in collaboration with:

ERASMUS UNIVERSITEIT ROTTERDAM
INSTITUTE OF HEALTH POLICY & MANAGEMENT

Maastricht University

Nivolumab for treating metastatic or unresectable urothelial cancer

Produced by	Kleijnen Systematic Reviews Ltd. in collaboration with Erasmus University Rotterdam (EUR) and Maastricht University
Authors	Nigel Armstrong, Health Economist, Kleijnen Systematic Reviews Ltd, UK Sabine Grimm, Health Economist, Maastricht UMC Bram Ramaekers, Health Economist, Maastricht UMC Xavier Pouwels, Health Economist, Maastricht UMC Shona Lang, Systematic Reviewer, KSR Ltd Debra Fayter, Systematic Reviewer, KSR Ltd Svenja Petersohn, Health Economist, Maastricht UMC Rob Riemsma, Reviews Manager, KSR Ltd Gill Worthy, Statistician, KSR Ltd Lisa Stirk, Information Specialist, KSR Ltd Janine Ross, Information Specialist, KSR Ltd Manuela Joore, Health Economist, Maastricht UMC Jos Kleijnen, Director, KSR Ltd, Professor of Systematic Reviews in Health Care, Maastricht University
Correspondence to	Nigel Armstrong, Kleijnen Systematic Reviews Unit 6, Escrick Business Park Riccall Road, Escrick York, UK YO19 6FD
Date completed	30/08/2017

Source of funding: This report was commissioned by the NIHR HTA Programme as project number STA 16/108/11.

Declared competing interests of the authors

None.

Acknowledgements

None.

Commercial in confidence (CiC) data are highlighted in blue throughout the report.

Academic in confidence (AiC) data are highlighted in yellow throughout the report.

Rider on responsibility for report

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

This report should be referenced as follows:

Armstrong N, Grimm S, Ramaekers BLT, Pouwels X, Lang S, Fayter D, Petersohn S, Riemsma R, Worthy G, Stirk S, Ross J, Joore MA, Kleijnen J. Nivolumab for treating metastatic or unresectable urothelial cancer: a Single Technology Assessment. York: Kleijnen Systematic Reviews Ltd, 2017.

Contributions of authors

Nigel Armstrong acted as project lead and systematic reviewer on this assessment, critiqued the clinical effectiveness methods and evidence and contributed to the writing of the report. Sabine Grimm acted as health economic project lead, critiqued the company's economic evaluation and contributed to the writing of the report. Bram Ramaekers, Xavier Pouwels and Svenja Petersohn acted as health economists on this assessment, critiqued the company's economic evaluation and contributed to the writing of the report. Debra Fayter, Shona Lang and Rob Riemsma acted as systematic reviewers, critiqued the clinical effectiveness methods and evidence and contributed to the writing of the report. Gill Worthy acted as statistician, critiqued the analyses in the company's submission and contributed to the writing of the report. Lisa Stirk and Janine Ross critiqued the search methods in the submission and contributed to the writing of the report. Manuela Joore acted as health economist on this assessment, critiqued the company's economic evaluation, contributed to the writing of the report and provided general guidance. Jos Kleijnen critiqued the company's definition of the decision problem and their description of the underlying health problem and current service provision, contributed to the writing of the report and supervised the project.

Abbreviations

Ab	Antibody
AE	Adverse Events
AIC	Akaike information criterion
ALT	Alanine transaminase
BI	Budget impact
BIC	Bayesian information criterion
BIRC	Blinded independent review committee
BNF	British National Formulary
BOR	best overall response
BSA	body surface area
BSC	Best supportive care
CDF	Cancer Drugs Fund
CD28	Cluster of differentiation 28
CE	Cost Effectiveness
CEA	Cost-effectiveness Analysis
CEAC	Cost effectiveness Acceptability Curve
CHMP	Committee for Medicinal Products for Human Use
CI	Confidence Interval
CIS	Carcinoma in situ
Cis	Cisplatin
CR	Complete response
CRD	Centre for Reviews and Dissemination
CrI	Credible interval
CS	Company's submission
CSR	Clinical study report
CHMP	Committee for Medicinal Products for Human Use
CT	Computer tomography
CTCAE	Common Terminology Criteria for Adverse Events (NCI)
CTLA-4	Cytotoxic T-lymphocyte-associated protein 4
"D" ¯ "res"	Residual deviance
DIC ¯	Deviance information criteria
DOR	Duration of response
DSU	Decision Support Unit
EAU	European Association of Urology
ECOG	Eastern Cooperative Oncology Group
EMA	European Medicines Agency
eMIT	Electronic market information tool
EORTC	European Organisation for Research and Treatment of Cancer
EORTC QLQ-C30	European Organisation for Research and Treatment of Cancer quality of life questionnaire
EPAR	European public assessment report
EQ-5D	European Quality of Life-5 Dimensions
EQ-5D-3L	European Quality of Life-5 Dimensions, three-level scale
ERG	Evidence Review Group
ESMO	European Society for Medical Oncology
EUR	Erasmus University Rotterdam
FDA	Food and Drug Administration
FP	Fractional polynomial
G-CSF	Granulocyte colony stimulating factor
GCP	Good Clinical Practice
Gem	Gemcitabine
GFR	Glomerular filtration rate
GP	General practitioner

HR	Hazard ratio
HRQoL	Health-related Quality of Life
HTA	Health Technology Assessment
IC	Indirect Comparison
ICD	International Classification of Diseases
ICER	Incremental Cost Effectiveness Ratio
IFNγR	Interferon gamma receptor
IPD	Individual patient data
IQR	Interquartile range
ITC	Indirect treatment comparison
ITT	Intention to Treat
IV	Intravenous
KM	Kaplan–Meier
KSR	Kleijnen Systematic Reviews
LPFT	Last patient first treatment
LYG	Life years gained
LYS	Life Year Saved
LYs	Life years
MedDRA	Medical Dictionary for Regulatory Activities
MeSH	Medical Subject Headings
MHC	Major histocompatibility complex
MHRA	Medicines and Healthcare Products Regulatory Agency
MICE	Multiple imputation by chained equations
MRI	Magnetic resonance imaging
MVAC	Methotrexate, vinblastine, doxorubicin and cisplatin
NA	Not applicable
NCI	National Cancer Institute
NF-κB	Nuclear transcription factor-κB
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NMA	Network meta-analysis
NMB	Net monetary benefit
NR	Not Reached/Not Reported
NSCLC	Non-small cell lung cancer
OR	Odds Ratio
ORR	Objective response rate
OS	Overall survival
PAS	Patient access scheme
PD	Progressive disease
pD	Number of effective parameters
PD-1	Programmed death 1
PD-L1	Programmed death ligand 1
PD-L2	Programmed death ligand 2
PFS	Progression-free survival
PH	Proportional hazards
PP	Post-progression
PR	Partial response
PRESS	Peer Review of Electronic Search Strategies
PROs	Patient-reported outcomes
PS	Performance status
PSA	Probabilistic Sensitivity Analyses
PSS	Personal Social Services
PSSRU HCHS	Personal and Social Services Research Unit Hospital and Community Health Services

PI3K	Phosphoinositide 3-kinase
Q2W	Every 2 weeks
Q3W	Every 3 weeks
QALY(s)	Quality-adjusted Life Year(s)
QoL	Quality of life
RCT	Randomised Controlled Trial
RECIST	Response Evaluation Criteria In Solid Tumours
SAE	Serious Adverse Events
SD	Stable disease/Standard deviation
SE	Standard error
Shp-2	Src homology 2 domain-containing protein tyrosine phosphatase 2
SLR	Systematic literature review
SmPC	Summary of product characteristics
STA	Single Technology Appraisal
STC	Simulated treatment comparison
TEAEs	Treatment-emergent adverse events
TESAEs	Treatment-emergent serious adverse events
TNM	Tumour-node-metastasis
TTD	Time to treatment discontinuation
TTR	Time to response
TURBT	Transurethral resection of the bladder tumour
UC	Urothelial carcinoma
UICC	Union for International Cancer Control
UK	United Kingdom
UMC	University Medical Centre
USA	United States of America
VAS	Visual analogue scale
VEGF	Vascular endothelial growth factor
VGPR	Very good partial response
WHO	World Health Organisation

Table of Contents

Table of Tables

Table of Figures

1. SUMMARY

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

The patient population described in the final scope issued by the National Institute for Health and Care Excellence (NICE) was '*Adults with metastatic or unresectable urothelial cancer whose disease has progressed after platinum-based chemotherapy'*. Nivolumab was to be compared to retreatment with first-line platinum-based chemotherapy (only for people whose disease has had an adequate response), paclitaxel, docetaxel or best supportive care. Outcomes included overall survival (OS), progression free survival (PFS), response rates, adverse effects of treatment and health-related quality of life (HRQoL).

There were several deviations between the decision problem addressed by the company submission and that of the final scope issued by NICE. For the population, the company submission (CS) was in agreement with the scope, although only one of the two pivotal nivolumab trials included patients from the UK. Both nivolumab studies were small (270 and 78 patients for CheckMate 275 and CheckMate 032 respectively); only six patients were from the UK. For the intervention, the CheckMate 275 trial was in line with the scope, but in the CheckMate 032 trial 23% patients switched to ipilimumab. For the comparator, both nivolumab trials were single arm studies and therefore no direct or indirect comparators were included. Simulated treatment comparisons (STC) were performed for comparisons of nivolumab to paclitaxel, docetaxel and best supportive care (BSC). Comparisons of nivolumab to cisplatin plus gemcitabine were included only as part of a scenario analysis. The ERG would have considered cisplatin and gemcitabine suitable for inclusion in the STC, especially given the limitations in the quantity and quality of evidence for nivolumab and all other comparator trials. For the outcomes, comparative data in the form of an STC was only provided for OS, PFS and objective response rate (ORR). There were no comparative analyses for adverse events or quality of life.

1.2 Summary of clinical effectiveness evidence submitted by the company

1.2.1 Direct evidence

The company conducted a systematic literature review (SLR) to inform the submission. The aim of the SLR was '*to understand the relative efficacy and safety of nivolumab compared to alternative therapies for adult patients with metastatic or unresectable urothelial cancer whose disease has progressed after platinum-based chemotherapy'*.

