrapolate progression-free survival had little impact on incremental costs (an increase from £[redacted] in the base case to £[redacted]) for NIVO+CHEMO vs XELOX. The incremental QALYs decreased slightly from [redacted] in the base case to [redacted]. The ICER increased compared to the base case, from £45,383 per QALY to £50,125 per QALY.

Table 22. Exponential distribution for PFS

Technology	Total costs (£)	Total life years	Total QALYs	Inc. costs (£)	Inc. life years	Inc. QALYs	ICER (£/QALY)
Nivolumab + XELOX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	-
XELOX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£50,125

e. PFS Generalised gamma

Using generalised gamma to extrapolate progression-free survival had little impact on incremental costs (a decrease from £[redacted] in the base case to £[redacted]) for NIVO+CHEMO vs XELOX. The incremental QALYs also decreased from [redacted] in the base case to [redacted]. The ICER increased compared to the base case, from £43,398 per QALY to £45,876 per QALY.

Table 23. Generalised gamma distribution for PFS

Technology	Total costs (£)	Total life years	Total QALYs	Inc. costs (£)	Inc. life years	Inc. QALYs	ICER (£/QALY)
Nivolumab + XELOX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	-
XELOX	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£45,876

7 References

1. National Institute for Health and Care Excellence. Appraisal Consultation Document - Nivolumab with platinum- and fluoropyrimidine-based chemotherapy for untreated HER2-negative advanced gastric, gastro-oesophageal junction or oesophageal adenocarcinoma. 2021.
2. Janjigian YY, Ajani JA, Moehler M, et al. LBA7 Nivolumab (NIVO) plus chemotherapy (Chemo) or ipilimumab (IPI) vs chemo as first-line (1L) treatment for advanced gastric cancer/gastroesophageal junction cancer/esophageal adenocarcinoma (GC/GEJC/EAC): CheckMate 649 study. Annals of Oncology. 2021;32:S1329-S30.
3. Bristol-Myers Squibb. Summary of Product Characteristics. OPDIVO 10 mg/mL concentrate for solution for infusion. . 2021. Available at: https://www.medicines.org.uk/emc/product/6888 [Accessed September 2021].
4. Chen L-T, Kang Y-K, Satoh T, et al. A phase III study of nivolumab (Nivo) in previously treated advanced gastric or gastric esophageal junction (G/GEJ) cancer (ATTRACTION-2): Three-year update data. Journal of Clinical Oncology. 2020;38(4_suppl):383-.
5. Chen L-T, Satoh T, Ryu M-H, et al. A phase 3 study of nivolumab in previously treated advanced gastric or gastroesophageal junction cancer (ATTRACTION-2): 2-year update data. Gastric Cancer. 2020;23(3):510-9.
6. Davidson M, Cafferkey C, Goode EF, et al. Survival in Advanced Esophagogastric Adenocarcinoma Improves With Use of Multiple Lines of Therapy: Results From an Analysis of More Than 500 Patients. Clin Colorectal Cancer. 2018;17(3):223-30.
7. Ford HER, Marshall A, Bridgewater JA, 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.
8. Chau I, Norman AR, Cunningham D, et al. The impact of primary tumour origins in patients with advanced oesophageal, oesophago-gastric junction and gastric adenocarcinoma--individual patient data from 1775 patients in four randomised controlled trials. Ann Oncol. 2009;20(5):885-91.
9. Chen T-T. Statistical issues and challenges in immuno-oncology. Journal for ImmunoTherapy of Cancer. 2013;1(1):18.
10. Rutherford MJ, Lambert PC, Sweeting MJ, et al. NICE DSU TECHNICAL SUPPORT DOCUMENT 21: Flexible Methods for Survival Analysis. 2020. Available at: http://nicedsu.org.uk/wp-content/uploads/2020/11/NICE-DSU-Flex-Surv-TSD-21_Final_alt_text.pdf [Accessed
11. Shankaran V, Xiao H, Bertwistle D, et al. A Comparison of Real-World Treatment Patterns and Clinical Outcomes in Patients Receiving First-Line Therapy for Unresectable Advanced Gastric or Gastroesophageal Junction Cancer Versus Esophageal Adenocarcinomas. Adv Ther. 2021;38(1):707-20.
12. Janjigian YY, Shitara K, Moehler M, et al. First-line nivolumab plus chemotherapy versus chemotherapy alone for advanced gastric, gastro-oesophageal junction, and oesophageal adenocarcinoma (CheckMate 649): a randomised, open-label, phase 3 trial. Lancet. 2021;398(10294):27-40.
13. Flatiron. US Flatiron Database: Refresh Number r2108 - data cut-off date: 7/31/2021.

14. Bristol-Myers Squibb. CheckMate 649 Supplementary data on CPS ≥5 subpopulation. 2021.
15. National Institute for Health and Care Excellence. Flexible Methods for Survival Analysis: TSD 21. 2020. Available at: http://nicedsu.org.uk/flexible-methods-for-survival-analysis-tsd/ [Accessed
16. 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. Available at: http://www.nicedsu.org.uk/NICE%20DSU%20TSD%20Survival%20analysis.updated%20March%202013.v2.pdf [Accessed 7 June 2016].
17. British National Formulary. Nivolumab solution for infusion. 2021. Available at: https://bnf.nice.org.uk/medicinal-forms/nivolumab.html [Accessed 31 January 2021].

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal

Nivolumab in combination with chemotherapy for untreated advanced gastric or gastroesophageal junction cancer

Appendix A: Utility Analysis Study Report

File name	Version	Contains confidential information	Date
		Yes	

Nivolumab in combination with chemotherapy for untreated advanced gastric or gastroesophageal junction cancer

Utility Analysis Report

October 2021

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BICR	Blinded Independent Central Review
CE	cost-effectiveness
EQ-5D	EuroQoL 5-Dimensions
FU	follow-up
HRQoL	health related quality of life
HSUV	health state utility value
HTA	health technology assessment
IPD	individual patient level data
MAR	missing at random
MCAR	missing completely at random
OS	overall survival
PD	progressed disease
PFS	progression-free survival
PRO	patient-reported outcome
QoL	quality of life
SD	stable disease
sd	standard deviation
se	standard error
SFU	survival follow-up
TTD	time to discontinuation
TTO	time trade-off
VAS	visual analogue scale

1. Introduction

Gastric cancer (GC) is a significant health problem in the UK. It accounted for 2% of all new cancer cases in 2017[1] and 6,600 new cases are reported every year (2015-2017).[2] Around half of new cases (51%) are in people aged 75 and over.[1] In the UK, at least 46-51% of people with GC are not diagnosed until their cancer reaches Stage III or IV (and 27-38% of patients with stage unknown).[3] Prognosis for these patients is poor, with few effective treatment options available.[1,4] All-stage five-year survival rates for GC are extremely poor compared with other cancer types, with one-year net survival ranging from 88.5% at Stage I to 21.4% at Stage IV. For patients with advanced disease chemotherapy or radiotherapy can improve symptoms and may improve survival,[5,6] but the aim of treatment is primarily palliative.[7]

Nivolumab is an immunotherapy, which stimulates the patient's immune system to target cancer cells through pre-existing, intrinsic processes, and has demonstrated significant improvements in survival compared to chemotherapy alone in patients with untreated, advanced GC.[8] Due to the novel mechanism of action of nivolumab, compared to conventional chemotherapy regimens, the quality of life (QoL) of patients treated with nivolumab in combination with chemotherapy (NIVO+CHEMO) may be different compared with patients receiving only chemotherapy (CHEMO). Thus, there is a need to assess nivolumab-specific utility values for the current indication: previously untreated patients with advanced gastric or gastro-oesophageal.

The pivotal study informing the clinical efficacy of NIVO+CHEMO in this indication is Checkmate 649, an open-label, multi-centre, XELOX/FOLFOX controlled Phase III trial (n=792 for chemotherapy arm, n=789 for nivolumab in combination with chemotherapy arm).

This analysis used longitudinal EQ-5D data from CheckMate 649 to estimate health state utility values (HSUVs) for use in a cost-utility model for the treatment of patients with previously untreated advanced gastric cancer. This state-transition model consists of the health states:

- Pre-progression
- Post-progression/progressed disease (PD)
- Death

In addition, an optional health state of "long-term remission" can be used to represent patients without progressed disease who have achieved ongoing relief from hazard of progression and excess mortality, and may be subject to alternative HSUVs than patients who remain at risk.

In addition to the baseline HSUVs, time to death was also considered as a potential modifier to patient utility, particularly in the post-progression health state, where the influence of heterogenous time to death was expected to be particularly acute. Treatment benefit, in the form of extended survival, was considered as justification for treatment-specific differences in HSUV due to the difference in impact of time to death over the surviving cohort.

2. Objectives

The primary objective of this analysis was to

- Develop HSUVs for a cost-utility model of treatment of previously untreated advanced gastric or gastro-oesophageal junction cancer.

A secondary objective of this analysis was to:

- Develop models of HSUV modification by time to death for use in the same cost-utility model.

3. Methodology

3.1. Data source

Individual patient-level data (IPD) from the CheckMate 649 trial were available, providing periodic patient self-assessment of utility via the EQ-5D-3L instrument. Patient baseline demographic and clinical characteristics were collected, and patient time to clinical progression and death were recorded if these events occurred within the follow-up period.

The schedule of assessment for patient-reported outcomes (PROs) was regular within CheckMate 649, but dependent upon treatment status, and could be summarised as follows:

- On treatment, assessment every 6 weeks (±3 days) from cycle 1 day 1, prior to any other study procedures.
- Follow-up visit 1 (FU1) at 30 days (±7 days) from last dose
- Follow-up visit 2 (FU2) at 84 days (±7 days) from FU1
- Survival follow-up visits (SFU[x]) every 3 months (±14 days) after FU2, by phone contact or in-person.

At each of these assessments, the EQ-5D-3L instrument was administered.

The tumour imaging assessment schedule within the protocol for CheckMate 649 was not dependent upon treatment status, and was (irrespective of follow-up visits):

- Every 6 weeks (±7 days) up to and including Week 48, then
- Every 12 weeks (±7 days)

Until PD, lost to follow-up, or withdrawal of consent.

As patients were permitted to remain on treatment after progression, and were permitted to discontinue treatment prior to progression, EQ-5D-3L data collected at both the 6 and 12 week timepoints are present in observation sets limited to the pre-progression period, and data sets limited to the progressed disease period. Within both of these data sets, observation frequency gave observations on treatment a disproportionate influence. This was mitigated by the relatively low rate of post-progression treatment and pre-progression discontinuation.

3.2. Statistical methods

3.2.1. Extraction of data

Patient-level data from Checkmate 649 was provided in a standardised Analysis Data Model (ADaM) format with the associated data dictionary. This data includes descriptions on study population demographics, patient baseline characteristics, time-to-event data for progression-free survival (PFS) and overall survival (OS), time to treatment discontinuation (TTD) and data per dimension on each questionnaire filled out by the patients for EQ-5D-3L assessment.

Each complete EQ-5D-3L record consisted of a score on each of the 5 dimensions, in addition to a visual analogue score (VAS) recorded while completing the EQ-5D-3L instrument. Completion of each of the 5 dimensions was required for inclusion in the used analytical dataset, but the VAS could be missing.

For each questionnaire, the treatment and progression status of the patient (per investigator and per BICR) was determined.

- Questionnaires completed prior to randomisation were marked as "baseline"
- For patients who received no study treatment, all records were marked as "off treatment"
- If the patient did not discontinue treatment, all records were marked as "on treatment"
- If the questionnaire was completed on a day prior to or equal to the date of discontinuation, the record was marked as "on treatment"
- Otherwise the record was marked "off treatment"

For progression status:

- If the patient did not have a PFS event within study follow-up, all records were marked as "pre-progression"
- If the questionnaire was completed on a day prior to but not equal to the date of progression, the record was marked as "pre-progression"
- If the questionnaire was completed on a day equal to or after the date of progression, the record was marked as "post-progression"

It was assumed that utility due to being on treatment would be maintained for the duration of treatment, until discontinuation, justifying the use of the final record. By contrast, disutility due to progression was assumed to be due to the health state of the patient, and not due to the assessment of the clinician, and so would be reflective of the state in which the patient was assessed to be at the time.

For records where the date of patient death was observed, the difference in time between the record and death was recorded; if this date was not available, the difference in time between the record and the OS censor was recorded as a minimum time from death.

3.2.2. Pre-processing of HRQoL data

The ED-5D-3L instrument has two components. The first is the assessment of five separate health state dimensions by asking the patient to agree with one of three statements about that dimension (

Table 1 shows the English translations of these statements). The second is the single-dimension assessment of general wellbeing on a scale of 0-100, the Visual Analogue Scale (VAS) score. Whilst neither assessment is entirely independent of the preferences of the population being assessed, the VAS is considered unanchored to the preferences of the general population and so is not intended to generalise, particularly between populations.[9] It is used to support the assessments of the 5D instrument, particularly with respect to perception of changing HRQoL.

Table 1. EQ-5D-3L health state dimension statements

Dimension	Scoring
Mobility	1 = I have no problems in walking about 2 = I have some problems in walking about 3 = I am confined to bed
Self-care	1 = I have no problems with self-care 2 = I have some problems washing or dressing myself 3 = I am unable to wash or dress myself
Usual activities	1 = I have no problems with performing my usual activities 2 = I have some problems with performing my usual activities 3 = I am unable to perform my usual activities
Pain / discomfort	1 = I have no pain or discomfort 2 = I have moderate pain or discomfort 3 = I have extreme pain or discomfort
Anxiety / depression	1 = I am not anxious or depressed 2 = I am moderately anxious or depressed 3 = I am extremely anxious or depressed

For each completed questionnaire, an index was derived from the five dimensions of the EQ-5D-3L health profile data to provide a single preference score of the self-assessed health states reported that would be representative of the preference of a patient in the UK population reporting the same health states. This index was as determined by Dolan et al (1997)[10] using the time-trade-off method. The resultant disutilities derived from the formula of Dolan et al[10] were subtracted from a baseline utility of 1 to give an index in the range of -0.595 to 1, where 1 represents perfect HRQoL, 0 represents no preference for further survival, and values less than 0 represent a negative preference for further survival.

For questionnaires where any dimension scored 2 or greater, the disutility value was greater than 0 and was determined by the following formula:

$$disutility = \alpha + \beta_1 M_O + \beta_2 S_C + \beta_3 U_A + \beta_4 P_D + \beta_5 A_D + \beta_6 M_2 + \beta_7 S_2 + \beta_8 U_2 + \beta_9 P_2 + \beta_{10} A_2 + \beta_{11} N_3$$

Using the coefficients from the Dolan[10] TTO disutility index, as presented in Table 2.

Table 2: Coefficients of Dolan[10] TTO disutility index

Variable Name	Variable	Associated Coefficient	Coefficient Value
-	-	α	0.081
M_O	1 if Mobility = 2; 2 if Mobility = 3	β_1	0.069
S_c	1 if Self-care = 2; 2 if Self-care = 3	β_2	0.104
U_A	1 if Usual Activities = 2; 2 if Usual Activities = 3	β_3	0.036
P_D	1 if Pain/Discomfort = 2; 2 if Pain/Discomfort = 3	β_4	0.123
A_D	1 if Anxiety/Depression = 2; 2 if Anxiety/Depression = 3	β_5	0.071
M_2	1 if Mobility = 3	β_6	0.176
S_2	1 if Self-care = 3	β_7	0.006
U_2	1 if Usual Activities = 3	β_8	0.022
P_2	1 if Pain/Discomfort = 3	β_9	0.140
A_2	1 if Anxiety/Depression = 3	β_{10}	0.094
N_3	1 if any score = 3	β_{11}	0.269

3.2.3. Study population descriptive summaries

Patients included in the study population were summarised in terms of the frequency and timing of their EQ-5D-3L responses. Baseline characteristics of included patients were descriptively summarised. Utility was summarised by treatment status and progression status.

3.3. Data exploration and 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 are missing completely at random (MCAR), as the remaining observations describe a reduced but unbiased sample of the overall distribution of observations.

3.4. Models of utility dependent upon progression status and TTD

Models of disutility dependent upon TTD were required in order to explain differences in HSUVs between treatments due to therapeutic effects. These were required for use in the cohort-level, semi-Markov, state transition model (STM).

The semi-Markov nature of the STM meant that it was not possible to track the proximity to the death state of patients in the pre-progression state who would "tunnel" through the post-progression state. Nevertheless, proximity to death was expected to affect the utility of patients in the pre-progression state, and this consideration determined the model form.

A linear mixed model was formed of utility independently dependent upon progression status and a transformation of TTD. A random offset was applied to each patient:

$$\boldsymbol{U} = \alpha \boldsymbol{p} + \boldsymbol{\beta} f(\boldsymbol{t}) + \gamma + \boldsymbol{\sigma}$$

Where

- $\boldsymbol{U}$ = utility value
- $\boldsymbol{p}$ = progression
- $f(\boldsymbol{t})$ = transformed time to death
- $\boldsymbol{\sigma}$ = random offset (per patient)
- α, $\boldsymbol{\beta}$, γ = coefficients

Within the state transition model, the mean baseline value γ is used to inform the HSUV for pre-progression. The progression term α is added to form the HSUV for post-progression. For patients transitioning to the death state in a given cycle, their disutility due to proximity to death accrued from model initiation is taken from the accrued QALYs at this cycle:

$$\Delta QALY = \int_0^{\tau} \boldsymbol{\beta} f(t) dt$$

Where τ is the time from model initiation.

3.5. Summary Values

The distribution of utilities observed at each observation, 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).[11,12]

Disutility due to adverse events was assumed to have been captured within the regular assessments. The on-treatment HSUV did not describe disutility due to events that resulted in treatment discontinuation; this disutility is assumed to be represented in the off-treatment HSUV.

4. Results

4.1. Study population descriptive summaries

Assessments of EQ-5D-3L status in CheckMate 649 were carried out every 6 weeks during the treatment phase, followed by two follow-up visits during the off-treatment phase approximately 5 and 17 weeks after last treatment. Thereafter, patients entered extended follow-up with vists every 3 months (~12 weeks).

There were 1,581 patients among the ITT population, randomised 789 to the NIVO+CHEMO arm and 792 to the CHEMO arm. As of the July 2021 data cut-off, [redacted] and [redacted] patients, respectively, had died from each arm within the trial period. Per investigator assessment, [redacted] and [redacted] patients had progressed or died without receiving subsequent therapy, whilst [redacted] and [redacted] were assessed as having progressed or died without receiving subsequent therapy per the BICR.

A total of [redacted] EQ-5D-3L completed questionnaires formed the "completed" dataset; a small number of these had missing associated VAS scores. These were predominantly in the "on treatment" period, but [redacted] were during the follow-up "off treatment" period. See Table 3 for a summary of the dataset.

Data availability for the "on-treatment" dataset for the NIVO+CHEMO arm is shown in Table 4 and displayed graphically in Figure 1. Completion rates were high throughout, with mostly >80% completion from baseline until week 145. Depending upon the assessor, up to 20% of observations at any week up to 151 were in patients with known progressed disease; this proportion rose over the first year on treatment, then remained roughly constant. The missingness pattern showed a mixture of monotonic and non-monotonic (sporadic) patterns (Figure 5). There was not a strong pattern of missingness immediately prior to discontinuation from treatment. Based on the degree and patterns of missingness, this dataset was considered suitable for evaluation as under complete-case analysis, under which missing data is assumed to be MCAR.

Data availability for the "off-treatment" dataset for NIVO+CHEMO is shown in Table 5 and displayed graphically in Figure 2. Completion rates were much lower in this dataset than in the "on treatment" dataset, with only 64% completed at Follow-Up 1. The proportion of observations among patients with confirmed progressed disease was lower than expected in this dataset, possibly as a result of commencing subsequent therapies. However, the observations displayed a similar proportion of progressed disease to the overall survival cohort, and so the sample did not appear to be biased in relation to progression status. The missingness pattern displayed considerably more monotonicity than the on-treatment dataset, and there was a notable pattern of missingness in the observation window prior to death (Figure 6). Given these features, it was considered unlikely that the data could provide unbiased estimates of mean utility under the assumption that observations were missing completely at random. Under the assumption that missingness was dependent only upon proximity to death and that the data was an unbiased weighted sample of utility conditional upon time to death, the mean profile of utility conditional upon time to death could be determined and used to provide patient-level estimates for the missing data.

Data availability for the "on-treatment" dataset for the CHEMO arm is shown in Table 6 and displayed graphically in Figure 3. The degree of missingness was comparable to the NIVO+CHEMO on-treatment dataset, with greater than 80% completion rate at all visits. The proportion of observations

in progressed disease was lower than in the NIVO+CHEMO arm, which is likely a reflection of the perceived benefit of nivolumab in patients apparently progressed that allowed for continued treatment post-progression. The missingness pattern (Figure 7) was very similar to that of the NIVO+CHEMO on-treatment dataset, and the same conclusion is drawn: that complete-case analysis under the assumption of MCAR is appropriate.

Data availability for the "off-treatment" dataset for the CHEMO arm is shown in Table 7 and displayed graphically in Figure 4. The rates of missingness were comparable to the off-treatment dataset for NIVO+CHEMO, but were slightly higher in the extended survival follow-up. The proportion of patients with progressed disease surviving diverged from the proportion of observations in patients with progressed disease in later follow-up, with a greater proportion of observations from patients with progressed disease. This may have been due to loss to EQ-5D follow-up of patients who commenced subsequent therapy prior to progression. Observing the missingness pattern in Figure 8, there were fewer long-term monotonic patterns than in the NIVO+CHEMO arm, but a large proportion of observations prior to death were missing. The same conclusions hold as for the NIVO+CHEMO arm: that the data were unlikely to give unbiased estimates of mean utility under the assumption of MCAR.

Table 3 : Dataset summary

	Treatment arm	
	NIVO+CHEMO	**CHEMO**
Intention to Treat population	789	792
Analysis population*	[redacted]	[redacted]
Total questionnaire entries	[redacted]	[redacted]
"Completed" questionnaires	[redacted]	[redacted]
"Completed" questionnaires assigned to on/off treatment analysis windows	[redacted]	[redacted]
"Completed" questionnaires at baseline**	[redacted]	[redacted]
"Completed" questionnaires assigned to "on treatment" period	[redacted]	[redacted]
"Completed" questionnaires assigned to "off treatment" period	[redacted]	[redacted]
**50 patients did not complete any EQ-5D-3L questionnaires; **106 patients had no baseline measurement*		

Table 4: EQ-5D-3L data availability – on treatment period (NIVO+CHEMO arm)

NIVO+CHEMO	Week on treatment												
	Baseline	**7**	**13**	**19**	**25**	**31**	**37**	**43**	**49**	**55**	**61**	**67**	**73**
Patients eligible, N													
Complete observations, N (%)													
Observations PD, Inv. (%)													
Observations PD, BICR (%)													
	79	**85**	**91**	**97**	**103**	**109**	**115**	**121**	**127**	**133**	**139**	**145**	**151**
Patients eligible, N													
Complete observations, N (%)													
Observations PD, Inv. (%)													
Observations PD, BICR (%)													
	157	**163**	**169**	**175**	**181**	**187**							
Patients eligible, N													
Complete observations, N (%)													
Observations PD, Inv. (%)													
Observations PD, BICR (%)													

PD: Progressed disease; Inv.: per investigator; BICR: per blinded independent central review

Figure 1. EQ-5D-3L questionnaire completion rates in on-treatment period (for NIVO+CHEMO arm)

Table 5: EQ-5D-3L data availability – off treatment period (NIVO+CHEMO arm)

	NIVO+CHEMO													
	Follow -up 1	**Follow -up 2**	**SFU1**	**SFU2**	**SFU3**	**SFU4**	**SFU5**	**SFU6**	**SFU7**	**SFU8**	**SFU9**	**SFU10**	**SFU11**	**SFU12**
Patients eligible, N	[redacted]													
Complete observations, N (%)	[redacted]													
Observations PD, Inv. (%)	[redacted]													
Observations PD, BICR (%)	[redacted]													
Patients PD (Inv.)	[redacted]													
Patients PD (BICR)	[redacted]													
SFU: Survival follow-up; PD: Progressed disease; Inv.: per investigator; BICR: Blinded Independent Central Review.														

[redacted]

Figure 2. EQ-5D-3L questionnaire completion rates: off-treatment period (for NIVO+CHEMO arm)

Table 6: EQ-5D-3L data availability – on treatment period (CHEMO arm)

CHEMO	Week on treatment												
	Baseline	**7**	**13**	**19**	**25**	**31**	**37**	**43**	**49**	**55**	**61**	**67**	**73**
Patients eligible, N													
Complete observations, N (%)													
Observations PD, Inv. (%)													
Observations PD, BICR (%)													
	79	**85**	**91**	**97**	**103**	**109**	**115**	**121**	**127**	**133**	**139**	**145**	**151**
Patients eligible, N													
Complete observations, N (%)													
Observations PD, Inv. (%)													
Observations PD, BICR (%)													
	157	**163**	**169**	**175**	**181**	**187**							
Patients eligible, N													
Complete observations, N (%)													
Observations PD, Inv. (%)													
Observations PD, BICR (%)													
PD: Progressed disease; Inv.: per investigator; BICR: per blinded independent central review													

Figure 3. EQ-5D-3L questionnaire completion rates in on-treatment period (for CHEMO arm)

Table 7: EQ-5D-3L data availability – off treatment period (CHEMO arm)

CHEMO														
	Follow -up 1	**Follow -up 2**	**SFU1**	**SFU2**	**SFU3**	**SFU4**	**SFU5**	**SFU6**	**SFU7**	**SFU8**	**SFU9**	**SFU10**	**SFU11**	**SFU12**
Patients eligible, N	[redacted]													
Complete observations, N (%)	[redacted]													
Observations PD, Inv. (%)	[redacted]													
Observations PD, BICR (%)	[redacted]													
Patients PD (Inv.)	[redacted]													
Patients PD (BICR)	[redacted]													
SFU: Survival follow-up; PD: Progressed disease; Inv.: per investigator; BICR: Blinded Independent Central Review.														