The company did not identify any randomised controlled trials (RCTs) for nivolumab. Two ongoing phase I/II single arm studies for nivolumab were identified (CheckMate 275 and CheckMate 032). Therefore no studies were found that directly compared nivolumab with any specified comparator.

Single arm data for nivolumab

Data from the individual trials indicated that for Check Mate 275 (n=275) nivolumab led to a confirmed ORR (BIRC) in 54 (20.0%) patients (95% CI: 15.4 to 25.3). In CheckMate 032 (n=78) nivolumab led to a confirmed ORR (BIRC) in 19 (24.4%) patients (95% CI: 15.3–35.4).

For CheckMate 275, at the latest database lock of 2 September 2016 (n=270 analysed), nivolumab led to a median OS of 8.57 months (95% CI: 6.05–11.27) and for CheckMate 032 (n=78) nivolumab led to a median OS of 9.72 months (95% CI: 7.26–16.16).

For CheckMate 275, at the latest database lock of 2 September 2016 (n=270 analysed), nivolumab led to a median PFS of 2.0 months (95% CI: 1.87–2.63) and for CheckMate 032 (n=78) nivolumab led to a median PFS of 2.78 months (95% CI: 1.45–5.85).

Health related-quality of life (HRQoL) data was limited either by the currently available follow-up data or patient numbers.

For CheckMate 275 (May 2016 database lock) 75.6% of patients discontinued treatment with nivolumab (disease progression, 53.3%; adverse events (AEs) unrelated to nivolumab, 12.6%; nivolumab toxicity, 5.2%). For CheckMate 032 (March 2016 database lock) 76.9% of patients discontinued study treatment (disease progression, 64.1%; nivolumab toxicity, 2.6%).

In the CheckMate 275 trial 51.1% of patients died (1.1% attributed to nivolumab toxicity), whilst in CheckMate 032 trial 46.2% of patients had died (2.6% attributed to nivolumab toxicity). In the CheckMate 275 trial 64.4% of patients had a drug related AE ([redacted] serious drug related AE), whilst in CheckMate 032 trial 83.3% of patients had a drug related AE (10.3% serious drug related AE).

Data for the CheckMate trials were pooled for the STC but the pooled results or method were not provided, despite a request in the clarification letter.

1.2.2 Indirect evidence

The identification of two single arm studies for nivolumab precluded any conventional mixed treatment comparison (MTC) or indirect meta-analysis. There were no studies that could provide a common comparator to support any indirect comparison or MTC. As a consequence the company decided to perform an unanchored (no common comparator) stimulated treatment comparison (STC).

Single arm data for comparators

Single arm data is provided as an alternative to the STC to allow naive comparisons to the single arm data of nivolumab. Data from the comparator trials indicated that paclitaxel (one trial, n=45) led to overall ORR (definition not reported) in four (9.0%) patients (95% CI: 2 to 21), gemcitabine and cisplatin (two trials, n=53) led to ORR (not defined) in 13 (39.4%) to eight (40.0%) patients (95% CI: NR), docetaxel and placebo (one trial, n=72) led to confirmed ORR (overall PR or CR) in eight (7.1%) patients (95% CI: NR) and docetaxel (one trial, n=45) led to ORR (best overall PR or CR) in four (8.9%) patients (95% CI: 2.5 to 21.2). ORR data for BSC was not identified.

BSC (one trial, n= 117) had a median OS of 4.6 months (95% CI: 4.1 to 6.6), paclitaxel (one trial, n=65) had a median OS of eight months (80% CI: 6.9 to 9.7), gemcitabine and cisplatin (one trial, n=65) had a median OS of 10.5 months (95% CI: 3 to 22.9), docetaxel and placebo (one trial, n=72) had a median OS of 7.03 months (95% CI: 5.19 to 10.41) and docetaxel (one trial, n=45) had a median OS of 9.2 months (95% CI: 5.7 to 11.7).

Docetaxel and placebo (one trial, n=72) had a median PFS of 1.58 months (95% CI: 1.48 to 3.09) and docetaxel (one trial, n=45) had a median PFS of 2.8 months (95% CI: 1.9 to 3.6). PFS data from other comparators were not available.

Simulated treatment comparison

The STC approach uses nivolumab IPD to attempt to model how patients might respond to treatment if they were more like those in a comparator trial based on key baseline characteristics. A prediction model is intended to adjust the difference in outcomes observed between the nivolumab and comparator studies given the high risk of bias that must exist in comparing observational data. The outcomes for which this method was applied were OS, PFS and ORR. Key characteristics were identified using literature searches and using discussions with clinical advisors. Eleven characteristics were initially identified, but no more than four characteristics were used per outcome. It was reported that stepwise model selection suggested that the best Cox Proportional hazards (PH) model for OS is based on Eastern

Cooperative Oncology Group (ECOG) performance status (PS), haemoglobin level, visceral metastases and liver metastases. For PFS the same approach showed the best model is based on ECOG PS, age, visceral metastases and liver metastases. Stepwise model selection suggested that the best logistic regression model for objective response is based on age and visceral metastases. The basis of selection was reported to be parsimony as indicated by the Akaike information criterion (AIC). No models other than the final and presumably most parsimonious models (no more than four covariates) were presented despite the consideration of 11 possible covariates. Since an unanchored STC relies on the major assumption that all effect modifiers and prognostic factors are accounted for, the NICE Decision Support Unit (DSU) Technical Support Document (TSD) 18 recommends caution in the application of the method. It also recommends a so-called 'out-of-sample' method for estimating the residual bias of any STC, due to effect modifiers or prognostic variables that are not accounted for in the prediction models. The company provided such an analysis in their response to the request for clarification.

Finally an evidence synthesis model was used to synthesise the results of the STC i.e. adjusted hazard ratios (HRs) (for OS and PFS) and odds ratios (for ORR) across all trials. For OS and PFS this enabled the adoption of an evidence synthesis model that did not require a PH assumption i.e. a fixed HR of nivolumab versus each comparator, but instead allowed the HR to vary over time, one HR per four-week period. This model, based on a paper by Jansen, 2011, is known as fractional polynomial (FP) and through variation in a set of up to two key parameters (P1 and P2) permits a wider variation in the form of the survival curves. Choice of FP model was reported to have been determined by best statistical fit, although the results of only two other sets of parameter values out of many possible were presented in Appendix D. The company also presented the results of analyses based on a PH model for OS and PFS i.e. fixed HRs in response to the request for clarification. The company were also requested in the clarification letter to present the results by Programmed death ligand 1 (PD-L1) subgroup, but they declined citing lack of baseline data in the comparator studies.

The systematic review identified 12 trials for inclusion in the STC; three were excluded as the dose and/or treatment regimens did not correlate with current UK clinical practice. In addition to the two nivolumab studies, two comparator studies were identified of paclitaxel, two of docetaxel, one of BSC, and two of cisplatin plus gemcitabine. Because not all studies reported all outcomes, only five were used for OS, one per comparator for all comparators except docetaxel for which there were two. The comparator studies were a mix of randomised controlled trials or single arm studies. For PFS only three were used, two for docetaxel and one for paclitaxel. For ORR six of seven studies were synthesised, only one paclitaxel study not being included. There was much variability in patient populations between the included studies of the STC.

The analysis based on the STC and using a fixed effect FP model with P1=0 and P2=0 found that for OS nivolumab is superior to all comparators but only at certain time points; the credible intervals for the HRs were quite wide and indicated the results were not always statistically significant. For OS nivolumab was statistically superior to: paclitaxel at time points between 44 and 72 weeks (HR 2.63, 95% CrI 1.17 to 5.52, 68 -72 weeks); docetaxel at time points between 20 and 72 weeks (HR 2.01, 95% CrI 1.14 to 3.37, 68 -72 weeks); BSC at time points between 20-72 weeks (HR 1.86, 95% CrI 1.17 to 2.85, 68 -72 weeks). Nivolumab was superior to cisplatin plus gemcitabine above 20 weeks but never reached statistical significance.

The analysis based on the STC and using a fixed effect FP model of PFS with P1=0 AND P2=0 was only possible for nivolumab compared to paclitaxel or compared to docetaxel. For PFS nivolumab was statistically superior to: paclitaxel at time points between 20 to 72 weeks (HR 7.26, 95% CrI 1.40 to

28.85, 68 to 72 weeks); docetaxel at time points between 8 to 12 weeks only (HR 1.72, 95% CrI 1.18 to 2.49).

The STC analysis of ORR using a fixed effect model found that nivolumab is significantly better than BSC (OR 106.70, 95% CrI 6.72 to 49820) or docetaxel (OR 3.12, 95% CrI 1.06 to 9.49), although the uncertainty was large. No significant differences were found for nivolumab compared to paclitaxel or gemcitabine plus cisplatin. In the random effects model nivolumab was only statistically superior to BSC (OR 108.1, 95% CrI 4.17 to 52240).

No formal comparison of AEs including no evidence synthesis was performed. However, the rate of neutropaenia was generally lower than for most comparators, the exception being BSC, and much lower than for cisplatin and gemcitabine. The rate for anaemia was a little lower except for being much lower than BSC and even lower again in comparison to cisplatin and gemcitabine. For leaukopaenia the rate was comparable i.e. 0% between all comparators where it was reported except against cisplatin plus gemcitabine. The rate of asthaenia was also lower than all comparators except cisplatin plus gemcitabine.

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

The company submission (CS) and response to clarification provided sufficient details for the ERG to appraise the searches for eligible trials. Searches were carried out in accordance with the NICE guide to the methods of technology appraisal sections 5.2.2 and 5.2.4 using a good range of databases. Additional searches of conference proceedings were reported, along with trials registers and the checking of reference lists of existing systematic reviews and health technology assessments (HTAs). The systematic review was performed to a good standard.