[redacted]

Figure 4. EQ-5D-3L questionnaire completion rates: off-treatment period (for CHEMO arm)

Figure 5. Missingness pattern of EQ-5D-3L data; on-treatment period (NIVO+CHEMO arm)

White blocks indicate expected missing data. Light shading indicates data expected and present. Darker shades indicate data not expected.

Figure 6. Missingness pattern of EQ-5D-3L data: off-treatment period (NIVO+CHEMO arm)

FU: Follow-up. Missing measurements shown for the scheduled follow-up visits 1 and 2; missingness for survival follow-up not shown.

Figure 7. Missingness pattern of EQ-5D-3L data; on-treatment period (CHEMO arm)

White blocks indicate expected missing data. Light shading indicates data expected and present. Darker shades indicate data not expected.

Figure 8. Missingness pattern of EQ-5D-3L data: off-treatment period (chemotherapy arm)

FU: Follow-up. Missing measurements shown for the scheduled follow-up visits 1 and 2; missingness for survival follow-up not shown.

4.2. Complete case analysis

Data were analysed with removal of observations with missing outcomes. This analysis required that the data was MCAR in order to provide unbiased estimates. This assumption was considered applicable for the on-treatment dataset, but was more doubtful in the off-treatment dataset, as these data were correlated imperfectly with progression status.

Table 8 gives the statistics of the Dolan indexed utility scores at baseline. In both arms, both mean and median are far below the age and sex matched general population norms. When observing the dimensions of the EQ-5D-3L individually (Appendix A), it is evident that a majority of patients reported some problems with pain or discomfort at baseline ([redacted] NIVO+CHEMO, [redacted] CHEMO). A large number of patients also reported some difficulty in performing usual activities ([redacted] NIVO+CHEMO, [redacted] CHEMO).

After application of the Dolan time trade-off utility decrement formula to the observations at each follow-up visit, the results were summarised in Figure 9 and Table 7. In both arms, there is a gradual increase in the mean utility of patients remaining on treatment over time, and in both cases this exceeds the matched general population norm after 1 year. Both arms have similar utility values at each time. The per-dimensional analysis shows that the proportion of patients with mobility issues stays approximately constant over time, as does the proportion with problems with self-care. The proportion reporting problems with usual activities slowly decreases over time in both arms. The largest change is in patients reporting problems with pain/discomfort, which shows a reduction from baseline at week 7 from [redacted] to [redacted] on NIVO+CHEMO and [redacted] to [redacted] on CHEMO. Anxiety/depression also shows a large decrease in the proportion reporting any problems, by week 31 this decreased from [redacted] to [redacted] in the NIVO+CHEMO arm and from [redacted] to [redacted] in the CHEMO arm.

Complete-case analysis of the off-treatment dataset was also undertaken, with the caveat that the missingness patterns suggested that the absolute values may be biased. Nevertheless, qualitative

comparison between the two arms was made. Table 10 and Figure 10 demonstrate the statistics of the utility index over follow-up visits. For both arms, the mean utility among the surviving and observed increased over time from a lower than baseline value shortly after discontinuing treatment. In the case of the NIVO+CHEMO arm, the matched general population mean was again exceeded at survival follow-up visit 3, where [redacted] observations were made. This high utility among long-term survivors might be explained by a difference in prognosis off-treatment between the NIVO+CHEMO and CHEMO arms.

Pooled summaries of these datasets are shown in Table 11. The on-treatment dataset overall mean was close to the matched general population mean, and so it would not be of substantial benefit to model the time-profile of utility observed, as this would be clipped by the general population value for clinical plausibility. The off-treatment dataset demonstrates a substantial difference between NIVO+CHEMO and CHEMO. A model involving time to death may explain this as a result of the extended treatment benefit from NIVO+CHEMO.

Summaries of the datasets per progression status are given in Table 12 and Table 13. Pre-progression utility was slightly higher in the NIVO+CHEMO arm, and this difference increased post-progression. The pre-progression utilities were comparable to the general population utility, whilst the post-progression were below baseline, except in the off-treatment dataset.

Table 8. Baseline utility

Statistic	NIVO+CHEMO (N=789)	CHEMO (N=792)
Baseline responses (N)	[redacted]	[redacted]
Mean (SD)	[redacted]	[redacted]
Median	[redacted]	[redacted]
Min, Max	[redacted]	[redacted]
General population mean*	[redacted]	[redacted]

**mean utility after age and sex matching to English population norms per the Health Survey for England (2008) as reported in Szende, Janssen & Cabases (2014).*[13]
SD: standard deviation

[redacted]

Figure 9. Summary of Dolan time trade-off utility over on-treatment measurement period

*EQ5D: EuroQol 5-dimension (3 level); Dolan: Dolan (1997) time trade-off index. Heavy bars are median; diamonds are mean values; hinges are at first and third quartiles; whiskers are to most extreme value no further than 1.5*IQR from hinge. Outliers shown individually. Limited to visits with ≥ 10 observations per arm.*

Table 9. On-treatment: Dolan index utility per visit

	NIVO+CHEMO	CHEMO
Visit	**Mean (95% CI)**	
Baseline	[redacted]	[redacted]
Week 7	[redacted]	[redacted]
Week 13	[redacted]	[redacted]
Week 19	[redacted]	[redacted]
Week 25	[redacted]	[redacted]
Week 31	[redacted]	[redacted]
Week 37	[redacted]	[redacted]
Week 43	[redacted]	[redacted]
Week 49	[redacted]	[redacted]
Week 55	[redacted]	[redacted]
Week 61	[redacted]	[redacted]
Week 67	[redacted]	[redacted]
Week 73	[redacted]	[redacted]
Week 79	[redacted]	[redacted]
Week 85	[redacted]	[redacted]
Week 91	[redacted]	[redacted]
Week 97	[redacted]	[redacted]
Week 103	[redacted]	[redacted]
Week 109	[redacted]	[redacted]

CI: confidence interval; SD: standard deviation. Limited to visits with ≥ 10 observations. Confidence interval by Normal approximation, may exceed 1

[redacted]

Figure 10 Summary of Dolan time trade-off utility over off-treatment measurement period

*EQ5D: EuroQol 5-dimension (3 level); Dolan: Dolan (1997) time trade-off index. Heavy bars are median; diamonds are mean values; hinges are at first and third quartiles; whiskers are to most extreme value no further than 1.5*IQR from hinge. Outliers shown individually. Limited to visits with ≥ 10 observations per arm.*

Table 10. Off-treatment: Dolan index utility per visit

	NIVO+CHEMO	CHEMO
Visit	**Mean (95% CI)**	
Follow-up 1	[redacted]	[redacted]
Follow-up 2	[redacted]	[redacted]
Survival follow-up 1	[redacted]	[redacted]
Survival follow-up 2	[redacted]	[redacted]
Survival follow-up 3	[redacted]	[redacted]
Survival follow-up 4	[redacted]	[redacted]
Survival follow-up 5	[redacted]	[redacted]
Survival follow-up 6	[redacted]	[redacted]
Survival follow-up 7	[redacted]	[redacted]
Survival follow-up 8	[redacted]	[redacted]
Survival follow-up 9	[redacted]	[redacted]
Survival follow-up 10	[redacted]	[redacted]
Survival follow-up 11	[redacted]	[redacted]
Survival follow-up 12	[redacted]	[redacted]

CI: confidence interval; SD: standard deviation. Confidence interval by Normal approximation, may exceed 1

Table 11. Mean Dolan TTO utilities over various periods

Dataset	NIVO+CHEMO	CHEMO	Pooled
	Mean [SE] (95% CI)		
Baseline	[redacted]	[redacted]	[redacted]
Week 7	[redacted]	[redacted]	[redacted]
Baseline + on treatment	[redacted]	[redacted]	[redacted]
On-treatment (excluding baseline)	[redacted]	[redacted]	[redacted]
On-treatment week 13+	[redacted]	[redacted]	[redacted]
All off treatment	[redacted]	[redacted]	[redacted]
Off treatment (FU)	[redacted]	[redacted]	[redacted]
Off treatment (SFU)	[redacted]	[redacted]	[redacted]

CI: confidence interval; FU: follow-up; TTO: time trade-off; SE: standard error; SFU: survival follow-up. SE modified by Prais-Winsten correction for autocorrelation.

Table 12. Mean utility by progression status (per BICR) – simple means; complete case analysis

Progression status	NIVO+CHEMO		CHEMO		Pooled	
	No. of questionnaires (No. of patients)	Mean Utility (SE)	No. of questionnaires (No. of patients)	Mean Utility (SE)	No. of questionnaires (No. of patients)	Mean Utility (SE)
All pre-progression	[redacted]					
All pre-progression (excluding baseline)	[redacted]					
On treatment pre-progression (excluding baseline)	[redacted]					
Off treatment pre-progression	[redacted]					
All Post-progression	[redacted]					
On treatment post-progression	[redacted]					
Off treatment post-progression	[redacted]					

SE: standard error of mean, corrected by Prais-Winsten estimate of autocorrelation

Table 13. Mean utility by progression status (per investigator) – simple means; complete case analysis

Progression status	NIVO+CHEMO		CHEMO		Pooled	
	No. of questionnaires (No. of patients)	Mean Utility (SE)	No. of questionnaires (No. of patients)	Mean Utility (SE)	No. of questionnaires (No. of patients)	Mean Utility (SE)
All pre-progression	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
All pre-progression (excluding baseline)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
On treatment pre-progression (excluding baseline)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Off treatment pre-progression	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
All Post-progression	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
On treatment post-progression	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Off treatment post-progression	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

SE: standard error of mean, corrected by Prais-Winsten estimate of autocorrelation

4.3. Utility conditional upon time to death

To investigate appropriate model forms for modelling utility as conditional upon time to death, the complete utility dataset was reduced to those where there was a known time until death (i.e. patient death time was observed in the CheckMate 649 data). This utility data was then plotted against time to death, and generalised additive model smoothers were used to identify suitable functional forms.

Figure 11 shows one of these investigations. It was intended that the model should show no effect of proximity to death once conditioned upon progression past a point to be identified by data inspection, in part for clinical plausibility and in part to reduce the discounting error of posthumously applying the accrued disutility due to proximity to death within the economic model. Data inspection revealed a clear kneeing point at 6 months prior to death. Prior to this point, utility would be modelled as constant conditional only upon health state, whilst after this point an apparently quadratic disutility profile would be applied. Utility near death was higher in the NIVO+CHEMO progressed population than in the CHEMO progressed population, but as this was not reflected in the constant utility region and could not be represented within the required model form without arm-independent models, the conservative choice was made to model upon the pooled arms.

[redacted]

Figure 11. Smoothed relationship between utility and time until death, progression per investigator

The final model form chosen was thus:

$$\boldsymbol{U} = \alpha \boldsymbol{p} + \beta_1 \boldsymbol{t_s} + \beta_2 \boldsymbol{t_s^2} + \gamma + \boldsymbol{\sigma}$$

Where $\boldsymbol{t_s}$ was a spline upon time to death defined as:

$$t_s = \begin{cases} 0 & t > 182 \\ \dfrac{182 - t}{182} & t \leq 182 \end{cases}$$

Predictions from this model are shown in Figure 12. The model is not reflective of the crossover in utility at more than one year until death; as the vast majority of data post-progression was collected with less than one year until death. This artifact in the smoother was considered likely to be due to low sampling. The prediction of utility near death in the pre-progressed state was lower than the smoothing model, but given the relatively low number of deaths directly from the pre-progressed state and that divergence did not start until within the final ~ 13 weeks, this was deemed acceptable.

[redacted]

Figure 12. Quadratic spline model of utility conditional upon time until death, progression per investigator

Thin lines are smoothing generalised additive models; dashed lines are mean-values predictions of the linear mixed model.

Table 14. Fixed-effect coefficients linear mixed model of utility upon time to death - progression per investigator

	Estimate	Std. Error	df	t value	Pr(>|t|)
Intercept (γ)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
$\boldsymbol{t_s^2}$ (β_2)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
$\boldsymbol{t_s}$ (β_1)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Progressed (α)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 15. Fixed-effect coefficients linear mixed model of utility upon time to death - progression per BICR

	Estimate	Std. Error	df	t value	Pr(>|t|)
Intercept (γ)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
t_s^2 (β_2)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
t_s (β_1)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Progressed (α)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Notably, the baseline utility from these models (γ,

Table 14, Table 15) is markedly lower than that using simple means. This is in part because mixed models are patient-level rather than marginal, and so do not take into account the relative time spent contributing to the cohort-level mean, where higher-utility patients are expected to dwell in a given health state for longer and record more utility samples. This patient-level effect is partially accounted for by the time to death spline, but also missing from this analysis were the utility values of patients who were not observed to die.

These data could be included in this spline model to influence the intercept value where the date of OS censoring was greater than 182 days from observation. The coefficients for these models are given in Table 16 and Table 17 and show an intercept lower than the simple means for pre-progression, but higher than for the observed deaths dataset. As these data affect but do not directly inform the spline coefficients which determine differential utility according to prognosis within health state, they were recommended for use in scenario analysis.

Table 16. Fixed-effect coefficients linear mixed model of utility upon time to death - progression per investigator – including censored data

	Estimate	Std. Error	df	t value	Pr(>|t|)
Intercept (γ)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
t_s^2 (β_2)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
t_s (β_1)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Progressed (α)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 17. Fixed-effect coefficients linear mixed model of utility upon time to death - progression per BICR – including censored data

	Estimate	Std. Error	df	t value	Pr(>|t|)
Intercept (γ)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
t_s^2 (β_2)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
t_s (β_1)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Progressed (α)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

5. Conclusions

The analysis presented in this report examined the datasets made available through CheckMate 649.

Data availability was very high throughout the on-treatment period, and, within the context of trials for oncology therapies, relatively high through the off-treatment period. However, the data pattern through the off-treatment period revealed a high rate of drop-out prior to death and so was expected to be biased.

Complete case analysis revealed that both the on-treatment and pre-progression datasets were comparable to general population norms. Post-progression, mean utility dropped to below baseline values. Means were generally higher in the NIVO+CHEMO arm than the CHEMO arm, however, the analysis was not powered to test for this difference.

As CheckMate 649 was an open-label study, patient-reported outcomes may be affected, as participants may systematically self-assess differently to patients blinded to their treatment. However, this may not be the case in all scenarios, or to a significant degree. A recent meta-epidemiological study revealed no significant difference in patient-reported relative outcomes in blinded and unblinded studies.[11] Given that perceptions of efficacy of CHEMO+NIVO in this population are consistent with the outcomes of this trial, it must be considered that patients receiving CHEMO+NIVO in practice are likely to have a similar perception of time spent on therapy as patients within this trial, and that the results are generally comparable with the observational general-population dataset provided in Szende et al.[13]

Models of utility conditional upon progression status and time to death were fitted. Time to death was highly influential upon utility and formed a quadratically declining profile within the last 6 months prior to death.

Use of these models within the cost-utility model for previously untreated gastric and gastro-oesophageal junction cancer was recommended.

6. References

1. Cancer Research UK. Stomach cancer incidence statistics. 2020. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/stomach-cancer/incidence [accessed 5 October 2020].
2. Cancer Research UK. Stomach cancer statistics. 2020. Available at: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/stomach-cancer [Accessed 5 October 2020].
3. Cancer Research UK. Stomach cancer. 2019. Available from: https://www.cancerresearchuk.org/about-cancer/stomach-cancer [accessed 6 October 2020].
4. Lordick F, Allum W, Carneiro F, et al. Unmet needs and challenges in gastric cancer: the way forward. Cancer Treat Rev. 2014;40(6):692-700.
5. Nagini S. Carcinoma of the stomach: A review of epidemiology, pathogenesis, molecular genetics and chemoprevention. World J Gastrointest Oncol. 2012;4(7):156-69.
6. Zali H, Rezaei-Tavirani M, Azodi M. Gastric cancer: prevention, risk factors and treatment. Gastroenterol Hepatol Bed Bench. 2011;4(4):175-85.
7. Smyth E, Verheij M, Allum W, et al. Gastric cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of Oncology. 2016;27(suppl 5):v38-v49.
8. Bristol-Myers Squibb Co. Clinical Protocol CA209649. A Randomized, Multicenter, Open-Label, Phase 3 Study of Nivolumab plus Ipilimumab or Nivolumab in Combination with Oxaliplatin plus Fluoropyrimidine versus Oxaliplatin plus Fluoropyrimidine in Subjects with Previously Untreated Advanced or Metastatic Gastric or Gastroesophageal Junction Cancer. (CheckMate 649: CHECKpoint pathway and nivoluMab clinical Trial Evaluation 649). 2018.
9. Foundation ER. EQ-5D-3L User guide v6.0. 2018. Available from: https://euroqol.org/publications/user-guides/ [accessed 31 July 2019].
10. Dolan P. Modeling Valuations for EuroQol Health States. Medical Care. 1997;35(11):1095-108.
11. Mislevy RJ. Journal of Educational Statistics. 1991;16(2):150-5.
12. Hardin J. Prais-Winsten regression. Stata Technical Bulletin. 1996;5.
13. Szende A, Janssen B, Cabases J. Self-Reported Population Health: An International Perspective based on EQ-5D. Dordrecht (NL): Springer; 2014.

Appendix A

Per dimension analysis

On-treatment

Among responses, all dimensions demonstrate a maximum proportion reporting some disutility at baseline, with a good trend of improvement over the first 3 months of treatment for both the CHEMO and NIVO+CHEMO arm. Self-care showed the lowest rate of problems at baseline and consequently had the lowest level of improvement. Mobility problems affected [redacted] and [redacted] of patients at the baseline for the CHEMO and NIVO+CHEMO arms, respectively. Patients on either the CHEMO and NIVO+CHEMO arms then showed a slight improvement with [redacted] still reporting some problems with mobility in both arms at week 13. Between weeks 13 and 37, there was an increase in the number of people reporting problems with mobility, but from this point on, there is then rapid improvement.

The majority of patients in both arms reported problems with pain/discomfort at the baseline, although there was a great improvement by week 13. In the NIVO+CHEMO arm [redacted] reported problems with pain/discomfort at week 13 down from [redacted] at the baseline. Compared to the CHEMO arm, this is a greater improvement as [redacted] of patients in the CHEMO arm reported problems with pain/discomfort at week 13, down from [redacted] at baseline.

Table 18. Per-dimension: on-treatment, NIVO+CHEMO, mobility

NIVO+CHEMO		Week on treatment												
Mobility		**Baseline**	**7**	**13**	**19**	**25**	**31**	**37**	**43**	**49**	**55**	**61**	**67**	**73**
Score	**1**													
	2													
	3													
Total														
Number reporting some problems														
		79	**85**	**91**	**97**	**103**	**109**	**115**	**121**	**127**	**133**	**139**	**145**	**151**
Score	**1**													
	2													
	3													
Total														
Number reporting some problems														

		157	163	169	175	181	187
Score	1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number reporting some problems		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 19. Per-dimension: on-treatment, CHEMO, mobility

CHEMO		Week on treatment												
Mobility		Baseline	7	13	19	25	31	37	43	49	55	61	67	73
Score	1													
	2													
	3													
Total														
Number reporting some problems														
		79	85	91	97	103	109	115	121	127	133	139	145	151
Score	1													
	2													
	3													
Total														
Number reporting some problems														

		157	163	169	175	181	187
Score	1						
	2						
	3						
Total							
Number reporting some problems							

Table 20. Per-dimension: on-treatment, NIVO+CHEMO, self-care

NIVO+CHEMO		Week on treatment												
Self care		**Baseline**	**7**	**13**	**19**	**25**	**31**	**37**	**43**	**49**	**55**	**61**	**67**	**73**
Score	**1**													
	2													
	3													
Total														
Number reporting some problems														
		79	**85**	**91**	**97**	**103**	**109**	**115**	**121**	**127**	**133**	**139**	**145**	**151**
Score	**1**													
	2													
	3													
Total														
Number reporting some problems														

		157	163	169	175	181	187
Score	1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number reporting some problems		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 21. Per-dimension: on-treatment, CHEMO, self-care

CHEMO		Week on treatment												
Self care		Baseline	7	13	19	25	31	37	43	49	55	61	67	73
Score	1													
	2													
	3													
Total														
Number reporting some problems														
		79	85	91	97	103	109	115	121	127	133	139	145	151
Score	1													
	2													
	3													
Total														
Number reporting some problems														

		157	163	169	175	181	187
Score	1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number reporting some problems		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 22. Per-dimension: on-treatment, NIVO+CHEMO, usual activities

NIVO+CHEMO		Week on treatment												
Usual Activities		Baseline	7	13	19	25	31	37	43	49	55	61	67	73
Score	1	[redacted]												
	2	[redacted]												
	3	[redacted]												
Total		[redacted]												
Number reporting some problems		[redacted]												
		79	85	91	97	103	109	115	121	127	133	139	145	151
Score	1	[redacted]												
	2	[redacted]												
	3	[redacted]												
Total		[redacted]												
Number reporting some problems		[redacted]												

		157	163	169	175	181	187
Score	1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number reporting some problems		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 23. Per-dimension: on-treatment, CHEMO, usual activities

CHEMO		Week on treatment												
Usual Activities		**Baseline**	**7**	**13**	**19**	**25**	**31**	**37**	**43**	**49**	**55**	**61**	**67**	**73**
Score	**1**													
	2													
	3													
Total														
Number reporting some problems														
		79	**85**	**91**	**97**	**103**	**109**	**115**	**121**	**127**	**133**	**139**	**145**	**151**
Score	**1**													
	2													
	3													
Total														
Number reporting some problems														

		157	163	169	175	181	187
Score	1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number reporting some problems		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 24. Per-dimension: on-treatment, NIVO+CHEMO, pain/discomfort

NIVO+CHEMO		Week on treatment												
Pain/Discomfort		Baseline	7	13	19	25	31	37	43	49	55	61	67	73
Score	1													
	2													
	3													
Total														
Number reporting some problems														
		79	85	91	97	103	109	115	121	127	133	139	145	151
Score	1													
	2													
	3													
Total														
Number reporting some problems														

		157	163	169	175	181	187
Score	1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number reporting some problems		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 25. Per-dimension: on-treatment, CHEMO, pain/discomfort

CHEMO		Week on treatment												
Pain/Discomfort		Baseline	7	13	19	25	31	37	43	49	55	61	67	73
Score	1													
	2													
	3													
Total														
Number reporting some problems														
		79	85	91	97	103	109	115	121	127	133	139	145	151
Score	1													
	2													
	3													
Total														
Number reporting some problems														

		157	163	169	175	181	187
Score	1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number reporting some problems		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 26. Per-dimension: on-treatment, NIVO+CHEMO, anxiety/depression

NIVO+CHEMO		Week on treatment												
Anxiety/Depression		Baseline	7	13	19	25	31	37	43	49	55	61	67	73
Score	1	[redacted]												
	2	[redacted]												
	3	[redacted]												
Total		[redacted]												
Number reporting some problems		[redacted]												
		79	85	91	97	103	109	115	121	127	133	139	145	151
Score	1	[redacted]												
	2	[redacted]												
	3	[redacted]												
Total		[redacted]												
Number reporting some problems		[redacted]												

		157	163	169	175	181	187
Score	1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number reporting some problems		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 27. Per-dimension: on-treatment, CHEMO, anxiety/depression

CHEMO		Week on treatment												
Anxiety/Depression		Baseline	7	13	19	25	31	37	43	49	55	61	67	73
Score	1													
	2													
	3													
Total														
Number reporting some problems														
		79	85	91	97	103	109	115	121	127	133	139	145	151
Score	1													
	2													
	3													
Total														
Number reporting some problems														

		157	163	169	175	181	187
Score	1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Number reporting some problems		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Off-treatment

Table 28. Off-treatment, per dimension, Mobility (CHEMO)

CHEMO		Off treatment visit												
Mobility		FU1	FU2	SFU1	SFU2	SFU3	SFU4	SFU5	SFU6	SFU7	SFU8	SFU9	SFU10	SFU11
Score	1													
	2													
	3													
Total														
Number reporting some problems														

		SFU12
Score	1	
	2	
	3	
Total		
Number reporting some problems		

Table 29. Off-treatment, per-dimension, Mobility (NIVO+CHEMO)

NIVO+CHEMO		Off treatment visit												
Mobility		FU1	FU2	SFU1	SFU2	SFU3	SFU4	SFU5	SFU6	SFU7	SFU8	SFU9	SFU10	SFU11
Score	1													
	2													
	3													
Total														
Number reporting some problems														

		SFU12
Score	1	
	2	
	3	
Total		
Number reporting some problems		

Table 30. Off-treatment, per dimension, self care (CHEMO)

CHEMO		Off treatment visit												
Self Care		FU1	FU2	SFU1	SFU2	SFU3	SFU4	SFU5	SFU6	SFU7	SFU8	SFU9	SFU10	SFU11
Score	1													
	2													
	3													
Total														
Number reporting some problems														

		SFU12
Score	1	
	2	
	3	
Total		
Number reporting some problems		

Table 31. Off-treatment, per-dimension, self care (CHEMO+NIVO)

NIVO+CHEMO		Off treatment visit												
Self Care		FU1	FU2	SFU1	SFU2	SFU3	SFU4	SFU5	SFU6	SFU7	SFU8	SFU9	SFU10	SFU11
Score	1													
	2													
	3													
Total														
Number reporting some problems														

		SFU12
Score	1	
	2	
	3	
Total		
Number reporting some problems		

Table 32. Off-treatment, per dimension, Usual Activities (CHEMO+NIVO)

CHEMO		Off treatment visit												
Usual Activities		FU1	FU2	SFU1	SFU2	SFU3	SFU4	SFU5	SFU6	SFU7	SFU8	SFU9	SFU10	SFU11
Score	1													
	2													
	3													
Total														
Number reporting some problems														

		SFU12
Score	1	
	2	
	3	
Total		
Number reporting some problems		

Table 33. Off-treatment, per dimension, Usual Activities (NIVO+CHEMO)

NIVO+CHEMO		Off treatment visit												
Usual Activities		FU1	FU2	SFU1	SFU2	SFU3	SFU4	SFU5	SFU6	SFU7	SFU8	SFU9	SFU10	SFU11
Score	1													
	2													
	3													
Total														
Number reporting some problems														

		SFU12
Score	1	
	2	
	3	
Total		
Number reporting some problems		

Table 34. Off-treatment, per dimension, pain/discomfort (CHEMO)

CHEMO		Off treatment visit												
Pain/Discomfort		FU1	FU2	SFU1	SFU2	SFU3	SFU4	SFU5	SFU6	SFU7	SFU8	SFU9	SFU10	SFU11
Score	1													
	2													
	3													
Total														
Number reporting some problems														

		SFU12
Score	1	
	2	
	3	
Total		
Number reporting some problems		

Table 35. Off-treatment, per dimension, Pain/Discomfort (NIVO+CHEMO)

NIVO+CHEMO		Off treatment visit												
Pain/Discomfort		FU1	FU2	SFU1	SFU2	SFU3	SFU4	SFU5	SFU6	SFU7	SFU8	SFU9	SFU10	SFU11
Score	1													
	2													
	3													
Total														
Number reporting some problems														

		SFU12
Score	1	
	2	
	3	
Total		
Number reporting some problems		

Table 36. Off-treatment, per dimension, anxiety/depression (CHEMO)

CHEMO		Off treatment visit												
Anxiety/Depression		FU1	FU2	SFU1	SFU2	SFU3	SFU4	SFU5	SFU6	SFU7	SFU8	SFU9	SFU10	SFU11
Score	1													
	2													
	3													
Total														
Number reporting some problems														

		SFU12
Score	1	
	2	
	3	
Total		
Number reporting some problems		

Table 37. Off-treatment, per dimension, anxiety/depression

NIVO+CHEMO		Off treatment visit												
Anxiety/Depression		FU1	FU2	SFU1	SFU2	SFU3	SFU4	SFU5	SFU6	SFU7	SFU8	SFU9	SFU10	SFU11
Score	1													
	2													
	3													
Total														
Number reporting some problems														

	SFU12
Score 1	
Score 2	
Score 3	
Total	
Number reporting some problems	

NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE

Single technology appraisal

Nivolumab in combination with chemotherapy for untreated advanced gastric or gastroesophageal junction cancer

Appendix B: Survival Analysis Study Report

File name	Version	Contains confidential information	Date
		Yes	

Nivolumab in combination with chemotherapy for untreated advanced gastric or gastroesophageal junction cancer

Survival Analysis Report

October 2021

Table of Contents

List of Tables

List of Figures

Abbreviations

AFT	accelerated failure time
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
BICR	blinded independent central review
CHEMO	fluoropyrimidine- and platinum-containing chemotherapy
CI	confidence interval
CPS	combined positive score
DBL	database lock
DSU	decision support unit
FOLFOX	oxaliplatin plus leucovorin and fluorouracil
GOJ	gastro-oesophageal junction
IO	immuno-oncology
IPI	ipilimumab
ITT	intention to treat
LTM	lifetable mortality
LTR	long term response
mCRC	metastatic colorectal cancer
MSI-H	microsatellite instability high
NIVO	nivolumab
OAC	oesophageal adenocarcinoma
ORR	objective response rate
OS	overall survival
PD-L1	programmed cell death ligand 1
PFS	progression-free survival
PH	proportional hazards
PLD	patient-level data
R-P	Royston-Parmar
TSD	technical support document
ToT	time on treatment
TTD	time to treatment discontinuation
XELOX	oxaliplatin plus capecitabine

Executive Summary

The objective of this analysis was to develop parametric extrapolations for blinded independent central reviewer (BICR)-assessed progression-free survival (PFS), investigator-assessed PFS, overall survival (OS), post-progression survival (PPS), and time to treatment discontinuation (TTD) following nivolumab therapy in combination with oxaliplatin plus fluoropyrimidine (NIVO+CHEMO) versus oxaliplatin plus fluoropyrimidine (CHEMO), based on available patient-level data (PLD). Evidence for the efficacy of NIVO+CHEMO versus CHEMO in the treatment of patients with previously untreated advanced or metastatic gastric/gastro-oesophageal junction (GOJ) cancer/oesophageal adenocarcinoma (OAC) is primarily derived from the CheckMate 649 trial. The primary outcome of CheckMate 649 was BICR-assessed PFS in patients with enriched expression of programmed cell death ligand 1 (PD-L1), defined as PD-L1 combined positive score (CPS) ≥ 5%. Other endpoints included OS and investigator-assessed PFS.

In CheckMate 649 (July 2021 database lock) 789 patients and 792 patients were included in the intention to treat (ITT) populations for NIVO+CHEMO and CHEMO, respectively. Models of time to event were formed based upon data from CheckMate 649 in order to inform state occupancy of a partitioned survival cost-utility model.

Data from the CheckMate 649 trial were analysed to produce models of survival outcomes to a lifetime horizon for use in cost-utility modelling. Semi-parametric piecewise models were chosen for PFS outcomes for both arms due to the high rate of change of hazard over the initial period after commencing treatment, followed by a more consistent period. Similarly, semi-parametric piecewise models were also chosen for the OS outcome for both arms. Fully parametric models of PPS were suitable for extrapolation, though may favour the CHEMO arm. The same distribution was recommended for both arms due to similarity of nonparametric estimates of the survival function, but scaling rules were rejected.

For investigator-assessed PFS, a semi-parametric model utilising Kaplan-Meier data to month 6.44 and a lognormal relative survival model thereafter was chosen for both arms. For OS, a semi-parametric model utilising Kaplan-Meier data to month 6.44 and a Gompertz relative survival model thereafter was selected for both arms. For PPS, a fully-parametric model utilising a log-logistic relative survival model was chosen for both arms. For TTD, direct use of the Kaplan-Meier estimate was chosen for both arms.

Given the clinical expectation that a small number of responsive patients may enter long-term remission after treatment with nivolumab, the long term survival benefit produced by this semi-parametric modelling approach may be conservative.

The ratio of PFS events due to death versus progression was described by a logistic model dependent upon time and log time. The PPS was calibrated to PFS and OS hazard ratios derived from an NMA.

1. Introduction

1.1. Context

CheckMate 649 (NCT02872116) is a Phase III, open-label, randomised trial which evaluates the efficacy of nivolumab (NIVO) plus ipilimumab (IPI) (NIVO+IPI), or NIVO in combination with oxaliplatin plus fluoropyrimidine (NIVO+CHEMO), versus oxaliplatin plus fluoropyrimidine (CHEMO), in patients with previously untreated advanced or metastatic gastric/gastro-oesophageal junction (GOJ) cancer/oesophageal adenocarcinoma (OAC).[1] CHEMO consists of investigator's choice oxaliplatin plus capecitabine (XELOX) or oxaliplatin plus leucovorin and fluorouracil (FOLFOX). The focus of this submission is on the cohort of patients that received combination of NIVO+CHEMO.

The primary outcomes of CheckMate 649 are overall survival (OS) in patients with enriched expression of programmed cell death ligand 1 (PD-L1), measured by combined positive score (CPS) ≥ 5, and progression-free survival (PFS) as assessed by blinded independent central review (BICR) in patients with PD-L1 CPS ≥ 5 (PFS population). Relevant secondary and exploratory objectives were OS in all randomised patients, PFS assessed by BICR in all randomised patients in the PFS population, and PFS as assessed by the investigator across PD-L1 CPS cut-offs.[1]

Clinical data to inform OS, PFS, post-progression survival (PPS) and distribution of time to treatment discontinuation (TTD) in an economic model were derived from CheckMate 649 July 2021 database lock (DBL).[2] The follow-up period for CheckMate 649 was considerably less than the lifetime time horizon required for an economic model, with a substantial proportion of patients remaining alive at trial clinical cut-off. Therefore, extrapolation of survival outcomes was used to inform over this time horizon.

1.2. Survival outcomes in immuno-oncology

Treatment of solid tumours with immuno-oncological therapies (IO therapies, immunotherapies) in advanced and metastatic settings has resulted in outcomes that are challenging to model with techniques historically used. In particular, the response characteristics of IO therapies differ from traditional systemic oncology treatments; notably, the delayed onset of treatment effects and the potential for long-term survival, including survival after progression.[3]

The exemplar for these unique response characteristics is ipilimumab treatment for unresectable or metastatic melanoma, the indication with longest follow-up in this field, where reports on survival for some patients having received ipilimumab treatment are approaching ten years,[4] and the rate of survival at year 3 has remained at a constant plateau until year 10. This durable survival profile is also present in patients receiving nivolumab for previously-treated non-small-cell lung cancer.[5] This study demonstrated survival benefit for patients treated with nivolumab over patients treated with docetaxel, and in landmark analysis based upon response at six months, high conditional survival rates were demonstrated from approximately month 24 post-landmark (30 months from treatment start) for patients with complete response or partial response, and also high conditional survival rates for patients with stable disease from month 30 (36 months from treatment start). Objective response is not considered a prerequisite for long term survival in patients receiving IO therapies,[6] and the observed phenomenon of long term survival after progression does not appear to be wholly attributable to subsequent therapies, and requires further investigation.[7]

In addition, the methods used for assessing progression in patients treated with IO therapies have been criticised. As well as the phenomenon of "pseudo-progression," whereby immune response to IO therapy results in an apparent growth in the target lesions prior to true response, the response evaluation criteria in solid tumours (RECIST), which measures the sum of the longest diameter of the target lesions, cannot represent "mixed" responses, i.e. response varying by lesion.[8]

As a result of these features, a number of assumptions that are conventional in modelling survival outcomes within the metastatic setting are challenged:

- Response may not be continuously variable between tumours and, by extension, through the population. Instead, there may be a mixture of primary responsive targets and primary resistive targets.[8] Thus, the population may be highly heterogenous in survival time distribution.
- There may be a delayed onset of treatment effects.[3] Thus, the survival/progression hazard of patients initially after treatment induction may not be indicative of the full distribution of event times after treatment effect onset.
- There may be extended survival post-clinical progression after withdrawing treatment for a subset of patients.[7] Thus, PFS/OS surrogacy may not hold validity.
- Both PFS and OS may "plateau" for an extended period, with no excess hazard of death over the general population (statistical cure).[4] Thus proper, unimodal survival time distributions may be unable to model the whole population.

These features of the response to IO therapy were considered when developing models to extrapolate the survival outcomes of CheckMate 649, as part of the model selection algorithm outlined by the NICE Decision Support Unit (DSU).[3] Guidance relating to the difficulties in modelling survival outcomes in immuno-oncology has been published recently by the DSU.[9] Consideration has been given to this guidance throughout, and models that are non-compliant with the recommendations of this guidance, e.g. models that incorporate no external data in extrapolation, were not considered appropriate.

1.3. Cost-effectiveness model

To provide evidence for the cost-effectiveness of the treatment of patients with NIVO+CHEMO, a *de novo* cost-utility economic model was developed. In brief the model has the following features:

- Discrete-time: 2 week timestep until lifetime horizon
- 3 or 4 disease states: pre-progressed at risk, progressed disease, dead, pre-progressed not at risk of progression [optional].
- Semi-Markov: rate of transition from pre-progressed at risk to progressed disease or death dependent upon time from model initiation; rate of transition from progressed disease to death dependent upon both time from progressed disease (excess hazard component) and time from model initiation (matched population baseline hazard component).
- Time on treatment (ToT) modelled independently and applied within the pre-progressed states.

In addition, a secondary model was developed in response to requests from the NICE evidence review group (ERG). This secondary model replaces the state-transition component of the Semi-Markov model with a three-state partitioned survival model defined by a progression-free survival and overall survival model.

2. Objectives

The objectives of this study were:

- To develop statistical models of PFS, OS, PPS and TTD for patients receiving NIVO+CHEMO or CHEMO alone for use in an economic model comparing the cost-effectiveness of these treatments.
- To assess the appropriateness of each extrapolation and select the most appropriate model, reflecting the approaches outlined by the NICE DSU and Bagust and Beale (2014).[10,11]

3. Methodology

In order to provide a robust and transparent assessment of parametric extrapolations for OS, PFS, and ToT, the methodologies suggested within the NICE DSU and Bagust and Beale (2014) were applied.[10,11]

3.1. Data Sources

3.1.1. NIVO+CHEMO and CHEMO populations

In CheckMate 649 (July 2020 DBL), 789 patients and 792 patients were included in the intention to treat (ITT) populations for NIVO+CHEMO and CHEMO, respectively. NIVO was given for up to 24 months in the absence of disease progression or unacceptable toxicity. Treatment beyond progression with nivolumab, alone or in combination with chemotherapy, was permitted.

The PD-L1 CPS ≥ 5 population was analysed, independently per PD-L1 CPS ≥ 5 with NIVO+CHEMO versus CHEMO alone. The use of the PD-L1 CPS ≥ 5 population for all outcomes required formation of an PD-L1 CPS ≥ 5 TTD outcome variable, whereby those patients who received no dose of study drug were modelled as discontinuing on day 1. Reported outcomes are shown in Table 1.

Table 1. Outcomes from CheckMate 649, PD-L1 CPS ≥ 5 population (July 2021 DBL)

Endpoint	NIVO+CHEMO (N=473)	CHEMO (N=482)
OS		
Events, N	[redacted]	[redacted]
Median (months)	[redacted]	[redacted]
95% CI	[redacted]	[redacted]
Investigator assessed PFS		
Events, N	[redacted]	[redacted]
Median (months)	[redacted]	[redacted]
95% CI	[redacted]	[redacted]
BICR assessed PFS		
Events, N	[redacted]	[redacted]
Median (months)	[redacted]	[redacted]
95% CI	[redacted]	[redacted]
TTD		
Events, N	[redacted]	[redacted]

Endpoint	NIVO+CHEMO (N=473)	CHEMO (N=482)
Median (months)	[redacted]	[redacted]
95% CI	[redacted]	[redacted]
BICR: blinded independent central review; CI: confidence interval; DBL: database lock; NIVO: nivolumab; OS: overall survival; PFS: progression-free survival; TTD: time to treatment discontinuation		

3.1.2. Other comparators

Sources of outcomes data for patients receiving comparator therapies were identified from a clinical SLR (described in Appendix D). Hazard ratios of PFS and OS versus XELOX/FOLFOX were obtained using a network meta-analysis (NMA) to derive transition rates by scaling the CHEMO arm outcomes data of CheckMate 649.

3.1.3. Lifetable data

Baseline hazards of mortality from lifetables were used in a relative survival framework. Within CheckMate 649, this hazard was evaluated on a per-patient basis, conditional upon nationality of the study clinic, sex, patient age, and, for US patients, race and ethnicity. The sources of lifetable data used for this analysis are given in Table 2.

Table 2. Sources of lifetable data

Clinic location	Lifetable conditions	Analysis year(s)	Source	No. patients in CheckMate 649
Argentina	Sex	2008–2010	INDEC: Instituto Nacional de Estadistica y Censos[12]	82
Australia	Sex	2016–2018	Australian Bureau of Statistics[13]	38
Brazil	Sex	2017	Instituto Brasileiro de Geografia e Estatística, IBGE†[14]	69
Canada	Sex	2016–2018	Statistics Canada[15]	60
Chile	Sex	2015–2017	Human Mortality Database[16]	125
China	Sex	2016	WHO Global Health Observatory data repository*[17]	208
Columbia	Sex	2010	Economic Commission for Latin America and the Caribbean (ECLAC)[18]	31
Czech Republic	Sex	2015–2018	Human Mortality Database[16]	15
France	Sex	2015–2018	Human Mortality Database[16]	54
Germany	Sex	2015–2017	Human Mortality Database[16]	77
Greece	Sex	2015–2017	Human Mortality Database[16]	36
Hong Kong	Sex	2015–2017	Human Mortality Database[16]	6
Hungary	Sex	2015–2017	Human Mortality Database[16]	18

Clinic location	Lifetable conditions	Analysis year(s)	Source	No. patients in CheckMate 649
Israel	Sex	2015–2016	Human Mortality Database[16]	31
Italy	Sex	2015–2017	Human Mortality Database[16]	40
Japan	Sex	2015–2018	Human Mortality Database[16]	109
Mexico	Sex	2010	Economic Commission for Latin America and the Caribbean (ECLAC)[18]	34
Peru	Sex	2010	Economic Commission for Latin America and the Caribbean (ECLAC)[18]	41
Poland	Sex	2015–2018	Human Mortality Database[16]	61
Portugal	Sex	2015–2018	Human Mortality Database[16]	19
Romania	Sex	2016	WHO Global Health Observatory data repository*[17]	62
Russian Federation	Sex	2010–2014	Human Mortality Database[16]	14
Singapore	Sex	2017–2018	Statistics Singapore[19]	19
Spain	Sex	2018	Instituto Nacional de Estadística[20]	42
South Korea	Sex	2015–2018	Human Mortality Database[16]	8
Turkey	Sex	2016–2018	Turkish Statistical Institute[21]	35
Taiwan	Sex	2010–2014	Human Mortality Database[16]	6
UK	Sex	2016–2018	Office for National Statistics[22]	38
USA	Sex Race = White/Black Ethnicity = Hispanic/non-Hispanic	2017	CDC National Center for Health Statistics: National Vital Statistics System[23]	203

* World Health Organization data expressed in 5-yearly intervals and capped at 85 years; mortality extending up to 100 years assumed to continue at the same hazard as the 80–85 years category

† Instituto Brasileiro de Geografia e Estatística data capped at 80 years; mortality extending up to 100 years assumed to continue at the same hazard

3.2. Outcome definitions

Outcomes in CheckMate 649 were defined as:

- PFS: The time between the date of randomisation and the earlier date on which either the overall response was assessed as progressed or the patient died of any cause. Patients who received subsequent therapy were censored at last tumour assessment on or prior to the

time of initiation upon subsequent therapy. Patients who were not observed as progressing or dying were censored at the last assessment prior to end of follow-up. Patients without any on-study tumour assessments and who did not die or died after initiation of subsequent anti-cancer therapy were censored at the randomisation date.

- OS: The time from randomisation until death from any cause. Patients who were not observed as dying were censored at the last date the patient was known to be alive.

In addition, TTD was defined as:

- TTD: The time from the date of treatment initiation until the date of permanent discontinuation from study treatment, if known. If not known, patients were censored at the time of the last dose of study therapy.

To inform the semi-Markov state transition model, a further state occupancy function was defined:

- Post-progression survival (PPS): The time between the date of progression and date of death for patients with a PFS event and without death on the same day. Patients without observed date of death were censored at the difference between last survival follow-up and progression date.

3.3. Methods of extrapolation

3.3.1. Overview of approach

In order to provide a robust and transparent assessment, the methodologies suggested within the NICE DSU and Bagust and Beale (2014) were applied.[10,11] The model selection algorithm was used to select a suitable model (Figure 1). An overview of the approach is detailed below:

- Characterise the available data from CheckMate 649
- Describe trends in the available data
- Assess viability of accelerated failure time and proportional hazards models
- Assess suitability of standard statistical models
- If standard statistical models are not indicated, consider (not mutually exclusive):
    - A relative survival framework
    - A piecewise framework
    - A flexible non-statistical framework (splines)
    - A mixture framework
- Assess appropriateness of parametric models of extrapolation on the basis of:
    - Goodness-of-fit statistics (Akaike Information Criterion [AIC]/ Bayesian Information Criterion [BIC])
    - Non-parametric or smoothed representations of PLD
    - Examination of log-cumulative hazard plots
    - Assessment of clinical validity
    - Consideration of external data (e.g. within-class in similar indications)
- Select most plausible models, and other valid models for sensitivity analysis.

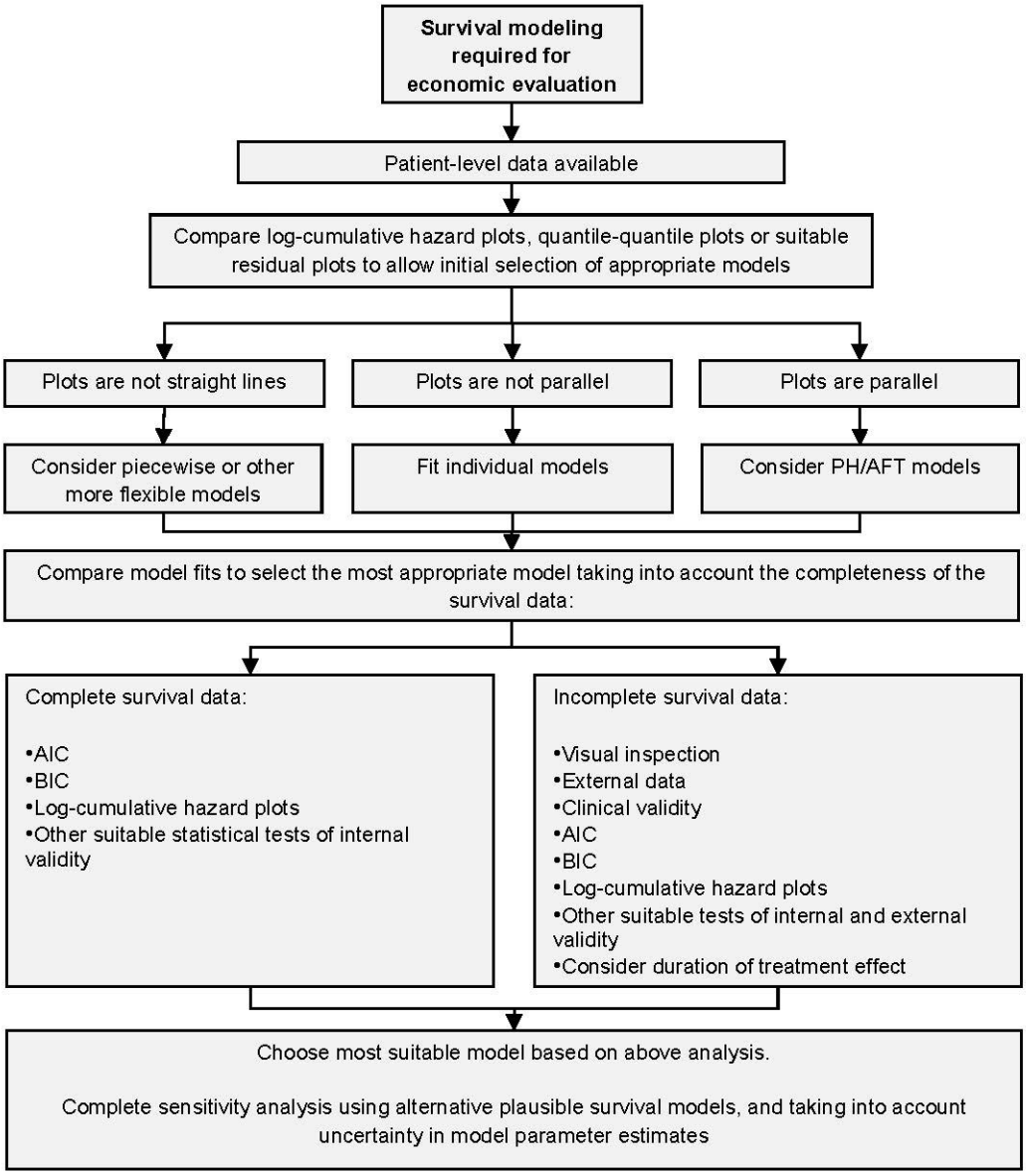

Figure 1. Survival model selection process algorithm

Source: NICE Decision Support Unit Document 14: Survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data.[11]
AFT: accelerated failure time; AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; PH: proportional hazards.

3.3.2. Description of patient-level data

The data were summarised by number of patients at risk, number of events, median (if reached), landmark survival at 6, 12, 24 and 36 months, and plots of Kaplan-Meier estimates of survival.

3.3.2.1. Matched general-population survival

A matched general-population survival curve was estimated using recent country, sex and age-specific lifetables. Based upon their exact age at randomisation, each patient was modelled as receiving piecewise-constant hazard of death to maximum age represented in their lifetable:

$$h_{LT_i}(t) = \ln\left(1 - q_{LT_i}(floor\left(t + age_{baseline_i}\right)\right)$$

Where $h_{LT_i}(t)$ is the instantaneous hazard of death per lifetable in units of 1/year for patient, i, $age_{baseline_i}$ is the age of the patient at randomisation, and $q_{LT_i}(x)$ is the annual probability of death from the lifetable stratum of patient i. The cumulative hazard due to lifetable:

$$H_{LT_i}(t) = \int_0^t h_{LT_i}(\tau)d\tau$$

was then converted to survival probability:

$$S_{LT_i}(t) = \exp\left(-H_{LT_i}(t)\right)$$

The mean survival probability over all patients within the original population was taken as the final matched general-population survival curve:

$$S_{LT}(t) = \frac{1}{n}\sum_{i=1}^{n} S_{LT_i}(t)$$

where n is the total number of patients at risk at t = 0. At the maximum extent of the country-specific lifetable (e.g. 100 years), a continuous, very high hazard was applied to cap survival; the impact of these capping hazards upon the marginal predicted hazard profile is visible in some plots.

3.3.3. Description of trends in available data

Nonparametric plots describing the scaled trends in the survival data were produced, per Ishak et al (2013).[24] Smoothed estimates of event hazard experienced over the follow-up were produced by three independent estimators:

- Kernel-smoothing of the cumulative hazard function using the "R" package *muhaz*.
- Formation of a flexible parametric Royston-Parmar spline model of cumulative hazard using the "R" package *flexsurv.*
- B-spline smoothing of the hazard function using the "R" package *bshazard.*

Default settings were used for *muhaz* and *bshazard*; for the Royston-Parmar spline models, a global optimum for knot position and value was sought for a range of numbers of internal knots. Subject to plausibility (e.g. positive hazard) the best fitting model was chosen by AIC.

In all cases, where applicable, the matched general population lifetable all-cause mortality function was plotted on the same axes as a reference.

3.3.4. Assessment of proportional hazards assumption

Complementary log-log (log cumulative hazard) plots of the CheckMate 649 trial arm data versus log time for OS and PFS were used to assess the appropriateness of the assumption of proportional hazards between outcomes on each therapy.