The ideal scenario to determine the relative benefits of nivolumab and its comparators would be a series of RCTs comparing nivolumab to its comparators. Failing this, a network meta-analysis of RCTs using a set of common comparators would be the preferred approach. However the submission relies on two single arm studies of nivolumab, which are entered into a STC together with the single arms of comparator studies. Single arm studies are basically observational studies and are considered low order for study quality. The methods used by the company to conduct the STC largely follow those described in NICE DSU TSD 18, but, as stated in the same TSD, given no comparative data (unanchored analysis) the results obtained should be treated with caution. The ERG found the following limitations in the STC analysis:

1. There was no STC analysis for AEs or HRQoL. Therefore the value of any potential extension to life cannot be judged in relation to any changes to the patients' quality of life.
2. The analysis relies on two small single arm nivolumab studies, one includes 78 patients and the other included 275. Therefore any statistical analyses have increased uncertainty due to the small sample size.
3. The numbers of patients are small for all comparator studies (33 to 117) and not all studies provided data for all outcomes.
4. There were no common comparators; therefore an unanchored STC had to be performed.
5. The company pooled the two nivolumab trials despite each one using different methods of outcome assessment, CheckMate 275 using BIRC and CheckMate 032 using investigator-assessed. The results of this pooling (and its variability) were not reported.
6. Ideally the results of the STC would be based on independent review (BIRC) assessment methods. Given that the BIRC method was only available for CheckMate 275 at a minimum it

would have been useful to perform the STC using only the CheckMate 275 data. This was suggested to the company but was not performed.

7. The major assumption for unanchored STC is that all effect modifiers or prognostic variables are accounted for. Not all of the key characteristics (possible effect modifiers or prognostic variables) for the STC were reported for all comparator trials, therefore imputations were required for these characteristics which were based on correlations to the baseline characteristics in the nivolumab trials.
8. The method used for the prediction models lacked transparency; the results at each stage of the stepwise selection process were not provided. In particular, it is not clear that the most parsimonious model is the best model. It would have been useful to see an STC that was based on prediction models with more covariates including all 11 considered. The only external test of validity of the STC i.e. the 'out-of-sample' method seemed to either show insufficient reduction in bias or be inapplicable given the use of the fractional polynomial model that was used for survival analysis. As stated on page 56 of TSD 18: '*The size of this systematic error can certainly be reduced, and probably substantially, by appropriate use of…STC. Much of the literature on unanchored … STC acknowledges the possibility of residual bias due to unobserved prognostic variables and effect modifiers; however, it is not made clear that the accuracy of the resulting estimates is entirely unknown, because there is no analysis of the potential magnitude of residual bias, and hence no idea of the degree of error in the unanchored estimates. It is, of course, most unlikely that systematic error has been eliminated. Hoaglin,[72, 73] in a series of letters critiquing an unanchored comparison by Di Lorenzo et al.[78] based upon a matching approach similar to MAIC, remarked that, without providing evidence that the adjustment compensates for the missing common comparator arms and the resulting systematic error, the ensuing results "are not worthy of consideration".*'[1]

Analysis of the single arm studies alone indicates that there is little difference in survival at least at the median between nivolumab at 8.74 and 9.72 respectively and either docetaxel and paclicaxel, at 9.2 or 8 months respectively. The value for gemcitabine plus paclitaxel was higher at 10.5 months.

The ERG found that the FP model for synthesising HRs for OS and PFS is supportable partly because of its flexibility in permitting a wide variety of functional forms from fixed HRs (PH assumption) to time varying HRs with different shaped survival curves. However, whilst the company stated that they chose the base-case models on the basis of best fit, the results of only two of many parameter sets were presented in Appendix D. The company did provide the results for PH models in response to the clarification request, but the method used has questionable validity and was not the one recommended in the paper on which the FP approach was based. The ERG was able to reproduce the base-case PF model (fixed effect, P1=0, P2=0) results for OS and PFS at least close enough that any difference could be explained by uncertainty. The ERG was also able to produce results that were based on unadjusted values of hazards for nivolumab by applying the fixed HR, one for each comparator trial reported in Appendix D i.e. as if estimated without the STC for these base case PF models. This confirmed that the model used for the adjustment had been a PH model as described by the company. However, the uncertainty in these unadjusted HRs was not estimable without the original nivolumab IPD. Finally, the ERG did find that the HRs estimated using a PH model according to Jansen, 2011 were different to those provided by the company by an amount that did not seem explicable by uncertainty.

No formal comparison of AEs including no evidence synthesis was performed, although it might be reasonable to conclude, based on few data from the comparators that the rate of key AEs was generally similar to or lower than the comparators.

In conclusion, it is difficult to be sure what the effectiveness of nivolumab is in comparison to the comparators in the scope. Evidence from directly examining the single arms of the trial data indicates little difference between the outcomes measured from the nivolumab and comparator studies. Such a naive comparison carries a high risk of bias. STC analysis was used to try and reduce this bias, but there is also no clear evidence that risk of bias was reduced by the STC analysis. Multiple limitations in the STC were identified and the test of validity recommended by TSD 18, the 'out-of-sample' method lacked success in reducing the bias (if it is applicable at all given the lack of data and FP model). The ERG was able to estimate the unadjusted hazards for nivolumab, but not with estimates of uncertainty. The effect of an analysis based on different combinations of covariates in the prediction model used to make the adjustment remains unknown.

1.4 Summary of cost effectiveness submitted evidence by the company

Systematic literature review

The company performed a SLR with the objective to identify evidence to support the development of a cost effectiveness model for nivolumab as a treatment for locally unresectable or metastatic urothelial cancer (UC). Although economic evaluations were identified with populations that matched the population described in the final scope of this appraisal, these did not consider the cost effectiveness of nivolumab.

Model structure and main modelling decisions

The company developed a de novo economic model using a cohort-based partitioned survival model. The model consists of three mutually exclusive health states: progression-free (PF) and post-progression (PP) disease states and death. Patients enter the model in the PF state and are treated with nivolumab or one of its comparators. Patients remain in the PF state until disease progression or death. The proportion of patients in each health state is determined by overall survival (OS) and progression-free survival (PFS) curves.

The model includes patients with metastatic or unresectable UC who have progressed following first-line platinum-based chemotherapy. Patient characteristics included in the model were age, gender, weight and body surface area (BSA) based on the CheckMate 275 study.

Nivolumab is implemented as per its EMA Summary of Product Characteristics (SmPC) posology and method of administration for second-line UC (i.e. 3mg/kg Q2W).

The company considered the following comparators in their base-case:

- Paclitaxel: 80mg/m^2 Q3W of a four week cycle
- Docetaxel: 75mg/m^2 Q3W
- Best supportive care (BSC)

The company also presented a scenario analysis, in which cisplatin plus gemcitabine was added as a comparator. The company justified this deviation from the scope (i.e. not including cisplatin plus gemcitabine in its base-case) by stating that there was limited evidence for retreatment with first-line platinum-based chemotherapy regimens for patients with locally advanced unresectable or metastatic UC.

The model adopts the perspective of the NHS and Personal and Social Services (PSS) in England and Wales. The cycle length is four weeks to account for the length of treatment cycles. A lifetime time horizon was adopted to capture all relevant costs and health-related utilities. All costs and utilities were discounted at a rate of 3.5% per year.

Treatment and relative effectiveness

Treatment effectiveness estimates were derived from the CheckMate 275 and CheckMate 032 studies. The time-to-event data of both studies were combined for the survival analyses, but the pooling method was not stated. Parametric time-to-event models were used to estimate overall survival (OS), progression-free survival (PFS) and time-to-treatment discontinuation (TTD) in the company's cost effectiveness model. A response-based approach was adopted to estimate OS and PFS, but not for TTD in the company's base-case. In response to clarification questions, the company also enabled a response-based analysis for TTD for scenario analysis. The response-based analysis was used because, according to the company, standard survival modelling approaches would not appropriately characterise the novel mechanism of action of nivolumab and standard parametric time-to-event models were not deemed flexible enough to characterise the change in hazard over time resulting from having (long-term) responders, and non-responders (no supporting evidence provided). In its response-based analysis, the company used a landmark analysis to prevent the occurrence of immortal-time bias. In this landmark analysis, OS and PFS of both groups (responders and non-responders) were estimated together until a specified landmark point (eight weeks in the company's base-case, 26 weeks explored in scenario analysis) based on the Kaplan-Meier estimates, after which different survival curves were fitted for each group and adjusted for background mortality. The parametric time-to-event models used to estimate OS and PFS after the landmark were selected based on statistical fit (Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC)) and visual inspection. Out of exponential, Weibull, Gompertz, lognormal, log-logistic and generalised gamma, the generalised gamma was chosen to estimate OS and PFS of both responders and non-responders. OS and PFS estimates obtained from the parametric time-to-event models estimated for responders and non-responders separately were combined by using a weighted average, with the weighting based on the proportion of responders in patients being progression-free and alive at the eight-week landmark point. This weighting was held constant throughout the model time horizon. The adjustment for background mortality was based on UK life tables and incorporated using a distribution around the mean UK age (instead of the mean age of the cohort).

The relative effectiveness of nivolumab versus the comparators was modelled through time-varying HRs obtained mainly via the STC. The STC was performed based on the pooled CheckMate 032 and CheckMate 275 trials dataset, in which response status was not taken into account. The HRs obtained from the STC were then applied to the combined parametric time-to-event models of nivolumab which took response status into account. The company explained that the predicted OS and PFS of the comparators were mostly lower than the observed OS and PFS, especially for docetaxel, because of the differences in patient characteristics between the comparator trials and the CheckMate studies. Data not available from the STC relied on the following assumptions: PFS for BSC was derived assuming that the HR for BSC versus paclitaxel was equivalent to that of BSC versus vinflunine for second-line UC patients, and then applying this HR to the paclitaxel PFS curve. This HR was held constant during the time horizon of the cost effectiveness model, due to the absence of alternative data. PFS estimates for cisplatin plus gemcitabine were derived by assuming equivalence of cisplatin plus gemcitabine PFS with that of paclitaxel. No evidence was provided to support this assumption.

Time-to-treatment discontinuation

TTD was estimated through a parametric time-to-event model that was selected based on statistical fit (AIC and BIC) with the pooled CheckMate studies, as well as other, unspecified, considerations. In the CS, TTD was estimated independent of response status but response-based TTD analysis was enabled in response to clarification questions. Even though the Gompertz and the log-logistic distributions showed a better fit, the generalised gamma distribution was selected to estimate TTD in the base-case

analysis, with the company claiming that this was to ensure consistency with OS and PFS and that these two distributions produced long tails with some patients still on treatment after five and 10 years. TTD of the comparators was based on their respective PFS curves because it was assumed that comparator treatment would continue until disease progression. For paclitaxel, only six cycles of treatment were assumed (24 weeks). BSC was assumed to be administered until death.

Adverse events

The company stated that grade 3-4 adverse events were incorporated in the model if their incidence was ≥5%. The impact of adverse events on quality of life and resource use and costs were incorporated in the first cycle of the model.

Health-related quality of life

None of the studies identified by the SLR were consistent with the NICE reference case and therefore EQ-5D-3L data valued with UK preference weights were taken from the CheckMate 275 trial. These utility estimates were stratified according to progression-free and post-progression health states. Utility estimates were derived using a mixed-effects model to reflect within subject variance, after interpolating for measurement times deviating from the measurement schedule and adjusted for missing data using multiple imputation. This resulted in health state utilities of 0.718 and 0.604 pre-progression and post-progression respectively.

The company applied disutilities to several AEs based on studies reporting utilities in patients with non-small cell lung cancer, pancreatic cancer and leukaemia. Disutilities were not treatment-specific and were applied as one-off events at the beginning of treatment, based on the proportion of patients experiencing the adverse event and the duration of the adverse event.

Resource use and costs

Resource use and unit costs data to inform the economic model were based on a number of sources, including CheckMate 275, national databases, published sources (both sources identified and not identified in the SLR), clinical advice and assumptions. British National Formulary (BNF) was used to obtain unit prices for nivolumab (40mg and 100mg), which were adjusted by a Patient Access Scheme (PAS), [redacted] Unit prices for docetaxel, paclitaxel and gemcitabine plus cisplatin were taken from the electronic market information tool (EMIT). The dose/number of vials required per administration were estimated based on dosage scheme and dose intensity (reflecting missed doses), using estimations of patient average weight and body surface area (both based on the CheckMate 275 study) and calculating dose intensity based on data from CheckMate 275 and CheckMate 032 assuming that all delayed doses represent missed doses. Dose intensity for all comparators was assumed equal to that of nivolumab. Administration costs were added to each dose. Monitoring cost (while on treatment) estimates were based on resources estimated using expert opinion and unit prices derived from NHS reference costs. Best supportive care costs were incurred until death after treatment discontinuation. Although not described in the CS, treatment dependent AE costs were incorporated as one-off event costs for patients on treatment during the first cycle of the model based on their occurrence.

Cost effectiveness results

In the deterministic base-case analysis, nivolumab was associated with larger QALY and LY gains and costs than docetaxel, paclitaxel, and BSC. With the PAS, nivolumab treatment resulted in deterministic incremental cost effectiveness ratios (ICERs) of £37,646, £44,960, £38,164, and £71,608 per QALY gained versus docetaxel, paclitaxel, cisplatin plus gemcitabine and BSC respectively.

Probabilistic sensitivity analysis (PSA) and deterministic sensitivity analysis (DSA) were undertaken and presented by the company. Patient age, weight and BSA, costs, resource use, utilities, TTD, PFS and OS were varied but relative effectiveness estimates were not included in these analyses. The PSA with 1,000 iterations resulted in ICERs of £54,220, £46,209, £44,698 and £103,568 per QALY gained for nivolumab versus docetaxel, paclitaxel, BSC and cisplatin plus gemcitabine The company reasoned that the PSA ICER increases were mainly driven by a reduction in PFS and OS in the PSA (compared with the deterministic analysis), but did not provide further insights into the mechanism by which this occurred.

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

Systematic literature review

The cost effectiveness searches in the CS were well documented and easily reproducible, and were carried out in line with the NICE guide to the methods of technology appraisal, using a good range of databases. Additional searches of conference proceedings and organisational websites were reported, along with the checking of reference lists of existing systematic reviews, meta-analyses and health technology assessments.

Model structure and main modelling decisions

The choice of partitioned survival analysis for this decision problem is in line with other appraisals in metastatic cancer, but it should be noted that the recent NICE DSU TSD 19 advocates for alternative model structures that can more accurately reflect interdependent survival functions and use transition probabilities for each possible transition between health states. Another criticism relates to the company's response-based analysis, which if deemed appropriate, should have been incorporated in the model via separate responder and non-responder health states. The ERG considers the adopted perspective, time horizon and discounting to be appropriate for this appraisal.

The patient population used in the model was deemed consistent with the population of the CheckMate 275 and CheckMate 032 studies, as well as the final scope issued by NICE for this appraisal. The company did not provide the comparison of nivolumab with cisplatin plus gemcitabine in the base-case, despite it being in the scope and despite ERG request. The company justified this by citing expert opinion that the population in the only available cisplatin plus gemcitabine study differed from the UK population in that the study population received MVAC in first line instead of cisplatin plus gemcitabine. The ERG considered this to be challengeable in that patients in the cited study would have had exposure to platinum-based therapy and that the precise combination of first-line treatment or naivety to gemcitabine might therefore be irrelevant. Furthermore, a relevant comparator should not be excluded based on issues with the data.

Treatment effectiveness, relative effectiveness and TTD

One of the main issues was that it was unclear whether pooling both CheckMate 032 and CheckMate 275 trials was appropriate and how this was done. The company failed to provide further details upon the ERG's request.

Furthermore, the ERG wishes to express strong concerns about the appropriateness of response-based analysis, implemented through landmark analysis. The need for response-based analysis was inadequately justified, with the company failing to demonstrate how standard parametric survival analysis methods failed to describe the mechanism of action of nivolumab in urothelial carcinoma. In contrast to what the company stated, most standard parametric time-to-event models do include changing hazards over time and some allow for non-monotonic changing hazard functions over time. No mathematical reasoning was provided and based on visual inspection of the conventional, not

response-based, conventional, survival analysis alone, it is the ERG's view that the need for response-based analysis could not be established. The ERG considers that a standard approach should be shown to be inappropriate in the particular decision problem at hand before discarding it.

If, however, the need for alternative methods to conventional survival analysis could be justified, it is the ERG's view that the methods recommended in NICE DSU TSD 14 should be considered before adopting a landmark analysis. However, the company stated that these alternatives, such as spline-based or mixture cure models, were not considered. In summary, the company (a) did not provide sufficient evidence to demonstrate that conventional parametric time-to-event models failed to describe nivolumab survival, (b) did not provide evidence to support that the committee's criticisms on previous nivolumab appraisals applied to the current appraisal, and (c) did not provide evidence to demonstrate that the landmark analysis provided more valid results than standard survival modelling analyses or alternative methods recommended in TSD 14 (for example, no expert opinion was used to validate the resulting survival curves).

The use of response-based landmark analysis introduced further assumptions and additional uncertainty into the cost effectiveness analysis. These assumptions include (a) the choice of the eight-week landmark, with alternative choices causing unpredictable changes in cost effectiveness (the company only provided one alternative landmark and declined to provide others upon request); (b) the use of Kaplan-Meier estimates for the period up to the landmark instead of fitting a parametric curve until then may result in overfitting; (c) fitting parametric models to the responder and non-responder groups also results in larger uncertainty about these fitted curves: the sample size used is significantly smaller because of the splitting up of the study population into two groups and because only the available data after the landmark is used; (d) responder and non-responder groups were then combined for the indirect comparison casting further doubt over whether the response-based analysis has any benefits, especially given that hazard ratios are derived from the overall population and are then applied in a combined responder and non-responder population. The combination of curves was implemented using a weighted average, with the weight being the proportion of responders at the landmark, which was held constant. This inflated the proportion of non-responders in later periods because the proportion of responders is expected to increase over time compared to the proportion of non-responders; (e) response-based and conventional approaches result in vast differences in the predicted life years for nivolumab, with a predicted mean of 2.80 life years in the response-based analysis and 1.84 life years in the conventional, not response-based, approach (deterministic estimates). No explanation for this deviation was provided, and none of the response-based model predictions were validated using expert opinion. The use of response-based, and landmark, analysis had by far the biggest impact on the ICERs, with ICERs being significantly decreased in all comparisons when using the response-based approach.

The ERG's concerns about the selection of parametric time-to-event models include the rejection of the proportional hazard assumption between responders and non-responders without sufficient justification, and the simultaneous selection of parametric time-to-event models for responders and non-responders, which stands in contrast to the company's statement that there was *'no requirement to assume the same distribution to be appropriate for both responder and non-responder curves'*. This led to selection of the generalised gamma distribution, despite it not making the best statistical fit for non-responders (the Weibull makes a better fit). The company provided an updated model allowing the selection of differential distributions for responders and non-responders. Of further concern is that, despite NICE DSU TSD 14 recommendations, the choice of parametric time-to-event models for the response-based approach was not supported by expert opinion. Furthermore, the company was inconsistent in not using response-based analysis for estimating TTD. For TTD, the company chose the generalised gamma distribution despite it not having the best statistical fit and justified their choice by stating that the better

fitting Gompertz and log-logistic distributions would result in implausible numbers of patients still on treatment at five years. The choice of differential parametric time-to-event curves for responder and non-responder OS, PFS and TTD was shown to significantly increase the ICERs in ERG scenario analyses.

The cost effectiveness analysis model suffers from significant uncertainty and bias induced by comparing single-arm studies through the STC. It is the ERG's opinion that the discrepancy in populations in which relative effectiveness estimates were derived (adjusted CheckMate 275 and CheckMate 032 population) and applied (i.e. the combined but separately estimated responder and non-responder survival curves) induced bias that could not be quantified and that the company declined to comment on, despite the ERG's request. The ERG would have preferred to apply separate HRs to responders and non-responders. However, the company did not provide these, stating that small numbers in responder and non-responder groups did not allow separate estimation of relative effectiveness.

The company did not sufficiently justify the need for time-dependent HRs to model the relative effectiveness of nivolumab versus the comparators, providing log-cumulative hazard plots that showed the separate CheckMate studies, while the HRs were derived based on the pooled CheckMate studies dataset. The ERG considers that therefore proportionality of hazards could not be ruled out. Time-independent HRs were provided by the company in response to clarification questions but these could not be replicated by the ERG. The use of the time-independent HRs produced by the ERG increased all cost effectiveness estimates in ERG scenario analysis. The ERG notes that using time-independent HRs has the advantage of preventing over-parameterisation which might occur when estimating time-dependent HRs with the relatively limited amount of data submitted by the company.

Assumptions that were not supported by clinical evidence were made around the relative effectiveness of nivolumab versus cisplatin plus gemcitabine and BSC in terms of PFS to make up for lack of data to inform these. Alternative assumptions in ERG scenario analysis only had a small effect on the ICERs in these comparisons.