In each case, a constant vertical spacing of the logarithm of cumulative hazard was considered indicative of constantly proportional hazards between the two event time distributions, and a constant horizontal spacing of achieved cumulative hazard in log time was considered indicative of the event time distributions having an accelerated failure time relationship.

3.3.5. Standard statistical models

TSD 14[11] outlines six statistical survival time distribution models that should be considered prior to undertaking alternative survival modelling methods:

1. Exponential
2. Weibull
3. Gompertz
4. Log-logistic
5. Lognormal
6. Generalised gamma

Fitting these probability distributions to survival data for extrapolation assumes that all times-to-event within the population are drawn from the same, optionally conditional, distribution. For a cohort-level marginal model of survival times, the unexplained, natural variance of survival time and the variance due to heterogeneity of the modelled population are incorporated into a single distribution.

The use of statistical models for extrapolation draws validity from the theoretical basis of these probability distributions as sums and products of random variables, analogous within a large population to the cumulative effect of a large number of random processes driving events, and the repeated observation of these distributions within the natural world. It is assumed that the observed portion of the distribution (generally, the earlier event times) may inform parameters of the distribution upon which the unobserved later event time distribution may be inferred.

The nominated distributions can be described using a characteristic function – the hazard profile. This is the instantaneous risk of event per unit time, conditional upon surviving to the evaluated time. For cohort-level models, these profiles must be considered with relation to the heterogeneity of the cohort and the selection pressures present. Individuals with lower hazard at baseline are more likely to have later event times, so even in a uniformly progressive disease this selection pressure can result in a decreasing average hazard among the survivors, even as all survivors have monotonically increasing hazard of event.

The hazard profiles given by the nominated models can be grouped as follows:

- The average hazard over the whole cohort is constant (exponential model).
- The average hazard over the whole cohort increases or decreases proportional to a function of time (Weibull, Gompertz model [both degenerate to an exponential model when the coefficient of proportionality is 0]).
- The average hazard over the whole cohort increases to a peak, then decreases long term (log-logistic, lognormal model).

The generalised gamma model can describe any of these profiles, and can degenerate to exponential, Weibull or lognormal models, depending upon its parameter values. As it has a greater number of parameters to fit than any of the other nominated models, it is at the greatest risk of overfitting to the data – i.e. fitting to random variation rather than the underlying distribution.

As stated above, the nominated models fitted to all data may be appropriate when the variance of all event times as a result of both random variation and population heterogeneity can be explained within a single distribution. As well as implying that the distribution of risk amongst the population must vary smoothly, this is more easily satisfied when all events are driven by a similar process. For

composite events e.g. all-cause mortality, risks of one cause may vary disproportionately to risks of other causes with respect to time or population risk factors, resulting in heterogenous or multimodal event time distributions. Use of these models also implies that no abrupt changes in circumstances arise, i.e. treatment effects are maintained or changed smoothly with respect to time, and changes in risk factors with respect to time are smooth and consistent across the population.

Parametric survival functions were fitted to PLD using the R statistics environment, version 4.0.2 (2020-06-22), using the parametric survival fitting package *flexsurv* (version 1.1.1). The functional forms of the fitted models are described in Table 3.

Table 3. Functional forms of parametric survival equations

Distribution	Survival Function	Hazard Function
Exponential	$e^{-\lambda t}$	λ
Weibull	$e^{-\left(\frac{t}{\lambda}\right)^k}$	$\frac{k}{\lambda}\left(\frac{t}{\lambda}\right)^{k-1}$
Log-logistic	$\frac{1}{1+\left(\frac{t}{\alpha}\right)^\beta}$	$\frac{(\beta\ /\ \alpha)(t\ /\ \alpha)^{\alpha-1}}{1+(t\ /\ \alpha)^\beta}$
Lognormal	$\frac{1}{2}-\frac{1}{2}erf\left(\frac{\ln t-\mu}{\sqrt{2}\sigma}\right)$	$\frac{\frac{1}{\sigma t}erf\left(\frac{\ln t}{\sigma}\right)}{erf\left(\frac{-\ln t}{\sigma}\right)}$
Gompertz	$e^{\frac{\lambda}{\theta}(1-e^{\theta t})}$	$\lambda e^{\theta t}$
Generalised gamma function	$1-\Gamma_{(\lambda t)^\theta}(\rho)$ Where $\Gamma_{(\lambda t)}(\rho)=\frac{1}{\Gamma(\rho)}\int_0^{\lambda t}u^{\rho-1}e^{-u}\,du$ is the incomplete gamma function	$\frac{\theta\lambda^{\rho\theta}t^{\rho\theta-1}\exp\{-(\lambda t)^\theta\}}{\Gamma(\rho)}/S(t)$

t = time
e = Euler's number
erf = Error Function
$\Gamma(\ \ )$ = Gamma Function
$\ln(\ \ )$ = natural logarithm
Other parameters are distribution specific.

It is notable that all the nominated distributions are unimodal, i.e. they can only describe a single "concentration" of death times. This makes them unsuitable for describing strongly heterogenous populations or risk profiles. However, survival time may be multimodal, driven by:

- Process e.g. all the population are subject to a high hazard early in the process, but those who survive this initial period are subject to a low hazard until a substantial further period has passed.
- Population e.g. the population consists of patients who are either primary responsive or primary resistive. Those who are primary resistive experience high hazard of event at all times from process start, but the sub-population is rapidly exhausted and the marginal hazard declines to that of the primary responsive.

A combination of the two cases may also occur. In all cases, an accumulation of risk factors later in the life of the patients may cause a secondary concentration in event times, producing a

characteristic distribution of time to event known as a "bathtub" curve. None of the nominated models are capable of reproducing this type of distribution without modification.

Upon inspection of the data, it was clear that all plausible models fitted would exhibit decreasing marginal hazard after peak, i.e. the average hazard across the survivors at each time. As it is clinically unfeasible that the marginal hazard for a cohort would decrease indefinitely, and in accordance with the recommendation of TSD 21 to include external data in extrapolative models,[9] it was decided that all models must be fitted in a relative survival context, incorporating matched lifetable hazard of mortality as a baseline hazard above which the parametric model would inform an additive disease-specific excess hazard. Thus, the maximum life expectancy for any individual in the study was capped at matched general population lifetable life expectancy.

This relative survival model was described by:

$$S(t) = S_{LT}(t)S_E(t)$$

where $S(t)$ denoted survival probability at time t, S_{LT} is the background survival probability, and S_E the survival probability associated with the excess disease-related risk. In evaluating the likelihood function of these models, the individual lifetable survival functions were used. Letting t_j and δ_j represent the follow-up time and event indicator (δ_j=1 if deceased, δ_j=0 if censored), respectively, for individual $j, j = 1, \dots, N$, and letting λ_{LT_j} and λ_E represent the individual background and unconditional disease specific hazard functions respectively, the likelihood function was:

$$L = \prod_{j=1}^{N} \left[S_{LT_j}(t_j)S_E(t_j)\right]^{1-\delta_j} \left[\left(\lambda_{LT_j}(t_j) + \lambda_E(t_j)\right)\left(S_{LT_j}(t_j)S_E(t_j)\right)\right]^{\delta_j}$$

3.3.6. Alternative models

Per the methodological process given in TSD 14,[11] upon rejection of the nominated models, "piecewise modelling and other novel survival modelling methods such as those demonstrated by Royston and Parmar[25] and Jackson et al[26] should be considered."

The following model classes were considered potentially appropriate for the heterogeneity of survival times in the observed data, and were examined for appropriateness and fitted when indicated:

- Royston-Parmar spline models.
- Piecewise (Gelber) semi-parametric models.
- Mixture/mixture-cure models.

These models are described in the following subsections.

3.3.7. Royston-Parmar spline models

Unlike the nominated statistical distribution models, which use all available data to inform the observed portion of a partially-observed standard statistical distribution under the assumption that the unobserved portion is drawn from the same distribution for extrapolation, Royston-Parmar natural parametric spline fits use a number of pre-selected points to inform intersection of a curve. Typically, there would be one point at the first event time and with another point at the last event, and any number of intermediate points (knots) defining the functional progression between. Constraints are placed upon the derivatives of the function for smoothness, and boundary conditions are enforced for extrapolation, generally that the gradient of the function remains constant in extrapolation.

It is acknowledged that automated spline fitting algorithms do not currently identify the best position for knots, and whilst their relatively arbitrary hazard profiles would allow for good fitting over its support, caution should be exercised when interpreting such fits due to the lack of a rationale over the extrapolation zone. As these spline fits degenerate to their base forms beyond the final knot, their parameters are determined by the slope and offset of the spline as it intersects with this final knot, subject to the boundary condition of zero curvature. As such, spline fits may only approach a reasonable extrapolation of survival data if the spline near the final knot would, unconstrained, have a low rate of change of gradient, indicating that the parameters of the degenerate form are approaching constancy.

The profile of the survival functions in this assessment tended to have a high rate of hazard change nearing the extrapolation region, and so sensitivity to knot location in relation to this was high, without substantially affecting goodness-of-fit. As such, these models were not considered informative for extrapolation.

3.3.8. Semi-parametric models

The nominated unimodal models of survival time distribution assume some form of continuous progression of marginal hazards.

If a survival curve shows rapid rate of change of hazard, it could signify the total loss of a relatively high-risk sub-population. If this is the case, this sub-population does not require representation in the extrapolation region of the curve as they are fully represented by the trial data, and so extrapolation would be better represented by fitting to the (relatively low-risk) sub-population that remain after removal of the high-risk sub-population. Alternatively, it could represent the end of a period of unusual hazard; typically, study entrants are healthy enough not to be at immediate risk of mortality, and so there can be a period of lower hazard immediately after randomisation.

The method of Gelber et al (1993)[27] allowed for modelling these situations where a parametric fit that included the early data was not capable of or appropriate for modelling the complete survival curve; particularly for informing the extrapolation beyond the end of follow-up. The initial early data were represented by the Kaplan-Meier estimate of the study data, which does not impose a parametric form on the data. The subsequent later data was then assumed to consist of a more homogenous population with more continuous development of hazard that better represents the hazard trends beyond the end of follow-up. It was defined as follows:

$$S(t) = \begin{cases} S_{KM}(t) & \text{if } t \leq \tau \\ S_{KM}(\tau) S_p(t-\tau) & \text{if } t > \tau \end{cases}$$

Where $S_{KM}(t)$ is the Kaplan-Meier survival estimator and $S_p(t)$ is the parametric (relative) survival model of the tail data, conditional upon survival to cut point τ.

Selection of the cut point at which to begin parametric fitting of survival data without explicit definition of sub-populations is a matter of compromise between providing sufficient data for a robust estimation of the hazard progression in the extrapolation zone, and removal of the data that represents sub-populations not present in the extrapolation zone, or that is simply more constraining for the continuous hazard functions of the parametric forms to fit. Some cut times presented obvious discontinuation in the survival data, for example progression assessment times for PFS. Selection of cut times within these periods was avoided.

The semi-parametric approach in this analysis utilised available PLD to inform patient outcomes over the trial period, avoided over-parameterisation, and provided a clinical rationale for the underlying evolution of hazards over time, in line with NICE DSU guidelines.[11] A number of potential cut points were evaluated for each outcome, as there was no point at which the hazard profile clearly adopted the form that carried through to maximum follow-up. These ranged across the first year of CheckMate 649 during the period of 6-weekly assessments, whilst avoiding assessment windows of ±1 week of the assessments. For the base case, a time of 6.44 months was used. This represented a time mid-way through the first year, near the inflexion point of the hazard profile and after the period of most obvious stepping of the PFS curves.

In all cases, the Kaplan-Meier portion of the model was represented directly, whilst the parametric portion of the model was evaluated under a relative survival framework per the standard statistical models.

3.3.9. Mixture/mixture-cure models

Mixture models assume that the population consists of two or more sub-populations whose event times are drawn from independent distributions. A particular form of mixture model is the mixture-cure model, which assumes that one subpopulation is at no risk of a specific event. In the relative survival context, it is the excess hazard of event due to disease that is absent in the "cured" subpopulation. The mixture-cure model was defined as follows:

$$S(t) = S_{LT}(t)[p + (1-p)S_E(t)]$$

where $S(t)$ denotes survival probability at time t, S_{LT} is the background survival probability, S_E is the survival probability associated with the excess disease-related risk, and p denotes the cure fraction.

Parameters associated with S_E and p were obtained using maximum likelihood estimation. Letting t_j and δ_j represent the follow-up time and event indicator (δ_j=1 if deceased, δ_j=0 if censored), respectively, for individual $j, j = 1, \ldots, N$, and letting f_{LT} and f_E represent the background and disease specific density functions respectively, the likelihood function was:

$$L = \prod_{j=1}^{N} [pS_{LT}(t_j) + (1-p)S_E(t_j)S_{LT}(t_j)]^{1-\delta_j} [p\, f_{LT}(t_j) + (1-p)(S_E(t_j)f_{LT}(t_j) + S_{LT}(t_j)f_E(t_j))]^{\delta_j}$$

All nominated statistical models were evaluated as potential fits for the "uncured" population, with the exception of the generalised gamma. As a result of the flexibility of the generalised gamma distribution finding the global maximum likelihood of a mixture requires a more robust approach than the less flexible models, for which gradient descent algorithms are sufficient, and introduces additional uncertainty to the model. Given the small size of the dataset to be fitted, this was considered unlikely to be beneficial, i.e. models were likely to be overfitted versus models on the two parameter distributions, per the information criteria.

As an alternative to assuming zero excess hazard in the "cured" population, a secondary distribution was specified in addition to the background survival distribution. To limit the number of potential models, this was limited in all cases to a Weibull model, whilst the first excess hazard was fitted in turn by each of the nominated statistical models except the generalised gamma. In the majority of cases, however, these models converged to implicit/statistical cure models – the excess hazard in one sub-population was negligible – and are not reported here.

3.4. Assessment of fit

Assessment of extrapolations was undertaken on the basis of the following criteria:

- Goodness-of-fit statistics
- Visual inspection of the parametric fit over the observed period
- Consideration of the log cumulative hazard plots
- Clinical plausibility of the long-term hazard profile

Goodness-of-fit was evaluated using AIC and BIC. Minimisation of these measures indicates goodness-of-fit whilst penalising overfitting, therefore a smaller value demonstrates a more appropriate fit.

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 limited by relatively short duration of follow-up.

In addition, the information criteria were originally developed to prevent over-parameterisation in terms of including more explanatory variables than necessary, particularly linearly additive covariates. Hence, the purpose of these information criteria did not originally include comparison between different functional forms or differing numbers of knots in a spline. An example of the difficulties arising from this can be seen in the differences between a Weibull and a Gompertz survival model; these models receive the same penalty from the information criteria for degrees of freedom but offer differing transformations of the time variable. Selection between the two models then uses the time transformation as an implicit degree of freedom "fitted" by the analyst. This is obvious in the case of two such closely related models, but the same concept of analyst degree of freedom is true through all model forms, excepting those that are nested. Therefore, the appropriateness of curve fits should not depend solely on the information criteria and these criteria should not take precedence over the assessment of plausibility of long-term hazard progression, particularly when long periods of extrapolation are necessary.

The log cumulative hazard plots were examined to identify how closely the curve fits adhere to the hazard profile of the observed data. The assessment was made visually, as with the time to event curves. Each model provided a prediction of the hazard through the observed time with an indication as to its direction in the extrapolated period.

Final model selection was ultimately at the analysts' discretion, as there were no a priori specified models. The models were selected holistically per the above criteria, with reference to statistical goodness of fit, clinical plausibility, clinical expert opinion, data trends, external data in related indications, internal plausibility in relation to selected models for other outcomes, and the principle of parsimony. To evaluate the sensitivity of model predictions to this selection, a number of alternative models were selected to form scenario analyses.

3.5. Logistic models of ratio of progression to death events

The state transition model which this analysis informed has two competing transitions from the pre-progression state:

- Pre-progression -> Progressed disease
- Pre-progression -> Death

For the purpose of indirect comparison, only data for PFS was available, and so it was decided that this would be modelled directly, and the assumed relationship between the two competing transition intensities would be made explicitly by the use of a model of the ratio of PFS events that were deaths versus those that were progressions.

A logistic model of this ratio based upon the observed events within CheckMate 649 was formed. A number of alternative model forms were evaluated and selection was made on the basis of AIC.

3.6. Calibration of post-progression survival using NMA-derived hazard ratios.

From the ITC, paired hazard ratios of PFS and OS versus XELOX/FOLFOX were available. By making the following assumptions, only the post-progression to death state transition rate remained to be calibrated:

- The model of PFS of XELOX/FOLFOX could be scaled directly by the hazard ratio over all model time
- The proportion of PFS events which were due to death remained as per the model of section 4.6 irrespective of comparator therapy

To calibrate the model to the indirectly compared treatments, the Kaplan-Meier estimator of the XELOX/FOLFOX OS outcome from the CheckMate 649 trial was scaled by the NMA-derived hazard ratio. This was then used to determine patient-level weights in the CheckMate 649 data, such that the Kaplan-Meier estimator of the weighted PLD would equal the scaled Kaplan-Meier of the unweighted PLD:

$$\left[\prod_{i:t_i \le t}\left(1-\frac{d_i}{n_i}\right)\right]^{HR} = \prod_{i:t_i \le t}\left(1-\frac{d_i w_i}{n_{wi}}\right)$$

Where d_i is the number of events at t_i, w_i are the weights assigned to those events, n_i is the number of patients at risk at t_i, and n_{wi} is the weighted sum of patients at risk at this time.

These weights were used to determine the log-likelihood of the OS predicted by the economic model:

$$LL = \sum_{i=1}^{N} \ln\big(f(t_i)(d_i)\big)\, w_i + \ln\big(S(t_i)(1-d_i)\big)\, w_i$$

Where N is the total number of patients, d_i is the event indicator of patient i, where 1 = death and 0 = censor; $f(t_i)$ is the modelled OS density function and $S(t_i)$ is the modelled OS survival function.

The state transition model was replicated in the statistical programming language **R** and configured for the XELOX/FOLFOX arm of CheckMate 649. The PFS transition was scaled by the hazard ratio derived from the ITC, and the model of proportion of patients dying upon exiting the pre-progression state was maintained as in the base case. The post-progression disease specific survival was then scaled by a hazard ratio and the log-likelihood evaluated. This hazard ratio was then varied until maximum log-likelihood of OS was reached and the final value was taken.

3.7. Calculation of effective number of doses of nivolumab received

In order to scale the proportion of population on treatment with nivolumab plus XELOX or FOLFOX at any time to the number of doses dispensed, the drug exposure data of CheckMate 649 was used to determine an effective rate of dosing. This was done by consideration of the individual time to treatment discontinuation, the number of cycles of therapy expected to have been received by this time I, and the actual number of doses received N.

For patients assigned to nivolumab plus XELOX, the expected number of cycles was calculated as:

$$E_{XELOX} = floor\left(\frac{t-1}{21}\right) + 1$$

Where t was the number of days on therapy with 1 being equal to first exposure. For FOLFOX, the expected number of cycles was calculated as:

$$E_{FOLFOX} = floor\left(\frac{t-1}{14}\right) + 1$$

Some study centres modified dose to alter cycle length, e.g. due to considerations due to COVID-19 or modification of the chemotherapy component of therapy. The following rules were applied to adjust the expected number of cycles:

- For each instance of a patient receiving XELOX 240 mg at a visit, the expected number of cycles was increased by 0.5
- For each instance of a patient receiving XELOX 480 mg at a visit, the expected number of cycles was decreased by 1/3
- For each instance of a patient receiving FOLFOX 480 mg at a visit, the expected number of cycles was reduced by 1.

The effective administration intensity I was then calculated as:

$$I = \frac{N}{E}$$

3.8. General statistical considerations

All analyses were undertaken on an x64-based PC running Windows 10 Pro (v1909+), within the "R" statistical software environment version 4.0.2 (2020-06-22) as provided by CRAN (https://cran.r-project.org/). Relevant external statistical packages used were:

- survival (v3.1-12) (R base)
- flexsurv (v1.1.1) (https://CRAN.R-project.org/package=flexsurv)
- muhaz (v 1.6.2.1) (https://CRAN.R-project.org/package=muhaz)
- bshazard (v 1.1) (https://CRAN.R-project.org/package=bshazard)

4. Survival extrapolation from CheckMate 649

4.1. BICR-assessed PFS

4.1.1. Data description and assessment of proportional hazards

BICR-assessed PFS was shorter for the CHEMO arm than for NIVO + CHEMO. More events were experienced in the CHEMO arm (Table 4), the median survival was shorter, and survival at 6-monthly intervals was lower.

For both arms, BICR-assessed PFS showed a protocol-driven discontinuous survival profile over the initial period, where periodic assessment of progression status resulted in sharp changes in progression-free state occupancy (Figure 2). This pattern persisted through the trial, but step sizes decreased over trial follow-up.

In diagnostic transformations of the survival function (Figure 3) inconstant hazard was clearly shown (inconstant gradient in panel b). A non-Weibull profile was suggested with non-constant log cumulative hazard spacing, rejecting proportional hazards (panel c); a Gompertz hazard profile was not substantially contraindicated (panel d) and the log-logistic and lognormal models were suggested as suitable by steady gradient in panels (e) and (f) respectively, with the exception of study entry.

Smoothed estimates of the hazard functions demonstrated a consistent profile (Figure 4). Both arms showed a non-monotonic hazard profile, increasing to a peak and then decreasing, with the potential to reach zero excess hazard by the bspline and kernel-smoothed estimators. The CHEMO arm demonstrated a taller and broader peak in hazard. The NIVO + CHEMO arm estimates are impacted by the terminal event at end of follow-up, and the late increase in hazard function is an artifact of this.

A similar distribution of time to end of BICR-assessed PFS is expected for both arms, but a simple scaling rule, such as proportional hazards or accelerated failure time, is not supported. The hazard function is non-monotonic, and is discontinuous, particularly in the early trial. Due to these discontinuities continuous models from randomisation were expected to have compromised fit to the trial data.

Table 4: Observed BICR-assessed PFS, CheckMate 649

Endpoint	NIVO+CHEMO (N=473)	CHEMO (N=482)
PFS per BICR		
Events, N	[redacted]	[redacted]
Median (months) (95% CI)	[redacted]	[redacted]
Landmark survival (%) (95% CI)		
6 months	[redacted]	[redacted]
12 months	[redacted]	[redacted]
18 months	[redacted]	[redacted]
24 months	[redacted]	[redacted]
36 months	[redacted]	[redacted]

BICR: Blinded independent central review; CI: Confidence interval; PFS: Progression-free survival
Survival per Kaplan-Meier estimator, confidence interval by Greenwood estimate of variance upon log-log survival

[redacted]

Figure 2. BICR-assessed PFS: Kaplan-Meier

BICR: Blinded independent central review; NAR: number at risk; PFS: Progression-free survival

[redacted]

Figure 3. BICR-assessed PFS: Ishak diagnostic plots

BICR: Blinded independent central review; PFS: Progression-free survival. Ishak et al (2013).[24]

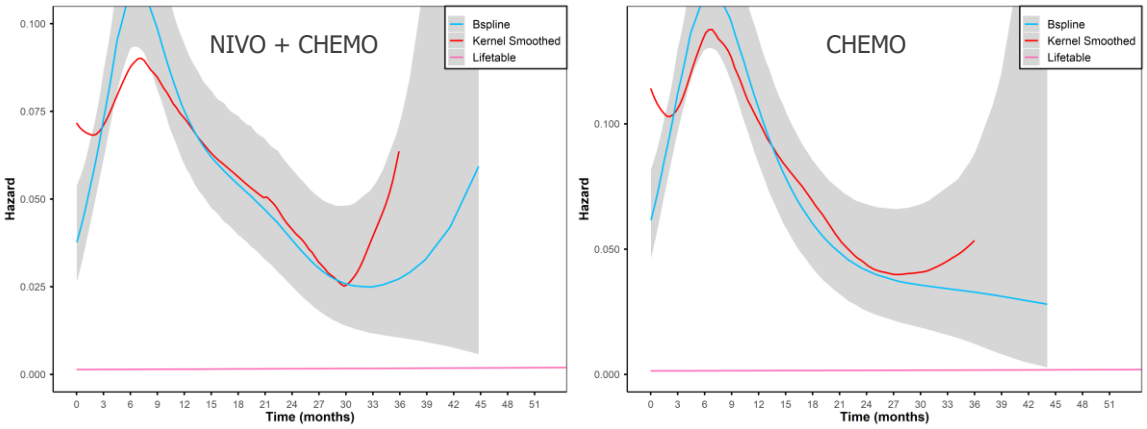

Figure 4. BICR-assessed PFS: Smoothed hazard function estimates.

Confidence interval is shown around b-spline estimator

4.1.2. Standard statistical models (relative survival framework)

Arm-independent models of BICR-assessed PFS were fitted assuming a single statistical distribution of excess hazard over all time.

For the NIVO+CHEMO arm (Figure 5, Figure 6), constant and monotonic hazard models (exponential, Weibull, Gompertz) did fit poorly, as expected, and as per investigator-assessed PFS. In this case, the other models also had difficulty in modelling the acute peak in hazard at around 6 months; all models entered a period of under-prediction prior to this point and over-prediction after. The log-logistic model was best capable of representing the curvature in the cumulative hazard function after this point, and so was considered the most appropriate for extrapolation; nevertheless, to improve within-trial goodness-of-fit, the semiparametric approach taken with investigator-assessed PFS was considered.

For the CHEMO arm (Figure 7, Figure 8), the same observations were made as for the NIVO+CHEMO arm, except that the deviation was more obvious for most models, particularly in over-prediction of survival after the hazard peak, which in this case occurred slightly later at approximately 9 months. Again, the log-logistic model best approximated the decreasing gradient of the cumulative hazard function approaching the end of follow-up.