The parameterisation of the fractional polynomial model that informs the NMA was found to have a large impact on cost effectiveness outcomes. In a PSA only varying the parameter values of the FP model between those parameter values that were provided as possible parameter combinations by the company resulted in substantial differences in incremental costs and QALYs for all comparators (for instance, incremental QALYs of nivolumab vs docetaxel had a credible interval of [redacted] to [redacted]).

Adverse events

Only the CheckMate 275 trial was used to inform the adverse event rates in the cost effectiveness model while the clinical effectiveness of nivolumab was estimated based on both CheckMate studies. The selection of sources for adverse events associated with comparators was not appropriately justified. The inclusion of both neutropenia and leukopenia was questionable, given that neutropenia is a subtype of leukopenia. There was an inconsistency in that not all included adverse events matched the inclusion criteria of having an incidence of ≥5%.

Health-related quality of life

The ERG identified several inconsistencies and choices lacking justification in the handling of health-related quality of life estimates. The main issues include inconsistencies in reported observations, the use of utilities derived only from CheckMate 275, the imputation of immature data, the use of multiple imputation instead of the mixed model to adjust for missing data, and inconsistencies in disutilities for adverse events with those used for a previous nivolumab appraisal.

Resource use and costs

Estimation of resource use and costs included a technical error in calculating the dose intensity; inconsistencies in using the average weight and BSA from CheckMate 275 (not using CheckMate 032) and in using the subsequent treatment proportions from CheckMate 275 (not using CheckMate 032). Further inconsistencies related to not using cost and resource use data from TA272 (identified in the SLR), and using different AE unit costs compared with a previous nivolumab appraisal. Some assumptions lacked justification, such as the assumption of an administration scheme that is inconsistent with UK clinical practice for cisplatin plus gemcitabine, the assumption that all delayed doses are missed doses for calculating nivolumab dose intensity, and assuming that the dose intensity for the comparators is equal to that of nivolumab.

Cost effectiveness results

Cost effectiveness results were not presented for one comparator identified in the scope (cisplatin plus gemcitabine) in the base-case. In their sensitivity analyses, the company did not explore important parameters regarding relative effectiveness. The number of iterations (1,000) used in the PSA was shown to not yield stable results. The company subsequently provided a PSA with 10,000 simulations, but this still did not achieve stability. Furthermore, there were marked differences between the deterministic and probabilistic results in the company's base-case, which the company did not provide explanation for. These differences were largely resolved by removing response-based analysis. The PSA did not include relative effectiveness estimates, but it did include inappropriate parameters, such as patient characteristics (age, weight) and comparator treatment costs. The company justified the exclusion of hazard ratios from the PSA by stating that sampling the time-dependent hazard ratios in each period independently would yield counter-intuitive results. However, it is possible to circumvent this problem, for example, by using a fixed set of random numbers. Because relative effectiveness estimates are by far the largest contributor to decision uncertainty, the PSA was deemed to be insufficient.

The ERG's concerns on validation include the lack of internal and cross validity efforts as well as sparse use of expert opinion; external validation efforts that are based on a lung cancer study only and therefore questionable in terms of their relevance; the use of only CheckMate 275 for validating model predictions; as well as transparency issues with the model.

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

1.6.1 Strengths

The searches for eligible studies in the CS were well documented and easily reproducible. Searches were carried out on a good range of databases. Supplementary searches of conference proceedings, and clinical trials registers, and the checking of references lists were undertaken by the company in order to identify additional studies not retrieved by the main searches.

Overall the systematic review process was well documented and appeared to be performed well.

The ERG considers the adopted perspective, time horizon and discounting used in the model to be appropriate for this appraisal. Incorporation of costs, resource use, and HRQoL data was appropriate, with a few minor errors and questionable judgements. The model structure followed that of past NICE technology appraisals in metastatic cancers. The company explored a range of different parametric time-to-event models to model survival data.

1.6.2 Weaknesses and areas of uncertainty

All nivolumab trial data were based on March, May and September 2016 database locks. More up-to-date data was requested but was not provided.

The ERG was concerned that limiting the MEDLINE and Embase clinical effectiveness searches to English language only publications may have introduced potential language bias.

No randomised controlled trials (RCTs) were identified for nivolumab.

There were no studies that directly compared nivolumab with any specified comparator. Furthermore, there were no studies that could provide a common comparator to support indirect comparison or MTC.

There are serious concerns regarding the representativeness of the nivolumab trial patients to the UK population. Firstly, only six patients from one trial were from the UK. Secondly, as few as 18.8% of patients in the UK might have and ECOG performance status of 0, as opposed to over 50% in the two nivolumab trials. Thirdly, there is a mismatch in terms of prior therapies, as many as over 75% of patients in the UK would have previously taken a gemcitabine platinum-based combination compared to fewer than 40% in the trials. Finally, there is a question of the applicability to those with locally advanced unresectable as opposed to metastatic disease given the very small proportion of such patients in the trials.

Risk of bias was not assessed appropriately for the single arm studies (which include those for nivolumab). Single arm studies are by definition low down in the hierarchy of study design and therefore the quality of these studies is low to start with and risk of bias tools have not been widely developed for this study design. With this is mind risk of bias was judged to be high for all data used in the STC given that only single arms were used.

No STC analysis for AEs or HRQoL was performed.

The STC analysis is compromised by many limitations (listed earlier) which impairs the ability to critique the presence of residual bias. Given that TSD 18 states that *without providing evidence that the adjustment compensates for the missing common comparator arms and the resulting systematic error, the ensuing results "are not worthy of consideration"* the ERG does not think the STC methods are sufficiently reported nor validated to sustain the companies claims.

The company did not provide the comparison of nivolumab with cisplatin plus gemcitabine in their base-case model, despite it being in the scope.

With regards to a response-based modelling approach, the use of unconventional, response-based, landmark survival analysis, without sufficient justification for its need necessitated further assumptions and thereby substantially increased uncertainty. Assumptions introduced include the choice of the eight-week landmark, with alternative choices causing unpredictable changes in cost effectiveness; the use of Kaplan-Meier estimates for the period up to the landmark instead of fitting a parametric curve until then, which may result in overfitting; increased uncertainty resulting from fitting parametric models due to decreased sample size; and the combination of responder and non-responder groups using a weighted average, with the weight being the proportion of responders at the landmark, which was held constant. If a response-based analysis is used, this should translate into separate responder and non-responder health states in the model, with differential estimation of relative effectiveness, TTD, HRQoL and resource use and costs. There is therefore an inconstancy in using such an analysis without including these health states. Furthermore, alternative methods to the employed landmark analysis are recommended in NICE DSU TSD 14, but these were not considered by the company.

With respect to the relative effectiveness, the company ruled out proportionality of hazards between responders and non-responders without sufficient justification. OS and PFS estimates derived using the pooled CheckMate studies and response-based analysis were not validated by clinical experts, posing a non-adherence to TSD 14 recommendations. This is of even greater concern because (1) best statistical fit was not the only criterion used for selecting the parametric time-to-event models and (2) model predictions using the response-based approach were significantly different from model predictions using the conventional approach. The application of hazard ratios to an artificially created a posteriori mixed responder and non-responder population while these were derived from the a priori Checkmate matched population poses an inconsistency. The use of time-dependent HRs was not appropriately justified and potentially caused over-parameterisation. Assumptions around the relative effectiveness of nivolumab versus cisplatin plus gemcitabine and BSC in terms of PFS were not supported by clinical evidence. The parameterisation of the fractional polynomial model contributed significant uncertainty, which was not sufficiently explored.

There were inconsistencies in resource use, costs and disutilities associated with adverse events compared with a previous nivolumab appraisal.

Uncertainty caused by the many modelling assumptions was not appropriately explored in deterministic and probabilistic sensitivity analyses. The PSA did not include the, perhaps, most influential and uncertain relative effectiveness parameters.

1.7 Summary of exploratory and sensitivity analyses undertaken by the ERG

A number of issues were identified by the ERG. The ERG was able to adjust/correct some of these in its base-case. This resulted in ICERs (probabilistic) of £87,709, £68,519 and £69,515 for nivolumab (with PAS) versus docetaxel, paclitaxel and BSC, respectively. Cisplatin plus gemcitabine dominated nivolumab.

Additional sensitivity analyses were performed to examine the potential impact of alternative assumptions on the cost effectiveness estimates. These included two scenario analyses: a) exploratory analyses performed using the ERG base-case, and b) exploratory analyses performed using the ERG base-case, except that a response-based approach was used.

The company's and ERG base-case results as well as those scenario analyses with the largest influence on the ICERs are shown in Table 1.1. The uncertainty about the treatment and relative effectiveness evidence is characterised by scenarios A.3 (using a naïve treatment comparison), which increases the ICERs. Using alternative parametric time-to-event models within the ERG base-case can decrease the ICERs significantly (A.1). Finally, using the response-based (B.1) approach significantly decreases the ICER, but these ICERs can increase significantly with the use of best-fitting parametric time-to-event models (B.3). In addition to these exploratory analyses, the ERG also demonstrated that alternative parameter values informing the fractional polynomial model for the NMA could have a vast impact on the ICERs.

Table 1.1: Scenario analyses with significant impact on ICERs

	Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	Nivolumab ICER (£/QALY)
Probabilistic Company base-case[a]	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£12,748	0.82	[redacted]	[redacted]	£54,131

	Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	Nivolumab ICER (£/QALY)
	Paclitaxel	£14,186	0.71	[redacted]	[redacted]	£45,482
	Cis+gem	£30,443	1.34	[redacted]	[redacted]	£100,417
	BSC	£8,811	0.57	[redacted]	[redacted]	£44,873
ERG base-case	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£12,493	0.74	[redacted]	[redacted]	£87,709
	Paclitaxel	£13,866	0.63	[redacted]	[redacted]	£68,519
	Cis+gem	£29,384	1.24	[redacted]	[redacted]	Dominated
	BSC	£8,696	0.56	[redacted]	[redacted]	£69,515
Alternative parametric TTE models (lognormal for OS, log-logistic for PFS) (A.1)[b]	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£13,173	1.01	[redacted]	[redacted]	£45,721
	Paclitaxel	£14,654	0.89	[redacted]	[redacted]	£39,286
	Cis+gem	£29,736	1.58	[redacted]	[redacted]	£72,732
	BSC	£9,235	0.72	[redacted]	[redacted]	£38,147
Naïve comparison data instead of STC results (A.3)[b]	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£13,005	0.77	[redacted]	[redacted]	£92,335
	Paclitaxel	£13,914	0.60	[redacted]	[redacted]	£64,914
	Cis+gem	£30,910	1.56	[redacted]	[redacted]	Dominated
	BSC	£8,630	0.52	[redacted]	[redacted]	£65,593
Response-based analysis (B.1)[c]	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£12,783	0.84	[redacted]	[redacted]	£53,273
	Paclitaxel	£14,163	0.73	[redacted]	[redacted]	£44,877
	Cis+gem	£30,310	1.39	[redacted]	[redacted]	£103,186
	BSC	£8,811	0.59	[redacted]	[redacted]	£44,183
Response-based analysis using alternative TTE models for OS, PFS and TTD (B.3)[c]	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£12,452	0.77	[redacted]	[redacted]	£77,597
	Paclitaxel	£13,948	0.67	[redacted]	[redacted]	£67,608
	Cis+gem	£29,880	1.25	[redacted]	[redacted]	£143,923
	BSC	£8,662	0.55	[redacted]	[redacted]	£64,282

	Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	Nivolumab ICER (£/QALY)
Note: [a] results have been reproduced by the ERG, based on the economic model submitted by the company in their clarification response; [b] using the ERG base-case ; [c] using ERG base-case except the change to conventional, not response-based approach ERG = Evidence Review Group; ICER = incremental cost effectiveness ratio; QALY = quality-adjusted life year						

2. BACKGROUND

In this section the ERG provides a review of the evidence submitted by Bristol-Myers Squibb in support of nivolumab, trade name Opdivo® for the treatment of metastatic or unresectable UC after platinum-based chemotherapy. We outline and critique the company's description of the underlying health problem and the overview of current service provision. The information is taken mainly from Chapter B.1.3 of the company submission (CS) with sections referenced as appropriate.

2.1 Critique of company's description of underlying health problem.

The underlying health problem of this appraisal is metastatic or unresectable UC in adult patients who have received platinum-based chemotherapy.

The company described the origin of UC from the urothelium or epithelial lining of the urinary tract which extends from the renal pelvis to the ureter, bladder and proximal urethra. Urothelial cancer can also be known as transitional cell carcinoma. As described in Table 3 on staging, the bladder is the main organ that is affected. Indeed, the CS states that UC *'accounts for approximately 90% of all bladder cancer'.*[2]

Common presenting symptoms of UC include painless haematuria (blood in the urine), dysuria, frequency, urgency, feeling of incomplete voiding, and straining. In addition, urinary, bowel and sexual functions are affected and therefore impacts on overall health-related quality of life (HRQoL), daily life and sleeping patterns.

The CS states that '*Locally advanced and metastatic disease refers to tumours that have grown through the bladder wall and/or have spread to lymph nodes or other distant sites.'*[2]

The CS outlines the impact of advanced or metastatic UC on patients. This includes symptoms of disease such as limited mobility, abdominal, bone or pelvic pain, anorexia, wasting and pallor.

The CS states that *'UC is the 10th most common cancer in the UK, and is 3–4 times more commonly found in males than females.*[36] *In 2014, there were 9,021 patients newly diagnosed with UC in England and Wales, of which 7,307 (73%) were in males and 2,756 (27%) were in females. The disease is also more common in older adults, with more than half (54%) of UC cases in the UK each year diagnosed in patients aged 75 and over.*

The majority of patients with UC are diagnosed in stages I and II (62%), with approximately 20% diagnosed at the advanced, metastatic stage.[36,2]

In section B.1.3.4, the CS states that '*Based on available data from Cancer Research UK and expert clinician feedback, the number of patients in England and Wales eligible for treatment with nivolumab, as per the licensed indication for locally advanced unresectable or metastatic UC whose disease has progressed following platinum-containing chemotherapy, is estimated to be 894 patients.'*[2]

ERG comment:

The ERG checked the references cited by the company to support the statements made above and considered the company to have provided an appropriate description of the underlying health problem. In addition the ERG would like to indicate death and survival statistics. Around 10% will survive their

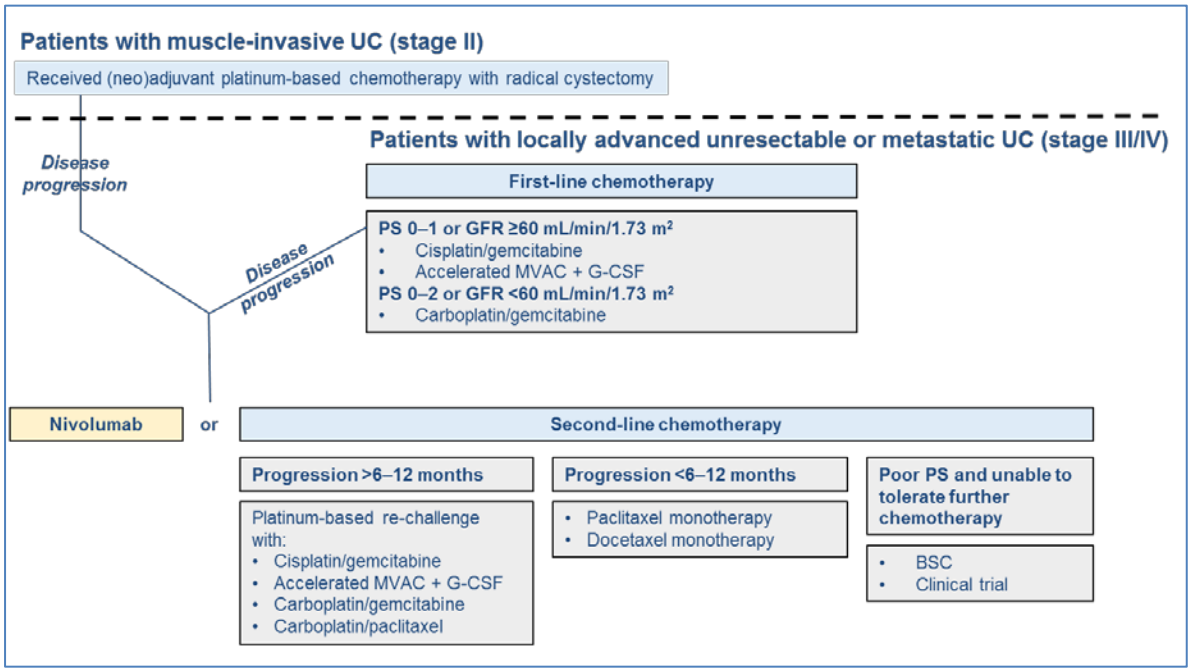
cancer for five years or more after diagnosis with T4 bladder cancer.[3] In 2014 there were 5,369 deaths from bladder cancer in the UK (3% of total cancer deaths).

The ERG notes that the projected numbers (894) eligible for nivolumab treatment were based on clinical expert opinion and could not be verified by the ERG, although the calculations for this figure (Table 56 of the CS) appear to be appropriate.

2.2 Critique of company's overview of current service provision

Figure 2.1 shows the CS current treatment pathway for persons with locally advanced or metastatic bladder cancer as well as the proposed position of nivolumab, based on NICE and EAU/ESMO guidelines and expert clinician feedback.[2]

Figure 2.1: Adapted treatment pathway to show potential position of nivolumab in the treatment of locally advanced or metastatic bladder cancer



Source: Figure 7 of CS

BSC = best supportive care; G-CSF = Granulocyte-colony stimulating factor; GFR = glomerular filtration rate; MVAC = methotrexate, vinblastine, doxorubicin and cisplatin; PS = performance status; UC = urothelial carcinoma

The company quote the NICE guidance for persons with locally advanced or metastatic bladder cancer. They state that '*For patients with locally advanced unresectable or metastatic UC whose condition has progressed after first-line therapy and who are physically fit [ECOG PS 0 or 1] with adequate renal function [GFR 60 ml/min/1.73 m^2], NICE recommends retreatment with cisplatin in combination with gemcitabine, or accelerated (high-dose) MVAC in combination with G-CSF. Patients for whom cisplatin-based chemotherapy is unsuitable (i.e. GFR <60 ml/min/1.73 m^2) may be treated with carboplatin plus paclitaxel in this setting.*'[2]

More specifically NICE guidance (NG2) states: 'Consider second-line chemotherapy with carboplatin in combination with paclitaxel or gemcitabine in combination with paclitaxel'.[4]

The company quote additional input from clinical experts.[5] Feedback from expert clinicians who were in UK clinical practice indicated that *'the vast majority of patients with locally advanced unresectable or metastatic UC following prior platinum-based chemotherapy would be treated with paclitaxel*

monotherapy, with docetaxel monotherapy also used in some centres. Of those patients considered fit enough to be offered second-line treatment with paclitaxel monotherapy, approximately one third to one half of these patients would typically refuse further chemotherapy treatment, and this figure may be even higher in some smaller centres. These patients would therefore currently opt for best supportive care (BSC), which may include painkillers, steroids and blood transfusions. Some patients would also be unsuitable for chemotherapy altogether, and would therefore be offered BSC instead of taxane-based chemotherapy.'[5]

In addition with reference to patients deemed physically fit, the expert clinicians added *'they would only consider retreatment with platinum-based chemotherapy for patients they considered fit enough and who had been progression-free for at least 9–12 months (or 6 months in some centres) following prior platinum-based chemotherapy; as such, this would very much be the minority of patients, representing only 5–10% of cases in the second-line setting.'*[5]

With reference to patients recommended for second line treatment of gemcitabine plus paclitaxel, the expert clinicians added that '*this regimen is used rarely in few centres across the UK and only for patients who have progressed quickly following first-line platinum chemotherapy and are very symptomatic'*[5]

The company suggest two potential positions for nivolumab in the treatment of for locally advanced or metastatic UC after failure of prior platinum-containing chemotherapy:[2]

1. In first-line locally advanced unresectable or metastatic disease, following disease progression after prior platinum-containing therapy received as (neo)adjuvant therapy with radical cystectomy in the muscle-invasive disease stage
2. In second-line unresectable or metastatic disease, following disease progression after prior platinum-containing therapy received in the locally advanced unresectable or metastatic disease stage.