[redacted]

Figure 5. BICR-assessed PFS, NIVO+CHEMO: Standard statistical models overlaid upon Kaplan-Meier

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; BICR: Blinded independent central review; LTM: Lifetable mortality; PFS: Progression-free survival. 95% confidence intervals obtained by data bootstrap (1000 repetitions)

Figure 6. BICR-assessed PFS, NIVO+CHEMO: Hazard profile of standard statistical models

BICR: Blinded independent central review; LTM: Lifetable mortality; PFS: Progression-free survival

Figure 7. BICR-assessed PFS, CHEMO: Standard statistical models overlaid upon Kaplan-Meier

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; BICR: Blinded independent central review; LTM: Lifetable mortality; PFS: Progression-free survival.. 95% confidence intervals obtained by data bootstrap (1000 repetitions)

Figure 8. BICR-assessed PFS, CHEMO: Hazard profile of standard statistical models

BICR: Blinded independent central review; LTM: Lifetable mortality; PFS: Progression-free survival

4.1.3. Semi-parametric models

Piecewise models using Kaplan-Meier data to a pre-specified cut point and a parametric relative survival model thereafter were fitted to the data. A number of potential cut points were considered, avoiding assessment windows due to the rapid change in hazard near the model start time implied by these periods. As for investigator-assessed PFS, a cut point of 6.44 months was chosen as a good compromise between allowing for the expression of study-entry effects and the development of response to therapy whilst maximising data for informing the extrapolative parametric model.

For the NIVO+CHEMO arm (Figure 9, Figure 10) the 6.44 month cut point was shortly after the peak in hazard, and so monotonically decreasing hazard models could be considered. The exponential model was a poor fit to the data, but all others were acceptable. The lognormal, generalised gamma and log-logistic models were the best statistical fits, indicating that although the Weibull model was at many times closer to the central Kaplan-Meier maximum-likelihood estimator in the tail, the low number of patients meant that the apparent straightening of the cumulative hazard function may have been due to chance, and the likelihood function favoured models that continued the trend in reducing excess hazard; the final event with low numbers of patients at risk does not greatly affect the likelihood function.

For the CHEMO arm (Figure 11, Figure 12) both the exponential and the Weibull models fitted poorly, under-predicting survival from ~18 months. Of the remaining, the log-logistic, lognormal and generalised gamma models fitted the data well, producing similar cumulative hazard curves into extrapolation; the Gompertz model decreased the cumulative hazard gradient more quickly and resulted in a worse fit.

Over both arms, there was little to distinguish between the log-logistic, lognormal and generalised gamma models, and so the lognormal model was preferred due to its low AIC/BIC.

Figure 9. BICR-assessed PFS, NIVO+CHEMO: Semi-parametric (piecewise) relative survival models overlaid upon Kaplan-Meier

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; BICR: Blinded independent central review; LTM: Lifetable mortality; PFS: Progression-free survival.. 95% confidence intervals obtained by data bootstrap (1000 repetitions)

Figure 10. BICR-assessed PFS, NIVO+CHEMO: Log cumulative hazard of semi-parametric (piecewise) relative survival models

BICR: Blinded independent central review; LTM: Lifetable mortality; PFS: Progression-free survival

Figure 11. BICR-assessed PFS, CHEMO: Semi-parametric (piecewise) relative survival models overlaid upon Kaplan-Meier

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; BICR: Blinded independent central review; LTM: Lifetable mortality; PFS: Progression-free survival. . 95% confidence intervals obtained by data bootstrap (1000 repetitions)

Figure 12. BICR-assessed PFS, CHEMO: Log cumulative hazard of semi-parametric (piecewise) relative survival models

BICR: Blinded independent central review; LTM: Lifetable mortality; PFS: Progression-free survival

4.1.4. Mixture-cure models

Models assuming that a proportion of the cohort are within a long term responsive fraction from randomisation demonstrated similar features to the parametric relative survival models (Figure 13, Figure 14) for the NIVO+CHEMO arm. As with investigator-assessed PFS, a long-term response/cure fraction was not always identified, particularly for the non-monotonic hazard functions. With the exception of the exponential and Gompertz models, fit to the trial data was acceptable.

For the CHEMO arm (Figure 15, Figure 16), the same conclusions are drawn, i.e. that all fits with the exception of the exponential and Gompertz models are acceptable, but the long-term response fraction is dependent upon the distribution. It may be as high as ~8% or may be undetectable. As with investigator-assessed PFS, these models did not introduce an exploration of uncertainty in a controlled way and so were not used.

Figure 13. BICR-assessed PFS, NIVO+CHEMO: Mixture-cure survival models overlaid upon Kaplan-Meier

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; LTR: Long term response; LT: Lifetable mortality. 95% confidence intervals obtained by data bootstrap (1000 repetitions)

Figure 14. BICR-assessed PFS, NIVO+CHEMO: hazard profiles of mixture-cure parametric relative survival models

LTR: Long term response. LT: lifetable

Figure 15. BICR-assessed PFS, CHEMO: mixture-cure survival models overlaid upon Kaplan-Meier

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; LTR: Long term response; LT: Lifetable mortality. 95% confidence intervals obtained by data bootstrap (1000 repetitions)

Figure 16. BICR-assessed PFS, CHEMO: hazard profiles of mixture-cure parametric relative survival models

LTR: Long term response. LT: lifetable

4.1.5. Model selection algorithm

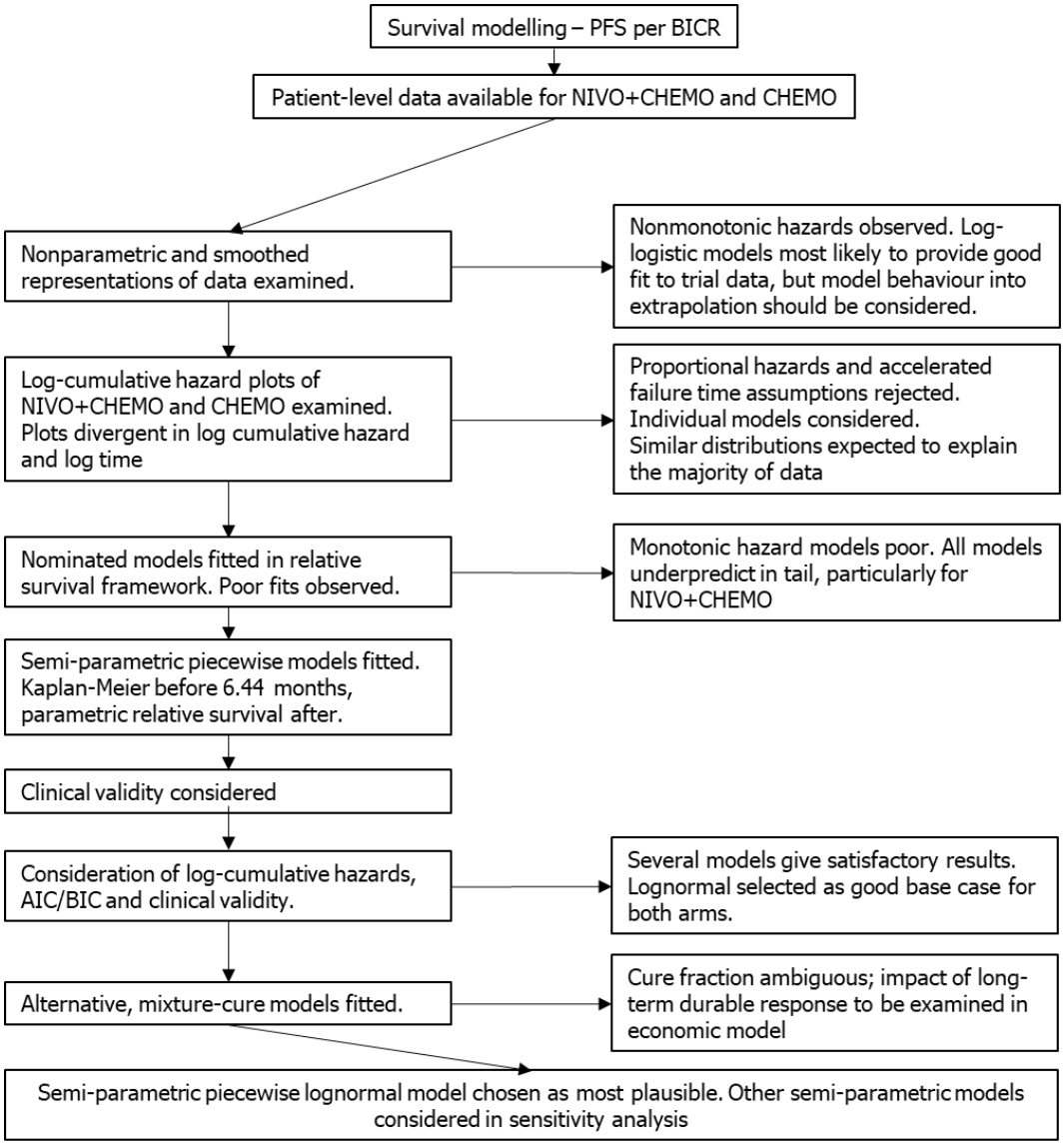

Figure 17. BICR-assessed PFS: Considerations and assessments made when modelling

4.1.6. Conclusions on BICR-assessed PFS

- Standard statistical models are a poor fit
- Semi-parametric models can fit well
- Mixture-cure models do not consistently identify a cure fraction among best-fitting models; the impact of long-term durable response is to be examined within the economic model based upon proportion of patients surviving to end of follow-up

- Semi-parametric lognormal is base case for both arms

4.2. Investigator-assessed PFS

4.2.1. Data description and assessment of proportional hazards

Investigator-assessed PFS was shorter for the CHEMO arm than for the NIVO+CHEMO arm. More events were experienced in the CHEMO arm (Table 5), the median survival was shorter and survival at 6-monthly intervals was uniformly lower.

For both arms, investigator-assessed PFS showed a protocol-driven discontinuous survival profile over the initial period, where periodic assessment of progression status resulted in marked changes in state occupancy (Figure 18). Some evidence of this persisted throughout the trial, but the step size in the survival function did decrease over time. The censoring pattern was poor, as a large number of patients were censored at last response assessment, represented by clustering of censor marks at the steps in the plot. These confirmed non-progressions would be expected to put the patients at a lower than average risk of progression in the days immediately after assessment, but by applying the censor at this time, events measured in the following days among those with a slightly later assessment are assumed equally likely to have occurred to the censored patients – as such, for both arms, the Kaplan-Meier estimator is likely to underestimate the PFS function.

In comparing diagnostic transformations of the survival function (Figure 19), inconstant hazard was demonstrated (inconstant gradient, panel b). The proportional hazards assumption was not well supported, as there was variable vertical spacing on the log-log transformed data (panel c), and a Weibull profile was contra-indicated by the decreasing gradient on this transform. A decreasing-hazard Gompertz transform was not contraindicated (panel d). An independent log-logistic model for each arm was not contraindicated (panel e). With the exception of the trial entry period, a log-normal model was considered plausible (panel f).

All smoothed estimators of the hazard function demonstrate a similar profile (*Figure 20). The hazard increases to a peak, before declining; this peak was between 7-8 months for NIVO+CHEMO and was between 8 and 11 months for CHEMO. The peak for CHEMO was substantially higher than for NIVO+CHEMO and was sustained for longer. The kernel-windowing estimator is known to have poor behaviour near the edges of the data domain, and so the tails are an artifact of the final events occurring after heavy censoring; after discounting the tail of this estimator, excess hazard monotonically decreases from the ~9 months peak for both arms. Sudden variations in the hazard function in the early trial for the Royston-Parmar smoother are the result of the fitter prioritising representation of the high-hazard assessment periods, resulting in an apparently discontinuous model.

In conclusion, time to end of PFS appears similarly distributed between both arms of the trial, but there is insufficient evidence for a scaling rule such as proportional hazards and given the difference in mechanism of action between the interventions on both arms (section 1.2), such a scaling rule is not plausible. However, given the similarity in the distribution of survival times, a similar modelling approach should be taken for each arm, and this may include using the same distribution for both arms, though with independently derived parameters. The hazard function for both arms is non-monotonic and decreases from a peak at around 9 months; a log-logistic may be appropriate, but the behaviour of the model as it approaches extrapolation should be scrutinised. The early survival function is highly stepped due to the review schedule, and significant time under the curve may be lost if these steps are smoothed by a continuous model.

Table 5: Observed investigator-assessed PFS, CheckMate 649

Endpoint	NIVO+CHEMO (N=473)	CHEMO (N=482)
PFS per investigator		
Events, N	[redacted]	[redacted]
Median (months) (95% CI)	[redacted]	[redacted]
Landmark survival (%) (95% CI)		
6 months	[redacted]	[redacted]
12 months	[redacted]	[redacted]
18 months	[redacted]	[redacted]
24 months	[redacted]	[redacted]
36 months	[redacted]	[redacted]

CI: confidence interval; NIVO: nivolumab; PFS: progression-free survival;
Survival per Kaplan-Meier estimator, confidence interval by Greenwood estimate of variance upon log-log survival

[redacted]

Figure 18. Investigator-assessed PFS: Kaplan-Meier

[redacted]

Figure 19. Investigator-assessed PFS: Ishak diagnostic plots

Dashed lines are simple linear regression upon the plotted data. Source: Ishak et al (2013)[24].

[redacted]***Figure 20. Investigator-assessed PFS: Smoothed hazard function estimates***

R-P: Royston-Parmar. Confidence interval is shown around b-spline estimator

4.2.2. Standard statistical models (relative survival framework)

Arm-independent models of PFS were fitted assuming a single statistical distribution of excess hazard over all time.

For the NIVO+CHEMO arm (Figure 21, Figure 22), constant and monotonic hazard models (exponential, Weibull, Gompertz) fitted poorly, as expected from the data inspection. The non-monotonic models (log-logistic, lognormal, generalised gamma) fitted considerably better to the early data, but tended to underpredict in the later trial, i.e. the marginal hazard function did not decrease sufficiently quickly. This can be indicative of a mixed population with a small "low-risk" fraction; the exhaustion of a "high risk" fraction can result in a rapid reduction in marginal hazard that cannot be represented by the distribution that dominates the early data. Of these models, the lognormal appears the best fitting both to the overall trial data and the tail and has the lowest BIC of all possible models as well as having low AIC, indicating good parsimony of fit. However, its prediction into extrapolation may be pessimistic.

For the CHEMO arm (Figure 23, Figure 24), the constant and monotonic hazard models fitted poorly. The non-monotonic models did fit better to the early data but, in contrast to the NIVO+CHEMO arm, tended to slightly overpredict survival during the second year. Overall, the fit to the data was slightly better than on the NIVO+CHEMO arm for these models, and the lognormal model was once again a

good fit according to the AIC and BIC and was relatively close to the prediction to the tail of the Kaplan-Meier.

To decrease the constraints upon these models in representing these later hazard changes, a piecewise approach was attempted in the following section, whereby the early model was represented directly by the Kaplan-Meier estimator, and the parametric relative survival model was fitted conditional upon survival to a cut time, reducing the necessity for the parametric model to represent the hazard function in the early trial to prioritise fitting to the later trial.

Figure 21. Investigator-assessed PFS, NIVO+CHEMO: Standard statistical models overlaid upon Kaplan-Meier

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; LTM: Lifetable mortality. 95% confidence intervals of models obtained by data bootstrap (1000 repetitions)

Figure 22. Investigator-assessed PFS, NIVO+CHEMO: Hazard profile of parametric relative survival models

LTM: Lifetable mortality.

Figure 23. Investigator-assessed PFS, CHEMO: Standard statistical models overlaid upon Kaplan-Meier

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; LTM: Lifetable mortality. 95% confidence intervals of models obtained by data bootstrap (1000 repetitions)

Figure 24. Investigator-assessed PFS, CHEMO: Hazard profile of parametric relative survival models

LTM: Lifetable mortality

4.2.3. Semi-parametric models

Piecewise models using Kaplan-Meier data to a pre-specified cut point and a parametric relative survival model thereafter were fitted to the data. A number of potential cut points were considered, avoiding assessment windows due to the rapid change in hazard near the model start time implied by these periods. As a compromise between maximisation of data for use in extrapolation and removal of the largest hazard discontinuities, a time of 6.44 months (i.e. the end of the 4th assessment window + 3 weeks) was chosen. Per the CheckMate 649 study protocol, regular 6-weekly assessments continued until week 48 (assessment 8), then replaced by 12-weekly assessments; however, delaying the parametric model until fully clear of this high-frequency period resulted in loss of approximately half of the patients available for modelling at the 6.44 month point without substantially affecting the model fit, and so this point was chosen as a compromise.

For the NIVO+CHEMO arm, the majority of models were acceptable from this cut point (Figure 25, Figure 26), with the exceptions being the constant-hazard (exponential) model and the Weibull model, the hazard function of which did not decrease rapidly enough to represent the flattening of

the tail of the data. Of the remainder, the generalised gamma model was lowest per AIC but not BIC and did not demonstrate the expected curvature in cumulative hazard implied by the cumulative hazard curve. Of the remainder, the Gompertz model was lowest per AIC and BIC, but the hazard profile of this model demonstrated a regression to zero excess hazard (i.e. lifetable hazard only) by approximately 10 years. As a statistical cure model, this hazard profile was not suitable for use in scenarios with a mixture-cure fraction in the economic model. Generalised gamma model did not demonstrate the expected curvature in cumulative hazard implied by the curvature of the hazard curve. Lognormal presented a good hazard profile and so was chosen as an alternative to the Gompertz.

Similar observations were made in models of the CHEMO arm as were made for the NIVO+CHEMO arm. The exponential and Weibull models were poor, but the remaining models fitted the data well. The Gompertz model did not have such good statistical goodness of fit in this case, which was taken into account when considering its applicability for the NIVO+CHEMO arm. The lognormal model was again chosen as a good representation of the data.

Figure 25. Investigator-assessed PFS, NIVO+CHEMO: Semi-parametric (piecewise) relative survival models overlaid upon Kaplan-Meier

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; LTM: Lifetable mortality. 95% confidence intervals obtained by data bootstrap (1000 repetitions). AIC and BIC for parametric portion of model only.

Figure 26. Investigator-assessed PFS, NIVO+CHEMO: hazard profiles of semi-parametric (piecewise) relative survival models

LTM: Lifetable mortality

Figure 27. Investigator-assessed PFS, CHEMO: Semi-parametric (piecewise) relative survival models overlaid upon Kaplan-Meier

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; LTM: Lifetable mortality. 95% confidence intervals obtained by data bootstrap (1000 repetitions). AIC and BIC for parametric portion of model only.

Figure 28. Investigator-assessed PFS, CHEMO: hazard profiles of semi-parametric (piecewise) relative survival models

LTM: Lifetable mortality

4.2.4. Mixture-cure models

Models assuming a proportion of the cohort are within a long-term responsive fraction from randomisation demonstrated similar features to the parametric relative survival models for the NIVO+CHEMO arm (Figure 29, Figure 30). Constant and monotonic hazard models did not fit the data well. The remining models, with the exception of the Gompertz, fitted well to the main bulk of the data but resulted in substantially different estimates of the cure fraction. For those models that predicted increasing hazard among the uncured fraction, a cure fraction of more than 10% was predicted, whereas those models that incorporate a long-term declining hazard (log-logistic, lognormal) predicted cure fractions of less than 3%. Given the comparability of goodness-of-fit

between the majority of models, but the high variation in cure fraction predicted, this data-driven approach to identifying a mixture-cure model was not felt to decrease the uncertainty in the decision problem.

Similar conclusions held for the CHEMO arm (Figure 31, Figure 32); the exponential and Gompertz models were very poor fits, and those models that predicted increasing hazard predicted higher cure fractions. However, in this case, models with a long-term decreasing excess hazard predicted no cure fraction. These models did not represent as well the rate of change of hazard in the period after 12 months, but the clinical plausibility of a > 5% fraction experiencing long-term stable disease receiving XELOX/FOLFOX is questionable. Again, the mixture models were not felt to clarify the decision problem and the piecewise models were used.

Figure 29. Investigator-assessed PFS, NIVO+CHEMO: Mixture-cure survival models overlaid upon Kaplan-Meier

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; LTR: Long term response; LT: Lifetable mortality. 95% confidence intervals obtained by data bootstrap (1000 repetitions)

Figure 30. Investigator-assessed PFS, NIVO+CHEMO: hazard profiles of mixture-cure parametric relative survival models

LTR: Long term response. LT: lifetable

Figure 31. Investigator-assessed PFS, CHEMO: Mixture-cure survival models overlaid upon Kaplan-Meier

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; LTR: Long term response; LT: Lifetable mortality. 95% confidence intervals obtained by data bootstrap (1000 repetitions)

Figure 32. Investigator-assessed PFS, CHEMO: hazard profiles of mixture-cure parametric relative survival models

LTR: Long term response. LT: lifetable

4.2.5. Model selection algorithm

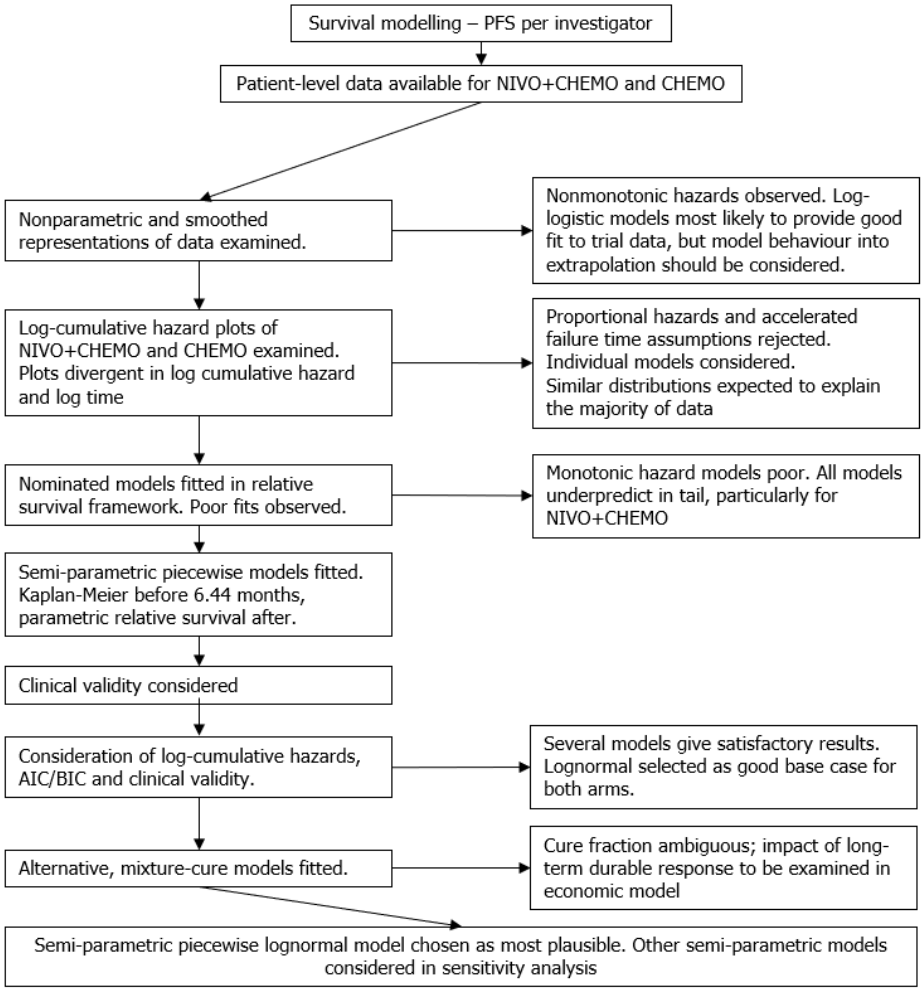

Figure 33. Investigator-assessed PFS: Considerations and assessments made when modelling

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; PFS: progression-free survival

4.2.6. Conclusions on investigator-assessed PFS

- Standard statistical models are a poor fit
- Semi-parametric models can fit well
- Mixture-cure models do not consistently identify a cure fraction among best-fitting models; impact long-term durable response to be examined within economic model based upon proportion of patients surviving to end of follow-up
- Semi-parametric lognormal is chosen as base case for both arms

4.3. Overall Survival

4.3.1. Data description and assessment of proportional hazards

Overall survival was shorter for the CHEMO arm than for the NIVO + CHEMO arm. More events were experienced in the CHEMO arm (Table 6), the median survival was significantly shorter and survival at 6-monthly intervals was lower.

In comparison to PFS, the Kaplan-Meier estimator of OS was continuous and did not show marked discontinuities (Figure 34). The survival profiles of both arms are marked by an initial period of high rate of mortality followed by a gradual reduction starting between 12 and 18 months.

This is visible on panel (b) of Figure 35, as the gradient (hazard) begins to decrease at this point. The reduction was greater in the NIVO+CHEMO arm. Due to this change in gradient, the exponential model was not indicated. In panel (c), the curves are curving relative to the reference linear regressions, and so Weibull models from randomisation were not indicated. Separation of the curves on this transform occurred after 3 months, indicating a non-proportional hazards relationship. This, in addition to the non-proportionality of the PFS outcome and the difference in methods of action of the two therapeutic regimes, excluded proportional hazards modelling. A gompertz model with decreasing hazard was not excluded by panel (d). Multiple crossings were seen on panel (e), indicating that a log-logistic model may be acceptable; by contrast, panel (f) showed a monotonic curve deviating from the linear regression, indicating against the log-normal distribution.

Smoothed estimates of the hazard functions demonstrated an increasing-decreasing hazard function in both arms ([redacted]Figure 36). Both peak and sustained hazard was higher in the CHEMO arm. Hazard on the NIVO+CHEMO arm trended towards the matched population hazard.

Table 6: Observed OS, CheckMate 649

Endpoint	NIVO+CHEMO (N=473)	CHEMO (N=482)
OS		
Events, N	[redacted]	[redacted]
Median (months) (95% CI)	[redacted]	[redacted]
Landmark survival (%) (95% CI)		
6 months	[redacted]	[redacted]
12 months	[redacted]	[redacted]
18 months	[redacted]	[redacted]
24 months	[redacted]	[redacted]
36 months	[redacted]	[redacted]

CI: Confidence interval; OS: Overall survival
Survival per Kaplan-Meier estimator, confidence interval by Greenwood estimate of variance upon log-log survival

[redacted]

Figure 34. OS: Kaplan-Meier

NAR: number at risk; OS: Overall survival

Figure 35. OS: Ishak diagnostic plots

OS: Overall survival. Source: Ishak et al (2013).[24]

Figure 36. OS: Smoothed hazard function estimates

Confidence interval is shown around b-spline estimator

4.3.2. Standard statistical models (relative survival framework)

Arm-independent models of OS were fitted assuming a single statistical distribution of excess hazard over all time.

For the NIVO+CHEMO arm (Figure 37, Figure 38) all models tended to underpredict in the tail, and either followed a pattern of early underprediction (Gompertz, log-logistic) or overprediction in the second year (Weibull). The Lognormal and Generalised gamma were the most successful models, but prediction in the tail was low.

For the CHEMO arm (Figure 39, Figure 40) the same pattern was seen, with these unimodal distributions unable to represent well the changes in marginal hazard observed within the trial. The lognormal showed the best visual fit, but underpredicts early in the trial, compromising its likelihood. In both arms, the log-logistic is the superior model by AIC and BIC; however, the representation of the hazard in the tail of the data could be improved.

Figure 37. OS, NIVO+CHEMO: Standard statistical models overlaid upon Kaplan-Meier

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; LTM: Lifetable mortality; OS: Overall survival. 95% confidence intervals obtained by data bootstrap (1000 repetitions)

Figure 38. OS, NIVO+CHEMO: Hazard profile of standard statistical models

LTM: Lifetable mortality; OS: Overall survival

Figure 39. OS, CHEMO: Standard statistical models overlaid upon Kaplan-Meier

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; LTM: Lifetable mortality; OS: Overall survival. 95% confidence intervals obtained by data bootstrap (1000 repetitions)

Figure 40. OS, CHEMO: Hazard profile of standard statistical models

LTM: Lifetable mortality; OS: Overall survival

4.3.3. Semi-parametric models

As the standard statistical models demonstrated difficulty in representing the changes in the hazard function within the trial, a delay to the start of the parametric extrapolative model was introduced in line with the PFS model. The same cut point was chosen for consistency, as this represented the start

of a time of more continuous changes in patient status and management both in the model and in the trial.

For the NIVO+CHEMO arm (Figure 41, Figure 42), the majority of models (with the exception of the exponential model) had acceptable fit to the data. The Weibull model was also on the lower boundary of the confidence interval in the tail under prediction, and so is not favoured. The Gompertz model was superlative per AIC and BIC, with a difference of almost 2 units from the log-logistic. This model represented the rapid reduction in hazard in the tail of the data.

For the CHEMO arm (Figure 43, Figure 44) all models were acceptable, though the exponential predicted low in the tail, similar to the Weibull for NIVO+CHEMO. As with the NIVO+CHEMO arm, the Gompertz model was superior by information criteria, by a difference of 2 units, due to its ability to represent the rapidly reducing hazard in the tail of the data.

Figure 41. OS, NIVO+CHEMO: Semi-parametric (piecewise) relative survival models overlaid upon Kaplan-Meier

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; LTM: Lifetable mortality; OS: Overall survival. 95% confidence intervals obtained by data bootstrap (1000 repetitions)

Figure 42. OS, NIVO+CHEMO: Hazard profile of semi-parametric (piecewise) relative survival models

LTM: Lifetable mortality; OS: Overall survival

Figure 43. OS, CHEMO: Semi-parametric (piecewise) relative survival models overlaid upon Kaplan-Meier

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; LTM: Lifetable mortality; OS: Overall survival. 95% confidence intervals obtained by data bootstrap (1000 repetitions)

Figure 44. OS, CHEMO: Hazard profile of semi-parametric (piecewise) relative survival models

LTM: Lifetable mortality; OS: Overall survival

4.3.4. Mixture-cure models

Models assuming a proportion of the cohort are within a long term surviving fraction from randomisation demonstrated better goodness of fit than the standard statistical models.

For NIVO+CHEMO, the rate of long-term survival (parameter "theta") varied depending upon the hazard function of the non-long-term survival fraction (Figure 45). Distributions showing monotonically increasing hazards (Weibull, Gamma, Gompertz) estimated long term survival fractions between 17% and 20%, whilst the long-tailed lognormal distribution estimated a fraction of 5% and had a correspondingly low mean. The log-logistic distribution and gamma distribution were very similar by information criteria but demonstrated fractions of 8.5% and 17.5% respectively. Hazard

profiles for these models (Figure 46) demonstrate the relative speed of convergence with general population mortality, with the short-tailed distributions converging around 60 months.

For the CHEMO arm (Figure 47, Figure 48) fits to the trial data were not as well centred, and for the lognormal and log-logistic distributions demonstrated negligible long-term survival fraction. Distributions with a non-negligible cure fraction had expected survival in excess of two years, which may be unexpectedly high from a clinical perspective.

Given the uncertainty in long-term survival fraction, these models are not recommended for the base case, but may be explored in sensitivity analysis.

Figure 45. OS, NIVO+CHEMO: Mixture-cure survival models overlaid upon Kaplan-Meier

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; LTR: Long term response; LT: Lifetable mortality; OS: Overall survival. 95% confidence intervals obtained by data bootstrap (1000 repetitions)

Figure 46. OS, NIVO+CHEMO: Hazard profile of mixture-cure parametric relative survival models

LTM: Lifetable mortality; OS: Overall survival

Figure 47. OS, CHEMO: Mixture-cure survival models overlaid upon Kaplan-Meier

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; LTR: Long term response; LT: Lifetable mortality; OS: Overall survival. 95% confidence intervals obtained by data bootstrap (1000 repetitions)

Figure 48. OS, CHEMO: Hazard profile of mixture-cure parametric relative survival models

LTM: Lifetable mortality; OS: Overall survival

4.3.5. Model selection algorithm

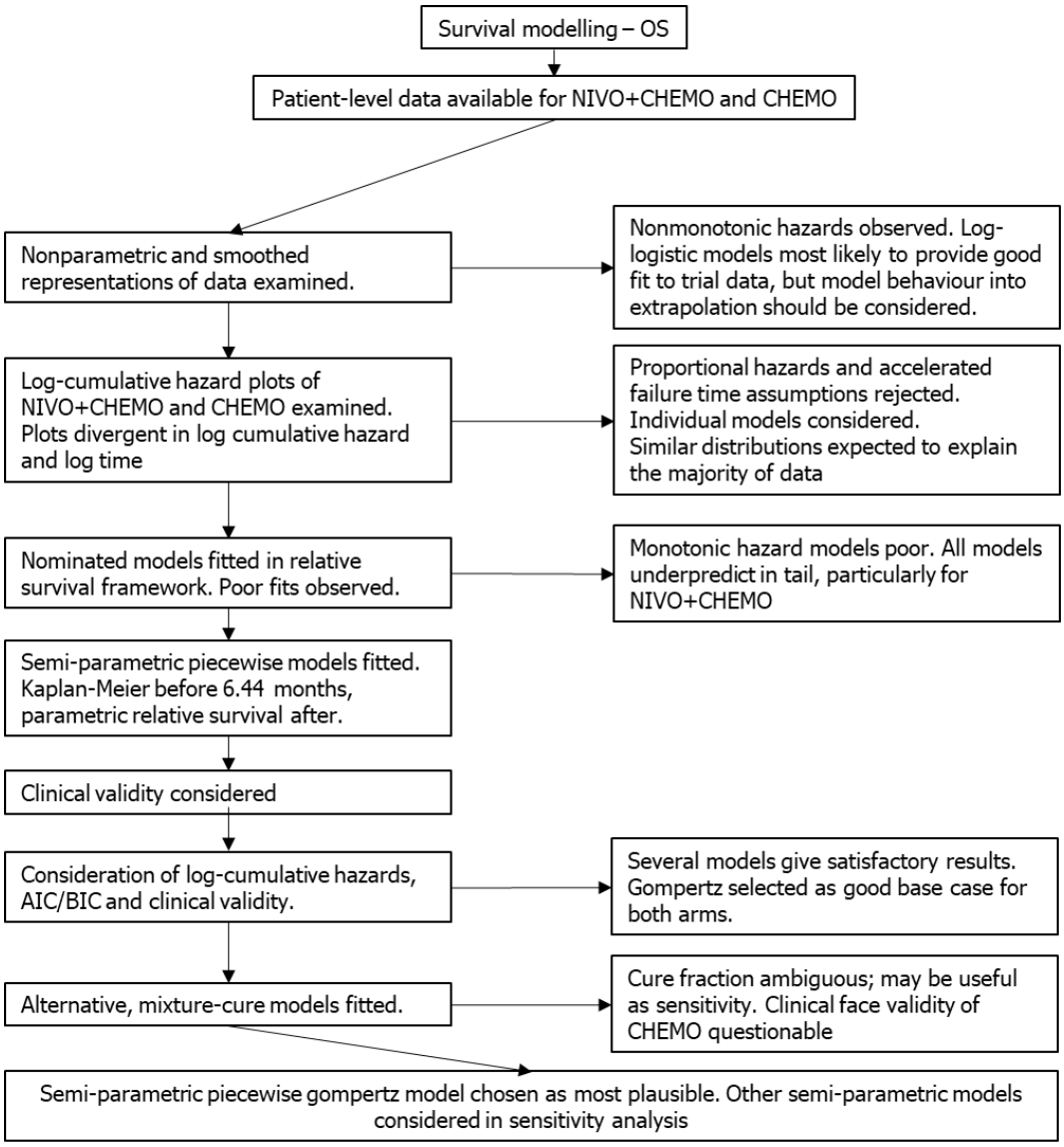

Figure 49 OS: Considerations and assessments made when modelling

4.3.6. Conclusions on OS

- Standard statistical models are a poor fit
- Semi-parametric models can fit well
- Mixture-cure models give a wide range of potential long-term survivor fractions for NIVO+CHEMO and may be suited for sensitivity analysis. On the CHEMO arm, mean survival of models with a non-negligible long-term survivor fraction may not have clinical face validity.
- Semi-parametric Gompertz is base case for both arms

4.4. Post-progression survival

4.4.1. Data description and assessment of proportional hazards

PPS was similar in the two arms (Table 7: Observed PPS (progression per BICR), CheckMate 649). At all times after progression, the survival rate was higher for the NIVO+CHEMO arm than for the CHEMO arm, and the median was slightly extended for NIVO+CHEMO. In the Kaplan-Meier (Figure 50), the difference in survival rate can be seen to be increasing on average, though panel (c) of Figure 51 does not indicate a clear proportional hazards relationship, with the spacing of the log cumulative hazard expanding in the later period. Curvature in panels (b) and (c) contraindicated the exponential and Weibull distributions. The Gompertz profile could be considered (panel d), but whilst the trend was generally to increasing hazard, this was not true for those surviving to 18 months post-progression in either arm. The repeated crossing of the regression line in panels (e) and (f) indicated that both the log-logistic and lognormal models were good candidates, dependent upon behaviour into extrapolation, which was of particular concern for the CHEMO model where overprediction was recognised as a potential issue.

Smoothed estimators of the hazard function also demonstrated a long-term decreasing hazard profile ([redacted]Figure 52), though in contrast to PFS, initial hazard was nearer to peak, and so monotonic models could be considered on the NIVO+CHEMO arm. The hazards on the CHEMO arm did not reduce until later than NIVO+CHEMO (if at all, decreasing long-term hazards were not demonstrated by all smoothers); in both cases a long-term reduction in excess hazard to zero was not inconsistent with the data across all smoothers, but was considerably more uncertain for the CHEMO arm.

As with PFS, models from the same family were preferred for both arms due to the similarity of the hazard profile over the trial period.

Table 7: Observed PPS (progression per BICR), CheckMate 649

Endpoint	NIVO+CHEMO (N=271)	CHEMO (N=286)
PPS		
Events, N	[redacted]	[redacted]
Median (months) (95% CI)	[redacted]	[redacted]
Landmark survival (%) (95% CI)		
6 months post progression	[redacted]	[redacted]
12 months post progression	[redacted]	[redacted]
18 months post progression	[redacted]	[redacted]
24 months post progression	[redacted]	[redacted]
36 months post progression		

CI: Confidence interval; PPS: post-progression survival
Survival per Kaplan-Meier estimator, confidence interval by Greenwood estimate of variance upon log-log survival

[redacted]

Figure 50. PPS: Kaplan-Meier

[redacted]

Figure 51. PPS: Ishak diagnostic plots

Source: Ishak et al (2013)[24]

Figure 52. PPS: Smoothed hazard function estimates.

PPS: post-progression survival; R-P: Royston-Parmar. Confidence interval is shown around b-spline estimator

4.4.2. Standard statistical models (relative survival framework)

Arm-independent models of PPS were fitted assuming a single statistical distribution of excess hazard over all time.

For the NIVO+CHEMO arm (Figure 53, Figure 54), all models fit the initial data well, with divergence occurring approximately 12 months after progression. From this point the exponential, Weibull and Gompertz models tended to under-predict survival. No model matched the apparent gradient of the Kaplan-Meier over the final 12 months, with all appearing to under-represent long-term survival. To compensate for this, the log-logistic model was chosen in the base case, in consideration of its superlative BIC and approximation of the gradient of the cumulative hazard function into extrapolation.

For the CHEMO arm, the same observations applied, i.e. that the exponential and monotonic fits under-predicted survival in the tail, but that in all cases the fits were unable to match the gradient through this period and so may extrapolate low. The generalised gamma appeared to be the best compromise in this case, but to match the selection for the NIVO+CHEMO arm, the log-logistic model was chosen for the base case.

Figure 53. PPS, NIVO+CHEMO: Standard statistical models overlaid upon Kaplan-Meier.

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; LTM: Lifetable mortality; PPS: post-progression survival. 95% confidence intervals obtained by data bootstrap (1000 repetitions)

Figure 54. PPS, NIVO+CHEMO: Hazard profile of parametric relative survival models

LTM: Lifetable mortality; PPS: post-progression survival

Figure 55. PPS, CHEMO: Standard statistical models overlaid upon Kaplan-Meier

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; LTM: Lifetable mortality; PPS: post-progression survival 95% confidence intervals obtained by data bootstrap (1000 repetitions)

Figure 56. PPS, CHEMO: Hazard profile of parametric relative survival models

LTM: Lifetable mortality; PPS: post-progression survival

4.4.3. Model selection algorithm

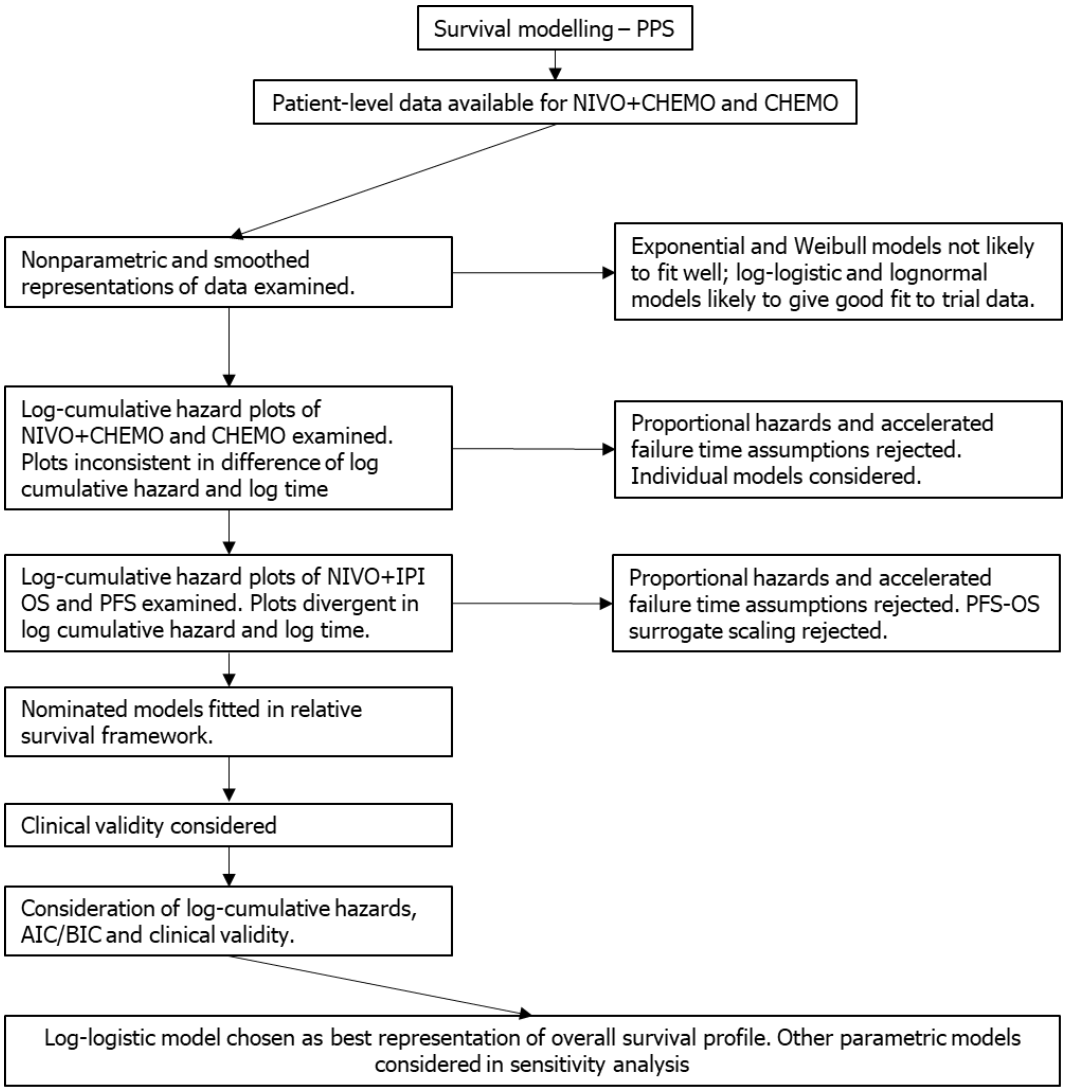

Figure 57. PPS: Considerations and assessments made when modelling

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; OS: overall survival; PFS: progression-free survival; PPS: post-progression survival

4.4.4. Conclusions on PPS

- Standard statistical models can fit adequately
- The log-logistic model is chosen for base case, which may be slightly optimistic for the CHEMO arm. The generalised gamma is also well supported

4.5. Time to treatment discontinuation

4.5.1. Data description and assessment of proportional hazards

The distribution of time to treatment discontinuation was lower for the CHEMO arm than the NIVO+CHEMO arm (Table 8), with a higher proportion of events observed, a lower median, and lower rates of treatment at all times. The Kaplan-Meier profile (Figure 58) demonstrated this separation until a protocol-mandated treatment discontinuation at 24 months. The Kaplan-Meier curve was visibly almost complete, and so in the first instance direct representation of this in the economic model without extrapolation was preferred.

In consideration of parametric models, Figure 59 demonstrated that a number of distributions within the exponential family could be acceptable, excepting that they deviate dramatically in the tail. Panel (b), with multiple crossings of the regression line, indicated that the exponential model may be adequate for the majority of the data, but the deviation after 24 months for both arms was unacceptable. Panel (c) also demonstrated this, though the effect was visually unweighted due to the logarithmic timescale; similarly, a collapse in the proportional hazards between the two arms occurred in this period and indicated against a scaling rule being used, particularly in the tail/extrapolation. Panel (d), demonstrated an approximation of the hazard function as almost constant, followed by a dramatic variation after 24 months, which was unlikely to be represented well by a constantly increasing or decreasing hazard function given by the Gompertz distribution. The curvature seen in panels (e) and (f) indicated against use of the log-logistic or lognormal distribution.

The smoothed hazard profiles shown in [redacted]Figure 60 were clearly affected by the discontinuity in hazard implied by the stopping rule on the NIVO+CHEMO arm, making their interpretation difficult. In contrast to PFS, hazard started near its peak, not ruling out monotonic hazard functions. The NIVO+CHEMO arm could be approximated by a constant hazard profile, but this did not seem to be the case with the CHEMO arm, where a decreasing marginal hazard was demonstrated.

Table 8: Observed TTD, CheckMate 649

Endpoint	NIVO+CHEMO (N=473)	CHEMO (N=482)
TTD		
Events, N	[redacted]	[redacted]
Median (months) (95% CI)	[redacted]	[redacted]
Landmark survival (%) (95% CI)		
6 months	[redacted]	[redacted]
12 months	[redacted]	[redacted]
18 months	[redacted]	[redacted]
24 months	[redacted]	[redacted]
36 months		

CI: Confidence interval; TTD: time to treatment discontinuation
Survival per Kaplan-Meier estimator, confidence interval by Greenwood estimate of variance upon log-log survival

[redacted]

Figure 58. TTD: Kaplan-Meier

TTD: time to treatment discontinuation

Figure 59. TTD: Ishak diagnostic plots

TTD: time to treatment discontinuation. Source: Ishak et al. (2013)[24]

Figure 60. TTD: Smoothed hazard function estimates

R-P: Royston-Parmar. Confidence interval is shown around b-spline estimator

4.5.2. Standard statistical models (relative survival framework)

Standard statistical models were fitted to the data to provide extrapolative models for scenarios where the Kaplan-Meier estimator was not to be used directly.

For NIVO+CHEMO (Figure 61, Figure 62), as expected from the non-parametric data inspection, the log-logistic and lognormal models fitted very poorly to the data, overestimating substantially from approximately 12 months onwards. The other models all performed similarly, and given the completeness of the data, did not deviate substantially in their extrapolation.

The same features were noted in the CHEMO arm (Figure 63, Figure 64), with the exception that as well as the log-logistic and log-normal models fitting poorly after 12 months, none of the models could demonstrate the change of hazard occurring at the end of the first year, resulting in overestimation at 12 months and subsequent underestimation. To attempt to reduce this error, a piecewise model was fitted, as for PFS.

Figure 61. TTD, NIVO+CHEMO: Standard statistical models overlaid upon Kaplan-Meier

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; LTM: Lifetable mortality; TTD: time to treatment discontinuation. 95% confidence intervals obtained by data bootstrap (1000 repetitions)

Figure 62. TTD, NIVO+CHEMO: Hazard profile of parametric relative survival models

LTM: Lifetable mortality; TTD: time to treatment discontinuation

Figure 63. TTD, CHEMO: Standard statistical models overlaid upon Kaplan-Meier

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; LTM: Lifetable mortality; TTD: time to treatment discontinuation.95% confidence intervals obtained by data bootstrap (1000 repetitions)

Figure 64. TTD, CHEMO: Hazard profile of parametric relative survival models

LTM: Lifetable mortality; TTD: time to treatment discontinuation

4.5.3. Semi-parametric models

As for the investigator-assessed PFS, a cut time of 6.44 months was used as this represented a good compromise between use of the data to inform the model parameters and the fit to features specific to later times.

For NIVO+CHEMO, fit for all models for the period prior to 24 months was improved, but behaviour after this time varied (Figure 65, Figure 66). The log-logistic and lognormal models showed very substantial overestimation after the 24 month stopping rule, whereas the other models effectively smoothed the step and intercepted the Kaplan-Meier to describe the patients remaining on chemotherapy alone until end of follow-up.

For CHEMO, the log-logistic, lognormal and Gompertz models were well fitted to the data (Figure 67, Figure 68). The lognormal model showed higher AIC and BIC than the other two models, but the cause of this was not immediately apparent as it did not result in survival predictions that deviated substantially from those of the log-logistic model within trial follow-up, and so it was considered feasible. The generalised gamma did not demonstrate the expected curvature in cumulative hazard in the period around 24 months and so was not recommended.