ERG comment:
The company's description of the treatment pathway and options was based on existing NICE guidance (NICE guideline NG2; Bladder cancer: diagnosis and management) which is appropriate and relevant to the decision problem.[4] In particular the second-line treatment options for the management of locally advanced or metastatic bladder cancer were most relevant for the position of nivolumab in the treatment pathway. The company provided an adapted pathway based on inputs from clinical experts, this appears to be sensible, assuming the expert opinions are correct (this data could not be verified by the ERG as it is not in the public domain).

The ERG draws the attention of the committee to the potential placement of nivolumab at second-line for patients with locally advanced unresectable or metastatic UC, which is in accordance with the scope. However, the placement following progression subsequent to muscle-invasive disease (stage II) is not within scope.

The ERG notes the following ongoing appraisals relevant to the decision problem, as mentioned in the scope:[6]

Atezolizumab for treating metastatic urothelial bladder cancer after platinum-based chemotherapy [ID939] Publication expected September 2017.

Pembrolizumab for previously treated advanced or metastatic urothelial cancer (ID 1019) Publication expected October 2017.

3. CRITIQUE OF COMPANY'S DEFINITION OF DECISION PROBLEM

Table 3.1: Statement of the decision problem (as presented by the company)

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	ERG comment
Population (s)	Adults with metastatic or unresectable urothelial cancer whose disease has progressed after platinum-based chemotherapy	Adults with metastatic or unresectable urothelial cancer whose disease has progressed after platinum-based chemotherapy	NA	CheckMate 275 was in line with the scope of the decision problem, but no patients were included from the UK. CheckMate 032 included a small proportion of patients who had not received platinum-based chemotherapy; only 8% patients were from the UK.
Intervention	Nivolumab	Nivolumab	NA	CheckMate 275 investigated nivolumab, however CheckMate 032 investigated nivolumab monotherapy, but 23% switched to ipilimumab.
Comparator (s)	• Retreatment with first-line platinum-based chemotherapy (only for people whose disease has had an adequate response) • Paclitaxel • Docetaxel • Best supportive care	• Paclitaxel • Docetaxel • Best supportive care	No data on retreatment with first-line platinum-based chemotherapy was identified in the clinical systematic literature review (SLR). However, the use of retreatment is limited to <10% of patients and is not a primary comparator for nivolumab in UC after platinum-based chemotherapy. Data from a trial involving cisplatin plus gemcitabine after the failure of MVAC (methotrexate, vinblastine,	Both included trials were single arm studies and therefore no direct or indirect comparators were included. Given the paucity of data generally the ERG believes evidence for all specified NICE comparators should have been included in the STC.

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	ERG comment
			doxorubicin and cisplatin) was identified and included as a scenario analysis, in the absence of clinical data to inform a comparison of nivolumab versus retreatment.	
Outcomes	The outcome measures to be considered include: • overall survival • progression-free survival • response rates • adverse events of treatment • health-related quality of life	The outcome measures considered include: • overall survival • progression-free survival • response rates (objective response rate, duration of response) • adverse events of treatment • health-related quality of life (via the EORTC QLQ-C30 and the EQ-5D-3L)	N/A	The ERG notes that comparative data in the form of an STC was only provided for overall survival, progression free survival and objective response rate. There was no formal comparison for adverse events or quality of life.
Economic analysis	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year. The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared.	The cost effectiveness of treatments are expressed in terms of incremental cost per quality-adjusted life year. A lifetime time horizon was adopted to capture all relevant costs and health-related utilities. All costs and utilities were discounted at a rate of 3.5% per year in alignment with the NICE guide to the methods of technology appraisal.	N/A	N/A

	Final scope issued by NICE	Decision problem addressed in the company submission	Rationale if different from the final NICE scope	ERG comment
	Costs will be considered from an NHS and Personal Social Services perspective.	Costs were considered from an NHS and Personal Social Services perspective.		
Subgroups to be considered	If appropriate, the appraisal should include consideration of the costs and implications of additional testing for biological markers, but will not make recommendations on specific diagnostic tests or devices. Guidance will only be issued in accordance with the marketing authorisation. Where the wording of the therapeutic indication does not include specific treatment combinations, guidance will be issued only in the context of the evidence that has underpinned the marketing authorisation granted by the regulator.	No subgroup analysis was undertaken.	The effect of nivolumab in relation to baseline tumour PD-L1 expression status was investigated as part of the pivotal clinical trials informing the clinical evidence base for nivolumab within this submission. However, the link between baseline tumour PD-L1 expression status and the efficacy of PD-1/PD-L1 targeting agents is yet to be fully established and the testing methodologies of PD-L1 expression status are yet to be fully validated; as such, no formal subgroup analyses have been presented within this submission. This is in line with the marketing authorisation for nivolumab which is not restricted based on PD-L1 expression status.	The company was requested in the clarification letter to perform these subgroup analyses in the STC, but declined to do so arguing that data on PD-L1 expression was not available in the comparator trials.[7]
Special considerations including issues related to equity or equality	None detailed.	Treatment access being available only via clinical trials currently represents an inequality for some patients.	The availability of a nationally funded treatment option on the NHS would help to move towards addressing this equity issue.	No comment.

Source: CS, Table 1, page 11-13.

CR = complete response; N.A.= not applicable; ORR = objective response rate; PR = partial response; PD-L1: programmed death-ligand 1; STC simulated treatment comparison

3.1 Population

The population defined in the scope is: 'Adults with metastatic or unresectable urothelial cancer whose disease has progressed after platinum-based chemotherapy'.[6]

The licensed indication for nivolumab is: *'Nivolumab (Opdivo®) is indicated for the treatment of locally advanced unresectable or metastatic urothelial carcinoma in adults after failure of prior platinum-containing chemotherapy'* (CS, page 16).'[2]

The submission relies on two single arm studies, the CheckMate 275 trial[8] and the CheckMate 032 trial.[9] Examination of the inclusion criteria for these trials indicated that the CheckMate 275 trial included patients with metastatic or surgically unresectable transitional cell carcinoma of the urothelium (bladder, urethra, ureter, or renal pelvis). Patients have progression or recurrence after treatment with at least one platinum-containing chemotherapy regimen or within 12 months of peri-operative treatment with a platinum agent in the setting of cystectomy for localized muscle-invasive UC. Patients must have an ECOG performance status of 0 or 1.[10] Therefore the ERG considers this a good match with regards to the final scope. However, none of the patients included in this trial were from the UK.

CheckMate 032 included patients with histologically confirmed locally advanced or metastatic disease of one of the following tumour types: triple negative breast cancer, gastric cancer, pancreatic cancer, small cell lung cancer, bladder cancer, ovarian cancer. Patients must have an ECOG performance status of 0 or 1.[11] Prior chemotherapy was not stipulated as an inclusion criterion and reading Appendix 3.8 of the Checkmate 032 CSR indicated that a proportion of patients did not previously receive a platinum-based chemotherapy. For the purposes of the CS *'a subgroup of the enrolled population in this trial is of relevance to this submission: the cohort of patients enrolled to receive nivolumab monotherapy for the treatment of locally advanced unresectable or metastatic UC who had progressed after at least one previous line of platinum-containing chemotherapy (n=86).'*(CS section B.2.2)[2] In Table 5 of the CS, previous platinum based therapies are found in two of three inclusion criteria for progression or recurrence, the third criteria states 'refusal of standard treatment with chemotherapy'. Therefore it appears that not all patients are required to have had at least one line of platinum therapy. This is indicated further by Table 6 of the CS which indicates that a maximum of 60.2% of patients received prior systemic therapies. Therefore the subgroup of patients from CheckMate 032 used in the CS is not in accordance with the population defined in the scope. In addition, only 6/78 (8%) of bladder cancer patients in CheckMate 032 were from the UK.

3.2 Intervention

The intervention is in line with the scope. The intervention described in the scope is 'Nivolumab'. The CS describes the recommended dose and schedule of nivolumab monotherapy in urothelial carcinoma as follows: '3 mg/kg administered as IV infusion over 60 minutes every 2 weeks (Q2W), which is consistent with the existing approved dose and schedule of nivolumab monotherapy in adults in other indications.' (CS, page 17).[2] Dose escalation or reduction is not recommended; dosing delay or discontinuation may be required based on individual safety and tolerability.

A marketing authorisation application for nivolumab was submitted to the European Medicines Agency (EMA) on the 25 August 2016. A positive opinion from the Committee for Medicinal Products for Human Use (CHMP) was received on the 21 April 2017. Full marketing authorisation was received from the EMA on Monday 5 June 2017.[12]

In the CheckMate 275 trial, nivolumab (BMS-936558) was administered intravenously over 60 minutes at 3 mg/kg every two weeks until progression or unacceptable toxicity. This is in line with the decision problem.[10]

In the CheckMate 032 trial, patients were given nivolumab (3 mg/kg administered by intravenous infusion every two weeks) as monotherapy or in combination with ipilimumab. For the purposes of the CS only the nivolumab monotherapy patients were included, however they could switch to ipilimumab. Eighteen (23%) of 78 patients (receiving nivolumab monotherapy) switched to combination treatment with ipilimumab upon disease progression.[9] Therefore the ERG considers that the intervention in CheckMate 032 is not in line with the intervention described in the final scope.

3.3 Comparators

The NICE scope indicates four possible comparators: retreatment with first-line platinum-based chemotherapy (only for people whose disease has had an adequate response), paclitaxel, docetaxel and best supportive care. The company submission presents evidence for three comparators only: paclitaxel, docetaxel and best supportive care.

Both included nivolumab trials were single arm studies and therefore no direct or indirect comparators could be included. The company submission used a simulated treatment comparison (STC) to provide comparisons of nivolumab to paclitaxel, docetaxel and best supportive care; cisplatin plus gemcitabine were included only as part of a scenario analysis.[1] Cisplatin plus gemcitabine were only included in a scenario analysis because the company submission stated they had limited generalisability to the decision problem, the specific reasons given were:

1) patients '*had received MVAC in first-line treatment and are therefore not considered to be directly comparable to those receiving cisplatin plus gemcitabine retreatment in current UK clinical practice, as they are gemcitabine naïve'* (section B.2.9.1 CS) Gondo *et al.* (2011).[13]

2) inclusion of '*chemotherapy-naïve patients in addition to patients who had previously undergone first-line treatment*' (section B.2.9.1 CS) Ozawa *et al.* (2007).[14]

3) '*the two trials did not use the standard dosing regimen typically used for cisplatin plus gemcitabine in the UK'* (section B.2.9.1 CS)[2]

According to NICE guidelines [NG2] gemcitabine and cisplatin or MVAC and G-CSF can be given as both first line and second line treatments, for locally advanced and metastatic bladder cancer.[4] Also, whilst it is true that for one trial patients who were chemotherapy naïve were included,[14] this was not the trial that informed OS.[13] Therefore the ERG would not consider cisplatin and gemcitabine to be unsuitable for inclusion in the STC, especially given the limitations of the nivolumab and other comparator trials.