Figure 65. TTD, NIVO+CHEMO: Semi-parametric (piecewise) relative survival models overlaid upon Kaplan-Meier

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; LTM: Lifetable mortality; TTD: time to treatment discontinuation.. 95% confidence intervals obtained by data bootstrap (1000 repetitions)

Figure 66. TTD, NIVO+CHEMO: hazard profiles of semi-parametric (piecewise) relative survival models

LTM: Lifetable mortality; TTD: time to treatment discontinuation

Figure 67. TTD, CHEMO: Semi-parametric (piecewise) relative survival models overlaid upon Kaplan-Meier

AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion; LTM: Lifetable mortality; TTD: time to treatment discontinuation.. 95% confidence intervals obtained by data bootstrap (1000 repetitions)

Figure 68. TTD, CHEMO: hazard profiles of semi-parametric (piecewise) relative survival models

LTM: Lifetable mortality; TTD: time to treatment discontinuation

4.5.4. Model selection algorithm

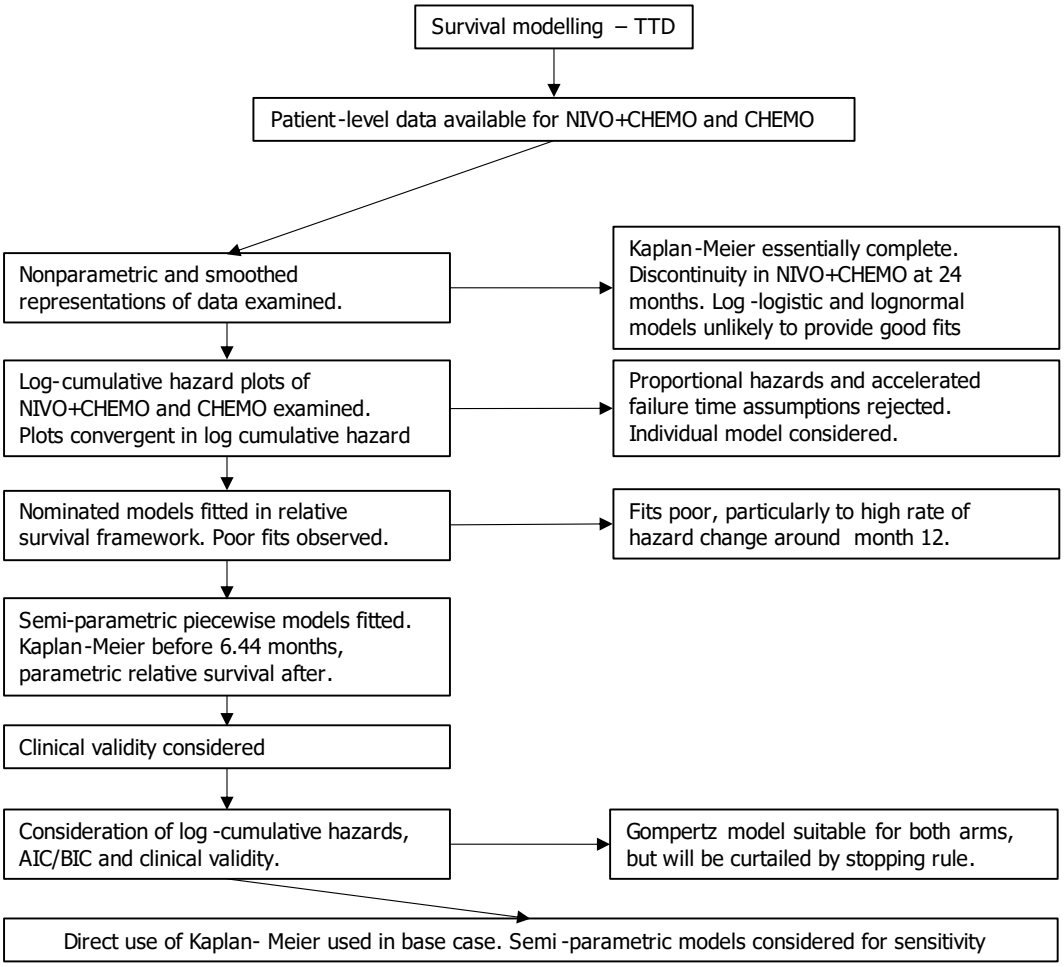

Figure 69. TTD: Considerations and assessments made when modelling

AIC: Akaike Information Criterion BIC: Bayesian Information Criterion; PFS: progression-free survival; TTD: time to treatment discontinuation

4.5.5. Conclusions on TTD

- Kaplan-Meier almost complete, to be used in base case
- Standard statistical models are a poor fit due to the inability to model rapid decrease in hazard ~ 12 months
- Semi-parametric models can fit well, but different models fit well to each arm and will require modifying by a stopping rule
- Direct use of Kaplan-Meier is base case

4.6. Ratio of transitions from pre-progression to death versus to post-progression

In order to incorporate relative treatment effects in the semi-Markov economic model, in the form of hazard ratios upon PFS and OS, direct modelling of PFS was undertaken, such that this model could be scaled in its entirety. The PFS state occupancy functions imply a rate of state exit, which is split into two components – those transitions to death versus those transitions to the post-progression state. The ratio between these two components was determined by logistic regression upon the PFS events.

Table 9: PFS events that are deaths, CheckMate 649

Endpoint	NIVO+CHEMO	CHEMO
PFS events that are deaths		
	Death / PFS events (%)	Death / PFS events (%)
Investigator-assessed	[redacted]	[redacted]
Investigator-assessed (not censoring for subsequent therapies)	[redacted]	[redacted]
BICR-assessed	[redacted]	[redacted]
BICR-assessed (not censoring for subsequent therapies)	[redacted]	[redacted]
BICR: Blinded independent central review; PFS: progression-free survival		

Table 9 shows the number of PFS events that were deaths for each arm. A large number of deaths occurred prior to progression when the patient had commenced a subsequent therapy.

A number of different logistic models were proposed to describe the time-profile of the ratio of death to progression events.

Table 10: Death upon progression logistic models: Progression per investigator, NIVO+CHEMO

Model	AIC	Deviance	DF of residual	Intercept		time		sqrt(time)		time * time		time*time*time		log(time)	
				Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE
intercept only															
time + intercept															
sqrt(time) + intercept															
time*time + intercept															
time*time*time + intercept															
log(time)															
time + sqrt(time) + intercept															
time + time*time + intercept															
time + time*time*time + intercept															
time + log(time) + intercept															
time + sqrt(time) + time*time + intercept															
time + log(time) + time*time + intercept															
time + sqrt(time) + log(time) + intercept															

Table 11: Death upon progression logistic models: Progression per investigator, CHEMO

Model	AIC	Deviance	DF of residual	Intercept		time		sqrt(time)		time * time		time*time*time		log(time)	
				Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE
intercept only	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]										
time + intercept	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]								
sqrt(time) + intercept	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]			[redacted]	[redacted]						
time*time + intercept	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]					[redacted]	[redacted]				
time*time*time + intercept	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]							[redacted]	[redacted]		
log(time)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]									[redacted]	[redacted]
time + sqrt(time) + intercept	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]						
time + time*time + intercept	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]			[redacted]	[redacted]				
time + time*time*time + intercept	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]					[redacted]	[redacted]		
time + log(time) + intercept	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]							[redacted]	[redacted]
time + sqrt(time) + time*time + intercept	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]				
time + log(time) + time*time + intercept	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]			[redacted]	[redacted]			[redacted]	[redacted]
time + sqrt(time) + log(time) + intercept	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]					[redacted]	[redacted]

[redacted]

Figure 70: Predictions of death upon progression model, time + log(time) + intercept

Thin lines – logistic model predictions. Dashed lines – smoothing generalised additive models

The ratio of progression to death events exhibited a U-shaped profile over time from randomisation (Figure 70), with an initial peak of deaths being superseded by a high rate of progression, followed by a long-term decrease in the proportion to PFS events due to progression. Logistic models involving log time performed well per AIC (Table 10, Table 11) and were selected to inform the economic model.

5. Discussion

Models of time to event were formed based upon data from CheckMate 649 in order to inform transition intensities of a semi-Markov state transition cost-utility model. This required survival analysis upon three outcomes: PFS, PPS and TTD.

PFS demonstrated a stepped profile due to fixed assessment times and had a non-monotonic hazard profile. It was unclear whether the excess hazard would decrease to zero over time for either arm, but it was possible for a small fraction of patients. The marginal hazard decreased rapidly once past the peak, which caused poor fitting for fully parametric models. Delaying the start of the parametric model alleviated this issue.

PPS demonstrated a higher initial hazard than PFS for both arms, and was more similar between the arms than PFS, but still demonstrated a benefit for the NIVO+CHEMO arm. The decreasing marginal hazard was rapid at some points, particularly in the CHEMO arm, but a log-logistic model was adequate for this transition; the CHEMO arm may be slightly overestimated.

TTD is almost complete and can be represented directly by the Kaplan-Meier estimator without extrapolation.

For PFS, there is some evidence of a small "plateau" for both arms. Identification of a cure fraction responsible for this is dependent upon the distribution of survival time of the at-risk, but the possibility for long-term remission, particularly after receiving nivolumab, should be considered.

Similarly, a small fraction of patients may express extended PPS. Given that the censoring distribution of PPS is non-random with respect to time from randomisation to progression, if patients who are slower to progress are more likely to have extended PPS, this fraction may become more significant over time.

6. Conclusion

Data from the CheckMate 649 trial was analysed to produce models of survival outcomes to a lifetime horizon for use in cost-utility modelling. Semi-parametric piecewise models were chosen for PFS outcomes on both arms due to the high rate of change of hazard over the initial period after commencing treatment, followed by a more consistent period. Fully parametric models of PPS were suitable for extrapolation, though may favour the CHEMO arm. The same distribution was

recommended for both arms due to similarity of nonparametric estimates of the survival function, but scaling rules were rejected.

- For investigator-assessed PFS, a semi-parametric model utilising Kaplan-Meier data to month 6.44 and a lognormal relative survival model thereafter was chosen for both arms.
- For OS, a semi-parametric model utilising Kaplan-Meier data to month 6.44 and a Gompertz relative survival model thereafter was chosen for both arms.
- For PPS, a fully-parametric model utilising a log-logistic relative survival model was chosen for both arms.
- For TTD, direct use of the Kaplan-Meier estimate was chosen for both arms.

Given the clinical expectation that a small number of responsive patients may enter long-term remission after treatment with nivolumab, the long-term survival benefit produced by this semi-parametric modelling approach may be conservative. The ratio of PFS events due to death versus progression was described by a logistic model dependent upon time and log time. The PPS was calibrated to PFS and OS hazard ratios derived from an NMA.

7. References

1. Bristol-Myers Squibb. Clinical Protocol CA209649. A Randomized, Multicenter, Open-Label, Phase 3 Study of Nivolumab plus Ipilimumab or Nivolumab in Combination with Oxaliplatin plus Fluoropyrimidine versus Oxaliplatin plus Fluoropyrimidine in Subjects with Previously Untreated Advanced or Metastatic Gastric or Gastroesophageal Junction Cancer. (CheckMate 649: CHECKpoint pathway and nivoluMab clinical Trial Evaluation 649). 2018.
2. Bristol-Myers Squibb. CA209649 Nivolumab+Chemotherapy Primary Clinical Study Report. 2020.
3. Chen T-T. Statistical issues and challenges in immuno-oncology. Journal for ImmunoTherapy of Cancer. 2013;1(1):18.
4. Schadendorf D, Hodi FS, Robert C, 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. 2015;33(17):1889-94.
5. Antonia SJ, Borghaei H, Ramalingam SS, et al. Four-year survival with nivolumab in patients with previously treated advanced non-small-cell lung cancer: a pooled analysis. The Lancet Oncology. 2019;20(10):1395-408.
6. Ascierto PA, Long GV, Robert C, et al. Survival Outcomes in Patients With Previously Untreated BRAF Wild-Type Advanced Melanoma Treated With Nivolumab Therapy: Three-Year Follow-up of a Randomized Phase 3 Trial. JAMA oncology. 2019;5(2):187-94.
7. Wolchok JD, Weber JS, Maio M, et al. Four-year survival rates for patients with metastatic melanoma who received ipilimumab in phase II clinical trials. Annals of oncology : official journal of the European Society for Medical Oncology. 2013;24(8):2174-80.
8. Restifo N, Smyth M, Snyder A. Acquired resistance to immunotherapy and future challenges. Nature Reviews Cancer. 2016;16:121-6.
9. Rutherford MJ, Lambert PC, Sweeting MJ, et al. NICE DSU TECHNICAL SUPPORT DOCUMENT 21: Flexible Methods for Survival Analysis. 2020. Available at: http://nicedsu.org.uk/wp-content/uploads/2020/11/NICE-DSU-Flex-Surv-TSD-21_Final_alt_text.pdf [Accessed
10. 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.
11. 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. Available at: http://nicedsu.org.uk/wp-content/uploads/2016/03/NICE-DSU-TSD-Survival-analysis.updated-March-2013.v2.pdf [Accessed September 2020].
12. INDEC: Instituto Nacional de Estadistica y Censos. Tablas abreviadas de mortalidad por sexo y edad 2008-2010. : total del país y provincias. 2013. Available at: https://www.indec.gob.ar/ftp/cuadros/publicaciones/tablas_mortalidad_2008_2010.pdf [Accessed November 2020].
13. Australian Bureau of Statistics. Statistics about life tables for Australia, states and territories and life expectancy at birth estimates for sub-state regions. 2019. Available at: https://www.abs.gov.au/statistics/people/population/life-tables/2016-2018#data-download [Accessed November 2020].
14. Instituto Brasileiro de Geografia e Estatística I. TÁBUA COMPLETA DE MORTALIDADE 2017. 2017. Available at: https://www.lifetable.de/data/BRA/BRA000020172017CU1.pdf [Accessed November 2020].
15. Statistics Canada. Life Tables, Canada, Provinces and Territories 1980/1982 to 2016/2019. 2019. Available at: https://www150.statcan.gc.ca/n1/pub/84-537-x/84-537-x2018002-eng.htm [Accessed November 2020].
16. University of California BU, and Max Planck Institute for Demographic Research (Germany),. Human Mortality Database. Available at: https://www.mortality.org/ [Accessed November 2020].
17. World Health Organization. WHO Global Health Observatory data repository,. Available at: https://apps.who.int/ [Accessed November 2020].

18. Economic Commission for Latin America and the Caribbean (ECLAC). Life Tables. Demographic Observatory, Latin America and the Caribbean (LC/PUB.2017/20-P). Santiago: United Nations. 2017. Available at: https://www.lifetable.de/ [Accessed November 2020].
19. Statistics Singapore. Complete Life Tables for Singapore Resident Population, 2017-2018. 2019. Available at: www.singstat.gov.sg [Accessed November 2020].
20. Instituto Nacional de Estadistica. Tablas de mortalidad de la población de España Available at: https://www.ine.es/dyngs/INEbase/operacion.htm?c=Estadistica_C&cid=1254736177004&menu=resultados&secc=1254736195384&idp=1254735573002 [Accessed November 2020].
21. Turkish Statistical Institute. Single age life table for Turkey by sex, 2016-2018. Available at: www.turkstat.gov.tr [Accessed October 2020].
22. Office for National Statistics. National life tables: UK (2018). 2020. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesunitedkingdomreferencetables/current [Accessed November 2020].
23. CDC National Center for Health Statistics: National Vital Statistics System. United States Life Tables 2017. 2019. Available from: https://www.cdc.gov/nchs/nvss/life-expectancy.htm#data [accessed November 2020].
24. Ishak KJ, Kreif N, Benedict A, et al. Overview of parametric survival analysis for health-economic applications. Pharmacoeconomics. 2013;31(8):663-75.
25. Royston P, Parmar MK. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Stat Med. 2002;21(15):2175-97.
26. Jackson CH, Sharples LD, Thompson SG. Survival models in health economic evaluations: balancing fit and parsimony to improve prediction. Int J Biostat. 2010;6(1):Article 34.
27. Gelber RD, Goldhirsch A, Cole BF, et al. Parametric extrapolation of survival estimates with applications to quality of life evaluation of treatments. Controlled clinical trials. 1993;14(6):485-99.

NICE National Institute for Health and Care Excellence

Nivolumab with platinum- and fluoropyrimidine-based chemotherapy for untreated HER2-negative advanced gastric, gastro-oesophageal junction or oesophageal adenocarcinoma [ID1465]

Consultation on the appraisal consultation document – deadline for comments 5pm on Thursday 11 November 2021. Please submit via NICE Docs.

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

Please return to: **NICE DOCS**

NICE National Institute for Health and Care Excellence

Nivolumab with platinum- and fluoropyrimidine-based chemotherapy for untreated HER2-negative advanced gastric, gastro-oesophageal junction or oesophageal adenocarcinoma [ID1465]

Consultation on the appraisal consultation document – deadline for comments 5pm on Thursday 11 November 2021. Please submit via NICE Docs.

number	Insert each comment in a new row. Do not paste other tables into this table, because your comments could get lost – type directly into this table.
Example 1	We are concerned that this recommendation may imply that ……..
1	30 months is too short a time for many patients to be considered cured, 36 – 48 months is seen within our patient support groups as more appropriate but is based on lived experiences not clinical data.
2	Agree that cured of cancer is different to being cured with same risk of dying of the general population, the effect of treatment alone puts them in a different group of population. *(Although not part of this appraisal this new evolving group of cancer surviving patients will be of a concern to the NHS in the near future)*
3	I am concerned that younger patients are being diagnosed more now than ever before and that setting the mean age of 64.15 years does not show the true age, younger patients although potentially fitter are more often diagnosed at a later stage and an effective treatment is lacking, but agree this treatment model is not suitable and that encouragement of a new model as suggested is appropriate and needed.
4	
5	
6	

Insert extra rows as needed

Checklist for submitting comments

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

Note: We reserve the right to summarise and edit comments received during consultations, or

Please return to: **NICE DOCS**

Nivolumab with platinum- and fluoropyrimidine-based chemotherapy for untreated HER2-negative advanced gastric, gastro-oesophageal junction or oesophageal adenocarcinoma [ID1465]

Consultation on the appraisal consultation document – deadline for comments 5pm on Thursday 11 November 2021. Please submit via NICE Docs.

not to publish them at all, if we consider the comments are too long, or publication would be unlawful or otherwise inappropriate.

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

Comments on the ACD received from the public through the NICE Website

Name	
Organisation	NICE Medicines Optimisation Team
Location	
Conflict	None
Notes	
Comments on the ACD:	
Section 1.2 Under "Why the committee made these recommendations" 2nd paragraph last sentence, should "FELOX" say "FOLFOX"?	

ERG response to company response to ACD 1

LIVERPOOL REVIEWS AND IMPLEMENTATION GROUP (LRiG)

Nivolumab in combination with chemotherapy for untreated advanced gastric, gastro-oesophageal junction, or oesophageal adenocarcinoma [ID1465]

Confidential until published

This report was commissioned by the NIHR Evidence Synthesis Programme as project number 133505

Completed 17 January 2022

Contains ACADEMIC IN CONFIDENCE DATA and COMERCIAL IN CONFIDENCE DATA

UNIVERSITY OF LIVERPOOL | LIVERPOOL REVIEWS AND IMPLEMENTATION GROUP

A MEMBER OF THE RUSSELL GROUP

1 INTRODUCTION

As part of their response to the appraisal consultation document (ACD) for nivolumab (NIVO) in combination with chemotherapy for untreated advanced gastric, gastro-oesophageal junction, or oesophageal adenocarcinoma, the company generated updated cost effectiveness results using a partitioned survival model based on the latest clinical effectiveness data ([redacted]) available from the CheckMate 649 trial. The ERG has assessed this model and ERG conclusions are contained in this report.

The company also used the latest CheckMate 649 trial data to update the original company model; the original model was a semi-Markov model that the NICE Appraisal Committee (AC) considered was unsuitable for decision making during the first AC meeting. The semi-Markov model assumes that all patients who are in the progression-free survival (PFS) health state at 30 months are essentially cured, experiencing mortality rates equal to general population mortality rates. The company has not presented any compelling new evidence to support this claim; a claim which the NICE AC considered implausible. As such, the ERG considers that the semi-Markov model remains unsuitable for decision making; see previous ERG report for a critique of the original (semi-Markov) company model.

2 ERG CRITIQUE OF THE COMPANY PARTITIONED SURVIVAL MODEL

For the comparison of nivolumab+capecitabine+oxaliplatin (NIVO+XELOX) versus XELOX, the company provided a partitioned survival model with curves for overall survival (OS), PFS and time to treatment discontinuation (TTD) based on CheckMate 649 trial data from the latest database lock ([redacted]).

The ERG commends the company for providing full documentation and transparent descriptions of the process used to select OS, PFS and TTD curves. The ERG was unable to generate curves as the CheckMate 649 trial K-M data was not made available to them. However, the ERG considers that the options presented by the company allow selection of curves that are suitable for decision making.

2.1 Costs and utilities

At the end of the Technical Engagement period, the ERG was satisfied that the costs and utility values that had been applied in the company's original semi-Markov model were appropriate for decision making. In the new partitioned survival model, the included costs are identical to the costs used in the original semi-Markov model, except for the cost of nivolumab which was updated to reflect the new Patient Access Scheme (PAS) price. In the new partitioned survival model, the utility values from the original semi-Markov model were updated in line with the latest evidence from the CheckMate 649 trial. As the changes to costs and utilities were minimal, the ERG is satisfied that both costs and utilities included in the partitioned survival model are appropriate for decision making.

2.2 Progression-free survival

The company followed Decision Support Unit (DSU) guidance[1,2] on curve selection and concluded that a semi-parametric approach to modelling was the most appropriate; the company therefore used the first 6.44 months of the CheckMate 649 trial PFS Kaplan-Meier (K-M) data and then applied a parametric distribution. The 6.44 month cut-off point was chosen to reflect the fact that high frequency assessments, which could influence the timing of PFS measurements, had ceased. The ERG considers the justification for this cut-off point is acceptable and that the semi-parametric approach produced curves that fit the CheckMate 649 trial PFS K-M data better than curves produced using a fully parametric approach. The semi-parametric PFS curves considered by the company with associated Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values are reproduced from the company response to the ACD in

Figure 1 (NIVO+XELOX) and Figure 2 (XELOX).

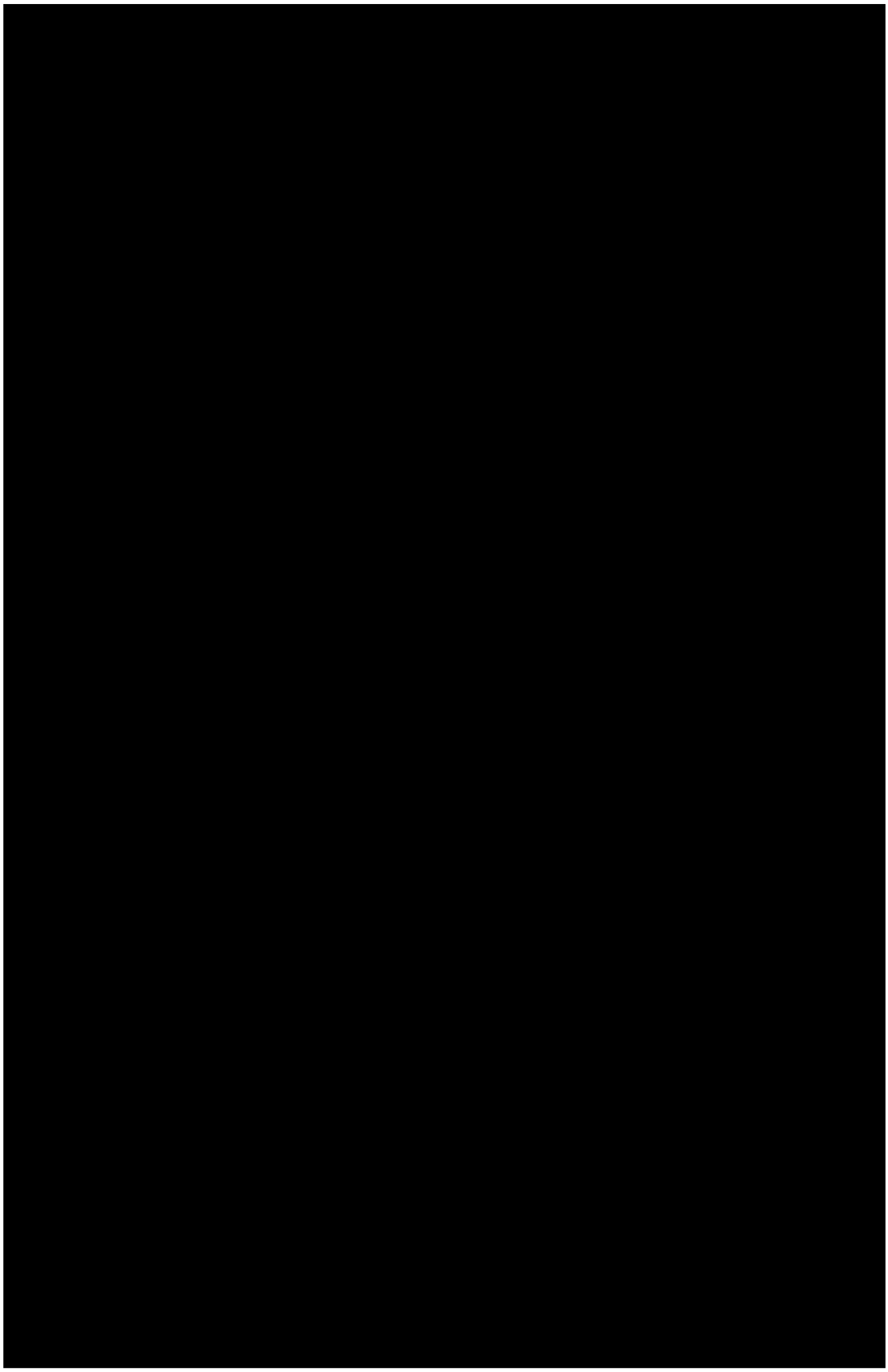

Figure 1 PFS (BICR), NIVO+XELOX CPS≥5%, semi-parametric models (cut at 6.44 months) considered by company

ACD=appraisal consultation document; BICR=blinded independent central review; CPS=combined positive score; PFS=progression-free survival
Source: Figure 12 in company response to ACD

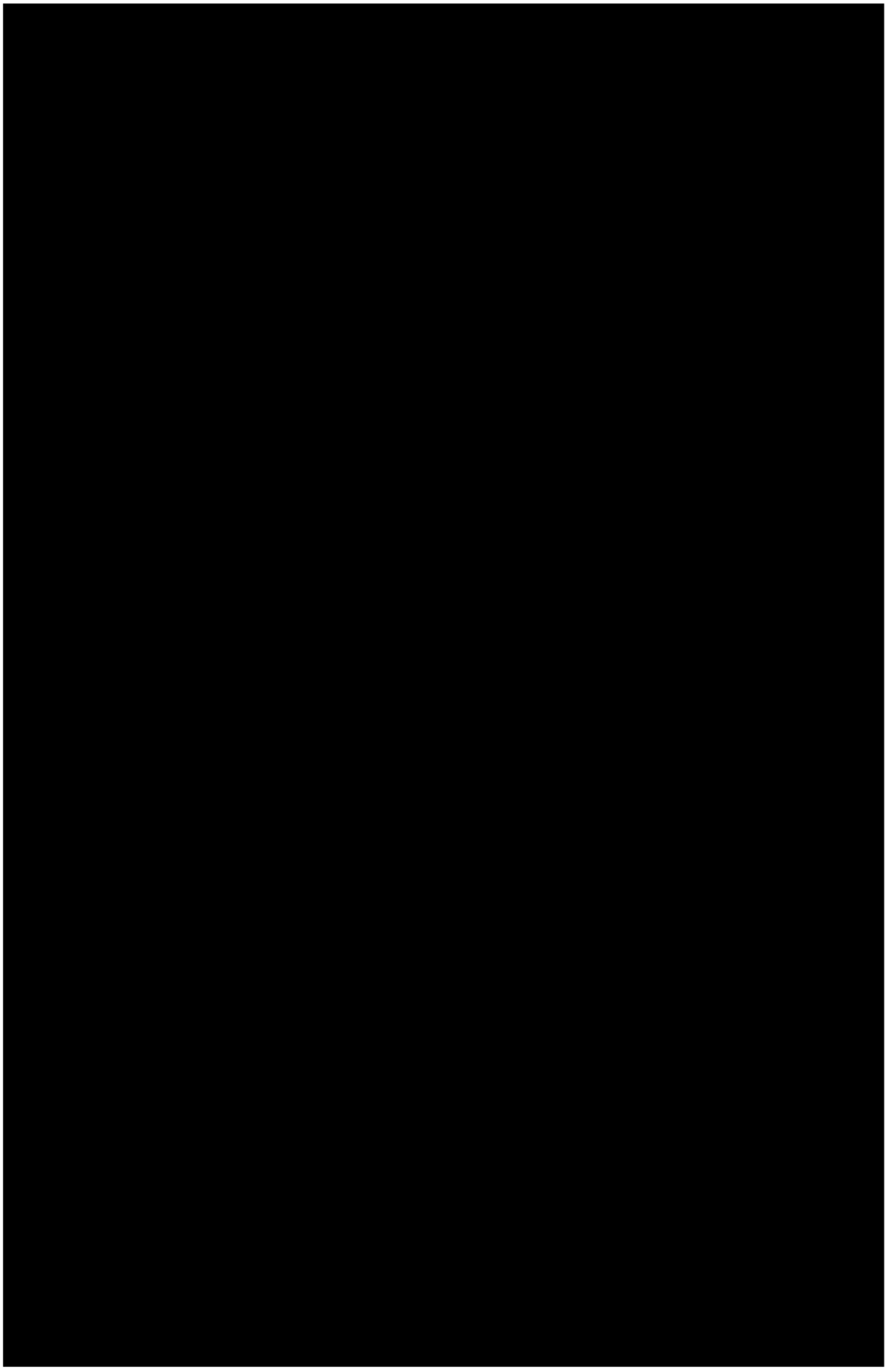

Figure 2 PFS (BICR), XELOX CPS≥5%, semi-parametric models (cut at 6.44 months) considered by company

ACD=appraisal consultation document; BICR=blinded independent central review; CPS=combined positive score; PFS=progression-free survival
Source: Figure 13 from company response to ACD

When interpreting the AIC/BIC statistics, it is important to consider that the statistics only provide information about the relative fit of the distributions to the available data. The statistics do not identify whether any single distribution is a good fit, nor can the statistics determine if a distribution has good predictive power. A rule of thumb for interpreting the relative goodness of fit of different distributions by their AIC score is that, if any distributions are within 2 points of each other, they are essentially indistinguishable, and that support remains for choosing a distribution that is up to 7 points higher than the distribution with the lowest AIC score.[3]

The company chose the log-normal distribution for use in the base case analysis for both NIVO+XELOX and XELOX as this distribution had the lowest AIC score. The log-logistic and generalised gamma distributions had AIC scores within 5 points of the log-normal distribution and all three curves are good visual fits to the available CheckMate 649 trial PFS K-M data. Choosing the log-logistic distribution decreases the company base case incremental cost effectiveness ratio (ICER) per quality adjusted life year (QALY) whilst the generalised-gamma distribution increases the size of the company base case ICER per QALY gained. The ERG considers that, as there is no other clinical evidence upon which to select a distribution for PFS outside of the CheckMate 649 trial PFS K-M data, it is appropriate to use the log-normal distribution in the base case.

2.3 Overall survival

The company has modelled OS using an innovative approach, i.e., the company has attempted to add the excess mortality of the population of interest to general population mortality. Such an approach is valid and the ERG considers would be a step forward in the modelling of OS in any disease area. However, in practice, the company has added all-cause mortality from the CheckMate 649 trial to all-cause general population mortality. This means that the model double counts some deaths. Further, the algorithms in the model that were used to add in the excess mortality experienced by the population of interest were not accurate. As evidence about excess mortality is not available, the ERG has corrected the company model so that OS directly follows the OS curves considered by the company and the mortality hazard can never fall below general population mortality. However, the ERG commends the company's novel modelling approach and would welcome its use in future appraisals if there are accurate data to support its adoption.

As was the case with PFS, the company followed DSU guidance[1,2] on curve selection and concluded that a semi-parametric approach was the most appropriate; the company used the first 6.44 months of the CheckMate 649 trial OS K-M data and then applied a parametric

distribution. The choice of a 6.44 month cut-off point was supported by the CheckMate 649 trial PFS data, but appears arbitrarily chosen for OS. However, the ERG considers that, as with PFS, the semi-parametric approach produces curves that fit the CheckMate 649 trial OS K-M data better than those produced using a fully parametric approach. The semi-parametric OS curves considered by the company with associated AIC/BIC values are reproduced from the company response to the ACD in

Figure 3 (NIVO+XELOX) and

Figure 4 (XELOX).

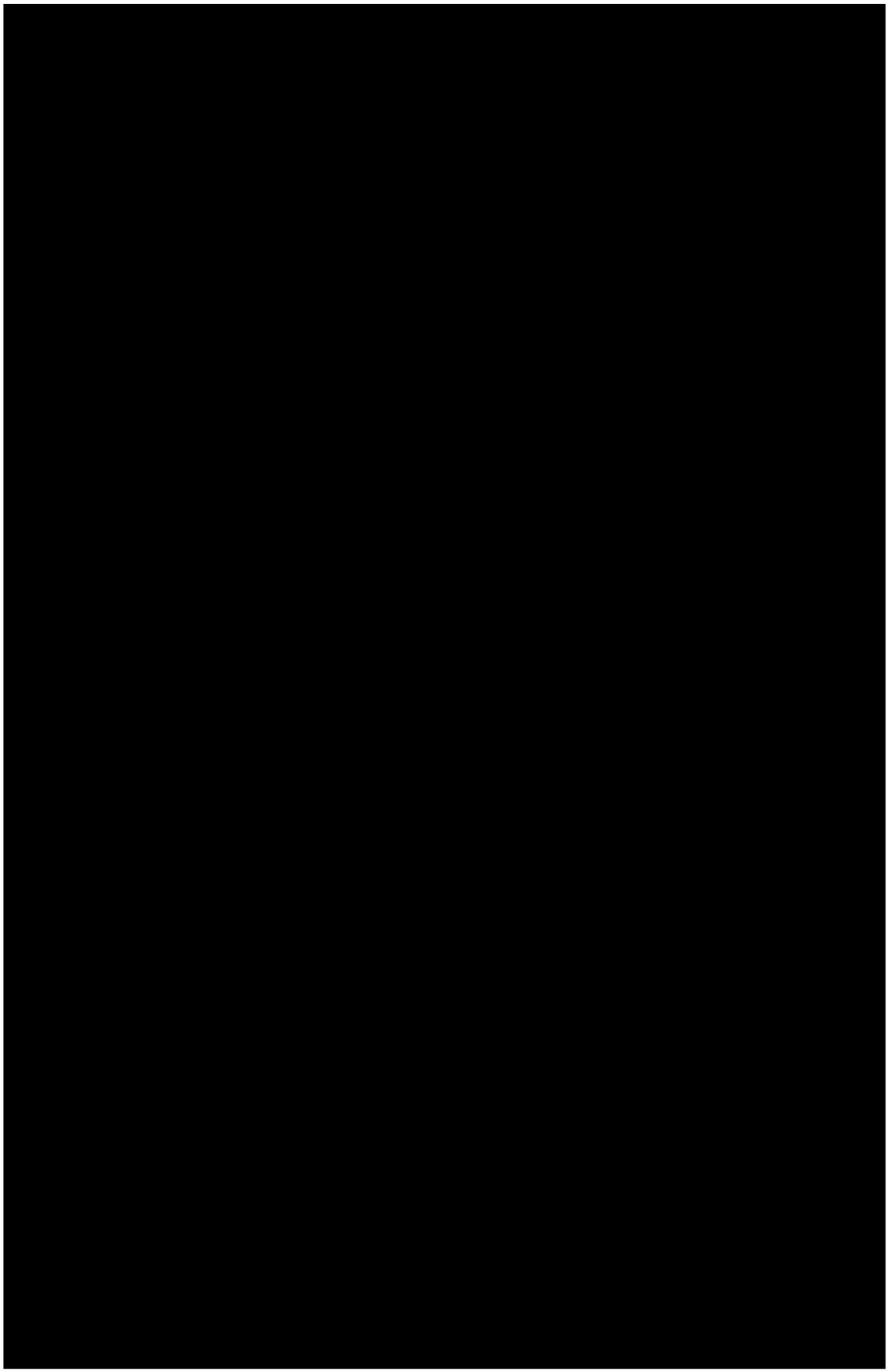

Figure 3 OS, NIVO+XELOX CPS≥5%, semi-parametric models (cut at 6.44 months) considered by company

ACD=appraisal consultation document; CPS=combined positive score; OS=overall survival
Source: Figure 7 from company response to ACD

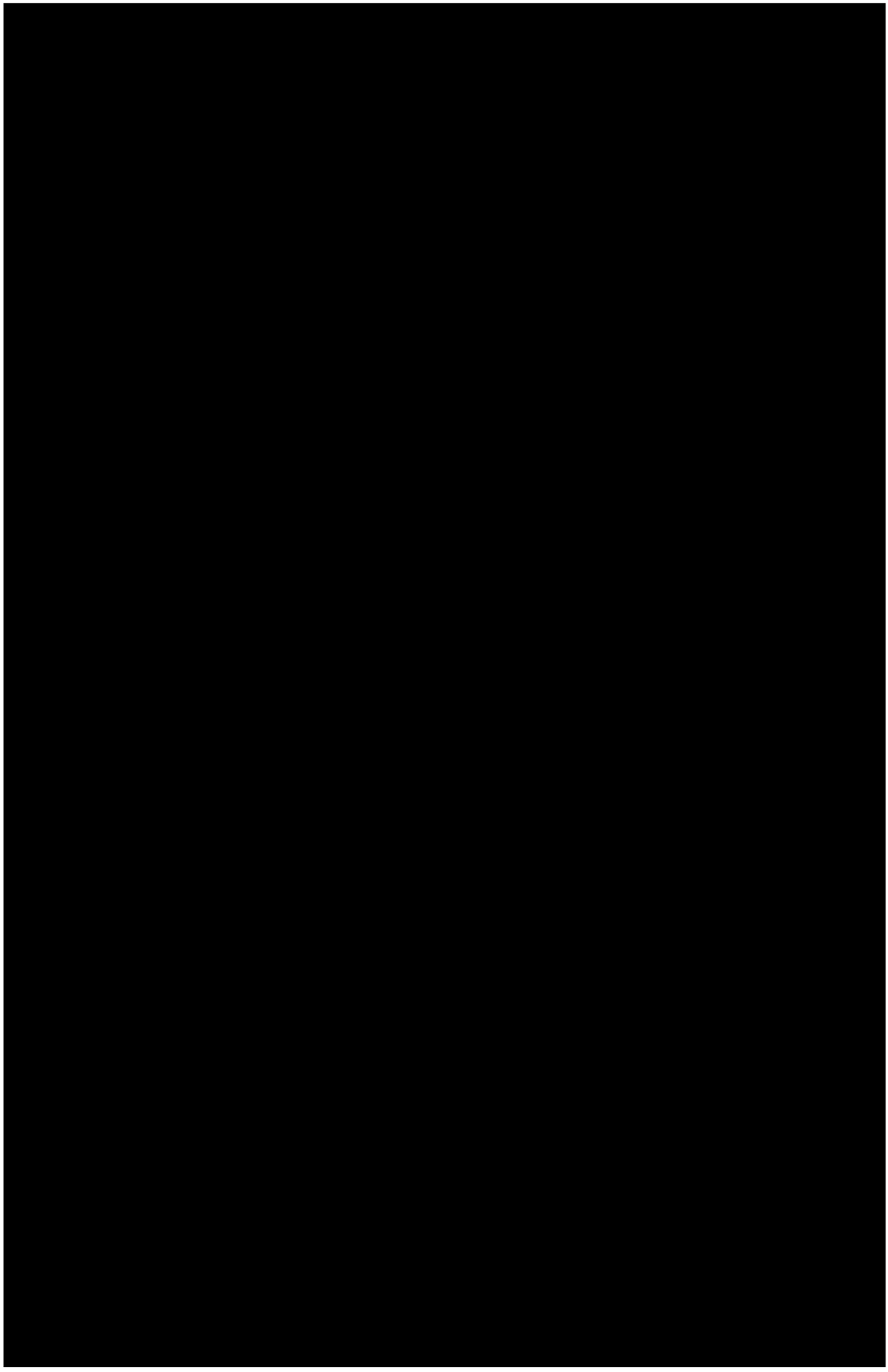

Figure 4 OS, XELOX CPS≥5%, semi-parametric models (cut at 6.44 months) considered by company

ACD=appraisal consultation document; CPS=combined positive score; OS=overall survival
Source: Figure 8 from company response to ACD

The company chose the Gompertz distribution in the base case analysis to model both NIVO+XELOX and XELOX as this distribution had the lowest AIC and BIC scores. The log-logistic and generalised gamma distributions had AIC scores within 5 points of the Gompertz distribution and so can be considered as potential alternative distributions. As shown in Table 1 and Table 2, at the end of year 1, compared to the CheckMate 649 trial OS K-M data, the Gompertz, log-logistic and generalised gamma distributions accurately predict OS for NIVO+XELOX but slightly underestimate OS for XELOX, and overestimate survival at the end of year 2 and year 3 for both NIVO+XELOX and XELOX (with the exception of the generalised gamma for NIVO+XELOX which almost exactly matches the OS at year 3). Looking at OS at the end of year 1 and year 3, the log-logistic distribution appears to be a poor fit for the XELOX arm and at 3 years appears to be the poorest fit for both arms. The ERG therefore considers that the log-logistic distribution should not be considered as a suitable distribution over the Gompertz or generalised gamma distributions.

In contrast to the limited PFS data available, other OS clinical effectiveness data exists; this evidence can be used to guide curve selection. For XELOX, Royal Marsden Hospital[4] evidence previously presented and discussed in the original company submission suggests that, at 5 years, 4.0% of patients could be expected to be alive and receiving chemotherapy. Both the Gompertz and generalised gamma distributions produce estimates below 4.0%, suggesting that both may be generating conservative estimates of OS for patients treated with XELOX.

Table 1 NIVO+XELOX OS at key time points by potential distribution

Distribution	1-year survival	2-year survival	3-year survival	5-year survival	10-year survival	20-year survival
Gompertz	57.2%	32.5%	21.9%	13.6%	9.2%	5.9%
Log-logistic	57.0%	32.8%	22.4%	13.5%	6.5%	3.0%
Generalised gamma	57.2%	33.2%	21.5%	10.6%	2.8%	0.5%
CheckMate 649 trial OS K-M	57.0%	31.0%	21.6%	-	-	-

K-M=Kaplan-Meier; OS=overall survival
Source: company partitioned survival model corrected by ERG with adjustment to stop mortality hazard falling below general mortality hazard

Table 2 XELOX OS at key time points by potential distribution

Distribution	1-year survival	2-year survival	3-year survival	5-year survival	10-year survival	20-year survival
Gompertz	45.6%	19.4%	9.8%	3.8%	1.5%	0.9%
Log-logistic	45.1%	19.4%	11.3%	5.7%	2.3%	0.9%
Generalised gamma	45.8%	19.4%	9.5%	2.9%	0.3%	0.2%
CheckMate 649 trial OS K-M	46.4%	18.6%	9.0%	-	-	-
Royal Marsden	44%	16%	-	4.0%	-	-

K-M=Kaplan-Meier; OS=overall survival
Source: company partitioned survival model corrected by ERG with adjustment to stop mortality hazard falling below general mortality hazard

It is highlighted in the ACD that the NICE AC considered that it was unreasonable to expect that the mortality hazard of patients with Stage IV gastric cancer would ever fall below the general population mortality hazard. However, the nature of all the distributions considered by the company is that they have falling (or constant) mortality hazards over time and, as general population mortality increases over time, the two hazards will inevitably cross. As stated, the ERG has fixed the distributions so that when the hazards cross, mortality hazards equalise and follow general population mortality hazards. Given the NICE AC's position that excess mortality for patients with Stage IV gastric cancer will never fall to zero, this makes all curves considered inaccurate. Analysis of (i) when mortality reaches general population mortality, (ii) the percentage of patients who reach this point and (iii) the percentage of patients still in PFS at this point allows consideration of how large this inaccuracy is likely to be and the impact it is likely to have on the size of the ICERs per QALY gained. This analysis is shown in Table 3 and

Table 4 for the Gompertz and generalised gamma semi-parametric distributions respectively.

Table 3 NIVO+XELOX mortality hazard analysis

Distribution	Time for mortality hazard to equal general population	Percentage alive when mortality hazard equals general population	PFS when mortality hazard equals general population
Gompertz	10.3 years	9.1%	4.1%
Generalised gamma	23.9 years	0.3%	0.1%

Source: company partitioned survival model corrected by ERG

Table 4 Table 4 XELOX mortality hazard analysis

Distribution	Time for mortality hazard to equal general population	Percentage alive when mortality hazard equals general population	PFS when mortality hazard equals general population
Gompertz	12.9 years	1.3%	1.2%
Generalised gamma	26.1 years	<0.0%	<0.0%

Source: company partitioned survival model corrected by ERG

As shown in Table 3 and

Table 4, for patients receiving NIVO+XELOX or XELOX, the Gompertz distribution predicts that the mortality hazard will reach background mortality earlier and affect a much higher percentage of patients than predicted by the generalised gamma distribution. For patients receiving NIVO+XELOX, the Gompertz distribution predicts 9.1% of patients will experience only general population mortality from 10.3 years (effectively cured), even although 54.9% of these patients had disease progression. If the Gopertz distribution is used, a significant percentage of patients have no excess mortality relatively early in the model time horizon; this means that the Gompertz distribution is unsuitable as it generates data that are in direct conflict with the NICE AC's view that patients would always experience excess mortality. In contrast, using the generalised gamma distribution means that only 0.3% of patients treated with NIVO+XELOX at 23.9 years, and almost no patients treated with XELOX at 26.1 years, lose the excess mortality from having Stage IV gastric cancer. The ERG therefore considers that whilst the evidence from the Royal Marsden Hospital[4] suggests that the generalised gamma distribution may slightly underestimate survival at 5 years for patients treated with XELOX, it is the generalised gamma distribution that, on balance, should be used in the base case analysis to generate OS estimates for both the NIVO+XELOX and XELOX arms.

2.4 Treatment effect waning

There is a 2-year stopping rule in the CheckMate 649 trial and in the SmPC[5] for nivolumab. The ERG considers that a scenario should be explored whereby any treatment effect from NIVO+XELOX compared to XELOX (i.e., where there is a lower mortality hazard for those

treated with NIVO+XELOX than those treated with XELOX) is not maintained for life. In line with previous submissions for nivolumab, the ERG has produced a scenario whereby the mortality hazard for those treated with NIVO+XELOX is equal to that of those treated with XELOX at 5 years (i.e., 3 years after treatment with nivolumab has stopped for all patients). The ERG highlights that this scenario is not evidence-based and as such does not form part of the ERG preferred base case.

2.5 Time to treatment discontinuation

The company has costed time to treatment discontinuation using CheckMate 649 trial TTD K-M data directly, i.e., without curve fitting. Given the completeness of the CheckMate 649 trial TTD K-M data, the ERG considers that this is an appropriate approach. The ERG notes that despite a clear drop in the number of patients using nivolumab at around 2 years, a small percentage of patients in the CheckMate 649 trial (approximately 5%, but there is limited information due to censoring) continued to receive nivolumab beyond 2 years, potentially for a further 6 months. The costs of these patients are included in the model and as any benefit to these patients in terms of prolonged PFS/OS is also included in the model, the ERG considers that removing these costs would be inappropriate. The ERG highlights that the inclusion of the costs/benefits for a small (<6 months) additional time on treatment with nivolumab for a small percentage of patients (<5%) is unlikely to influence cost effectiveness results.

2.6 ERG preferred ICERs per QALY gained

The ERG has corrected the company base case (Gompertz distribution for OS) by removing the attempts the company made to adjust for general population mortality. The corrected company base case generates an ICER of £41,738 per QALY gained for the comparison of NIVO+XELOX versus XELOX. In the scenario where treatment effect waning at 5 years is applied to the corrected company base case, the ICER for this comparison is £49,840 per QALY gained.

The ERG preferred base case generates an ICER of £58,816 per QALY gained for the comparison of NIVO+XELOX versus XELOX; these results are based on removal of adjustments for general population mortality and using a semi-parametric generalised gamma distribution for OS (K-M data cut-off: 6.44 months). In the scenario where treatment effect waning at 5 years is also applied to the ERG preferred base company base case, the ICER for this comparison is £70,681 per QALY gained.

Company base case and ERG cost effectiveness results following amendments to the company model are provided in Table 5.The instructions for implementing all ERG changes to the company model are provided in the Appendix.

Table 5 ERG revisions to company model and preferred ICER per QALY gained, PD-L1 CPS ≥5: NIVO+XELOX vs XELOX (PAS price for nivolumab, list prices for other drugs)

Scenario/ERG amendment	NIVO+XELOX			XELOX			Incremental			ICER	
	Cost	QALYs	Life Years	Cost	QALYs	Life years	Cost	QALYs	Life years	£/QALY gained	Change from base case
Company base case (Gompertz distribution for OS)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	***£45,383***	
ERG corrected company base case (semi-parametric Gompertz distribution for OS) without treatment effect waning	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£41,738	-£3,645
ERG corrected company base case (semi-parametric Gompertz distribution for OS) with treatment effect waning	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	£49,840	+£4,457
ERG preferred scenario: corrected company base case (semi-parametric generalised gamma distribution for OS) without treatment waning	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	***£58,816***	***+£13,433***
ERG preferred scenario: corrected company base case (semi-parametric generalised gamma distribution for OS) with treatment waning	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	***£70,681***	***+£25,298***

CPS=combined positive score; ERG=Evidence Review Group; ICER=incremental cost effectiveness ratio; NIVO=nivolumab; OS=overall survival; PAS=Patient Access Scheme; PD-L1= programmed cell death ligand 1; QALY=quality adjusted life year; XELOX=capecitabine+oxaliplatin

3 REFERENCES

1. Latimer N. Technical support document 14: Survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. Published March 2013; Available from: www.nicedsu.org.uk/wp-content/uploads/2016/03/NICE-DSU-TSD-Survival-analysis.updated-March-2013.v2.pdf. Accessed 14 January 2022.
2. Rutherford MJ, Lambert PC, Sweeting MJ, Pennington B, Crowther MJ, Abrams KR, *et al.* NICE DSU technical support document 21: Flexible methods for survival analysis. Published 23 January 2020; Available from: www.nicedsu.org.uk/wp-content/uploads/2020/11/NICE-DSU-Flex-Surv-TSD-21_Final_alt_text.pdf. Accessed 14 January 2022.
3. Burnham KP, Anderson DR. AIC differences. Model Selection and Inference: A Practical Information-Theoretical Approach New York: Springer-Verlag New York; 2002. p. 70-2.
4. Davidson M, Cafferkey C, Goode EF, Kouvelakis K, Hughes D, Reguera P, *et al.* Survival in advanced esophagogastric adenocarcinoma improves with use of multiple lines of therapy: Results from an analysis of more than 500 patients. Clin Colorectal Cancer. 2018; 17(3):223-30.
5. Bristol-Myers Squibb. Summary of Product Characteristics. OPDIVO 10 mg/mL concentrate for solution for infusion. 2020; Available from: https://www.medicines.org.uk/emc/medicine/30476. Accessed 10 November 2020.

4 APPENDIX

4.1 Microsoft Excel revisions made by the ERG to the company's partitioned survival model

Sheet	Cell	Instructions
"Survival used"	J24	=(I23-I24)/I23 Copy cell formula Paste to range J25:J1354
	K25	=IF(J25<'ACM Calculation'!Q13,1,0) Copy cell formula Paste to range K25:K1354
	R24	=(Q23-Q24)/Q23 Copy cell formula Paste to range R25:R1354
	S25	=IF(R25<'ACM Calculation'!Q13,1,0) Copy cell formula Paste to range S25:S1354
"ACM Calculation"	Q12	=((O11*dblBaseMaleUsed+(1-dblBaseMaleUsed)*P11)-(O12*dblBaseMaleUsed+(1-dblBaseMaleUsed)*P12))/(O11*dblBaseMaleUsed+(1-dblBaseMaleUsed)*P11) Copy cell formula Paste to range Q13:S1342
"Treatment Trace"	X11	=IF(Treatment_Waning=0,IF(F11=1,intCohort*MIN('Survival Used'!I23,1),0),intCohort)
	X12	=IF(F12=1,IF(Treatment_Waning=0,IF(F11=1,IF('Survival Used'!K24=1,X11*(1-O12),(F12*intCohort*MIN('Survival Used'!I24,1))),0),IF(H12>Waningyear,X11*(1-(('Control Trace'!X11-'Control Trace'!X12)/'Control Trace'!X11)),IF(F11=1,IF('Survival Used'!K24=1,X11*(1-O12),(F12*intCohort*MIN('Survival Used'!I24,1))),0))),0) Copy cell formula Paste to range X13:X1342
	P11	=MIN(F11*intCohort*'Survival Used'!E23,X11) Copy cell formula Paste to range P12:X1342
"Control Trace"	X11	=IF(F11=1,intCohort*MIN('Survival Used'!Q23,1),0)
	X12	=IF(F11=1,IF('Survival Used'!S24=1,X11*(1-O12),(F12*intCohort*MIN('Survival Used'!Q24,1))),0)

		Copy cell formula Paste to range X13:X1342
	P11	=MIN(F11*intCohort*'Survival Used'!M23,X11) Copy cell formula Paste to range P12:X1342
"Model Control"	M70	="Treatment_Waning"
	N70	Name cell "Treatment_Waning" Set = 1 to apply waning
	M71	="Waning start year"
	N71	Name cell "Waningyear" Set = 5 for ERG waning scenario