3.4 Outcomes

The company states that it assessed all the outcomes of the decision problem (overall survival, progression-free survival, response rates, adverse effects of treatment, health-related quality of life). However there were no direct or indirect comparators and the company submission used a STC to provide evidence of effectiveness to the comparators listed above. For the STC only three outcomes were considered; overall survival, PFS and ORR (section B.2.9 CS).[2]

There was no comparative data for adverse events or for quality of life. Note that adverse events and quality of life were reported for the two trials, but since these were single arm trials these results were

not informative. Adverse event data were provided in the response to clarification.[7] However, unlike for effectiveness, no evidence synthesis was performed for either of these two outcomes.

3.5 Other relevant factors

As stated by the company: 'A PAS [patient access scheme] is already in place with the Department of Health for inclusion in this technology appraisal, representing a simple discount of [redacted] on the list price of nivolumab' (CS, page 18).[2]

According to the company this STA fulfils the end-of-life criteria because:

- No studies identified in the SLR reported in Appendix D of the CS provided evidence of OS estimates for this patient population that approached 24 months.
- The economic analysis predicted mean life years (LYs) per patient with nivolumab of 2.78 years (33.36 months). In comparison, predicted mean LYs per patient with comparator therapies were 1.19 years (14.28 months) with paclitaxel, 1.40 years (16.80 months) with docetaxel and 1.01 years (12.12 months) with BSC. Nivolumab was therefore predicted to offer an extension to life of considerably greater than three months versus each of these comparators. Furthermore, in the context of the average survival of patients receiving paclitaxel, docetaxel or BSC, the survival gains offered by nivolumab represent a significant extension to life.

ERG comment: It appears that life expectancy is less than 24 months. However, given the absence of comparative trial data it is impossible to be confident of the extension to life resulting from treatment with nivolumab versus any of the comparators. The company bases the claim of extension to life on the economic model, which is informed by the STC, which attempts to estimate the treatment effect of nivolumab versus the comparators. However, as indicated in Section 4.3 and 4.4, the STC methods used to make the adjustment to reduce bias are not completely transparent, are accompanied by several limitations and are likely to result in residual bias (as argued in the methods guide followed by the company, NICE DSU TSD 18).[1] It is clear is that there is little difference in survival at least at the median between nivolumab (CheckMate 275 and CheckMate 032 trials)[10, 11] at 8.74 and 9.72 respectively and either docetaxel and paclicaxel, at 9.2 or 8 months respectively.[15, 16] The value for gemcitabine plus paclitaxel was even higher at 10.5 months.[13] It is true that the differences in these values are subject to potential bias given that the trial data represents observational data, but it is also true that the evidence provided by the STC to reduce this bias is far from clear.

4. CLINICAL EFFECTIVENESS

4.1 Critique of the methods of review(s)

The company conducted a systematic review to identify relevant direct and indirect clinical evidence on the use of nivolumab in metastatic or unresectable UC. This section critiques the methods of the review including searching, inclusion criteria, data extraction, quality assessment and evidence synthesis.

4.1.1 Searches

The following paragraphs contain summaries and critiques of all searches related to clinical effectiveness presented in the company submission. The Canadian Agency for Drugs and Technologies in Health (CADTH) evidence based checklist for the Peer Review of Electronic Search Strategies (PRESS), was used to inform this critique. [17] The submission was checked against the Single Technology Appraisal (STA) specification for company/sponsor submission of evidence. [18] The ERG has presented only the major limitations of each search strategy in the report.

The company submission stated that systematic review searches were undertaken in March 2017. Search strategies were reported in Appendix D of the CS for the following databases: Embase, MEDLINE, MEDLINE In-Process, MEDLINE Epub Ahead of Print and the Cochrane Library CENTRAL, DARE, NHS EED and HTA databases. In response to clarification the company confirmed that PubMed was not searched for this review and therefore should not have been listed in Appendix D.1.1.

Additional searches of the following conference proceedings were reported for the last four years: American Society of Clinical Oncology (ASCO), Genitourinary Cancers Symposium (GUCASYM), American Urological Association (AUA), European Association of Urology (EAU), European Society of Medical Oncology (ESMO).

The CS reported that bibliographies of eligible studies were searched for further relevant studies, and the reference lists of any systematic reviews and HTAs were scanned for further studies. ClinicalTrials.gov and WHO International Clinical Trials Registry Platform (WHO ICTRP) were also searched for ongoing clinical trials.

ERG comment:

- The database searches were clearly documented and reproducible, using a wide range of resources to identify published and unpublished literature. Database hosts and dates of searches were all reported. The database searches used combinations of indexing terms appropriate to the resource searched, free text and a number of synonyms for the condition. Study design filters were not applied.
- The search strategies contained some redundancy in their structure, but this will not have affected recall of studies.
- A typographical error in the Cochrane Library database searches noted by the ERG was amended, and searches were re-run by the company in response to clarification. No new relevant records were found.
- The ERG was concerned that limiting the MEDLINE and Embase clinical effectiveness searches to English language only studies may have introduced potential language bias. Current best practice states that 'Whenever possible review authors should attempt to identify and assess for eligibility all possibly relevant reports of trials irrespective of language of publication'.[19]

- All conference searches were conducted via Embase. The ERG has some concerns that relevant abstracts may have been omitted by searching using a biomedical database rather than directly searching conference proceedings, however this is unlikely to have affected the recall of relevant studies.

4.1.2 Inclusion criteria

A systematic literature review was conducted to identify clinical evidence on the efficacy and safety of nivolumab for the treatment of unresectable or metastatic UC. The full text documents were then assessed against the eligibility criteria by two independent reviewers, with disagreements adjudicated by a third reviewer.

The eligibility criteria used in the search strategy for clinical effectiveness are presented in Table 4.1.

Table 4.1: Eligibility criteria used in search strategy for clinical effectiveness

	Inclusion criteria	Exclusion criteria
Population	• Male and female adults aged 18 and over • Any ethnicity • Trials assessing patients with Stage III or Stage IV advanced, metastatic or unresectable urothelial carcinoma • Eligible patients must have progression or recurrence: o After treatment with at least 1 platinum-containing chemotherapy regimen for metastatic urothelial cancer or surgically unresectable locally advanced urothelial cancer, OR o Within 12 months of peri-operative (neo-adjuvant or adjuvant) treatment with platinum agent in the setting of cystectomy for localised muscle-invasive urothelial cancer • Trials with mixed populations of patients receiving first and second line treatment will only be eligible if results are reported separately for second line treatment or if more than 50% of the population are receiving second line treatment	• Paediatric population • Patients with Stage I or II urothelial carcinoma • Patients undergoing first-line treatment • Trials without a defined population • Trials with an unclear population
Interventions	• Retreatment with platinum-based chemotherapy (e.g. cisplatin plus gemcitabine, accelerated MVAC (methotrexate, vinblastine, adriamycin/doxorubicin and cisplatin), carboplatin plus gemcitabine or carboplatin plus paclitaxel) • Gemcitabine plus paclitaxel • Docetaxel monotherapy • Paclitaxel monotherapy • Gemcitabine monotherapy	

	Inclusion criteria	Exclusion criteria
	• Vinblastine monotherapy • Vinflunine monotherapy • Best supportive care	
Comparators	• Placebo • Any intervention of interest • Any other treatment that may facilitate an indirect comparison • Best supportive care	
Outcomes	• Overall survival (OS) • Progression-free survival (PFS) or time to tumour progression (TTP) • Objective response rate (ORR) • Complete response (CR) • Partial response (PR) • Duration of response (DoR) • Treatment-related adverse event (AEs): o Rates of overall Grade 3 or 4 AEs o Rates of specific Grade 3 or 4 AEs including: 1. Neutropenia 2. Anaemia 3. Thrombocytopenia 4. Febrile Neutropenia 5. Asthenia (Fatigue) 6. Nausea 7. Vomiting 8. Diarrhoea 9. Pruritus 10. Pneumonia 11. Lung infiltration 12. Alanine aminotransferase (ALT) increase 13. Hepatitis • Discontinuation/withdrawals due to AE • Health-related quality of life (HRQoL)	No outcomes of interest
Study design	• Randomised controlled trials • Non-randomised prospective controlled clinical trials or single-arm trials • Systematic reviews – will be eligible for reference checking only • Conference abstracts only to provide supplementary information	• Retrospective trials • Case reports • Case series of fewer than 5 people • Editorials, letters or news articles • Conference abstracts – as the primary trial reference
Language restrictions	English language only	Non-English

	Inclusion criteria	Exclusion criteria
Publication year	NR	NR
Source: CS, Table 7, pages 39-40		

ERG comment:

- The population of the systematic review is in line with the scope.
- The interventions and comparators for the inclusion criteria are appropriate for identifying treatments to facilitate a network analysis of nivolumab versus the comparators of the scope. A separate review for nivolumab only does not appear to have been performed. It is noticeable that nivolumab is not included as an intervention; the ERG assumes this is an oversight by the company given that nivolumab studies are included.
- All the outcomes outlined in the decision problem were included; however the company has limited the inclusion of adverse events to those that are grade 3 or 4. This will preclude assessment of 'all adverse events'.
- Randomised controlled trials, non-randomised controlled trials and single arm trials were all included in the review.

4.1.3 Critique of data extraction

According to Appendix D.1.4 of the CS data extraction was '*carried out by two independent reviewers with disagreements adjudicated by a third reviewer*'.[20]

ERG comment: The ERG believes that overall the data extraction was carried out appropriately.

4.1.4 Quality assessment

According to Appendix D.1.5 of the CS quality assessment was '*carried out by two independent reviewers with disagreements resolved through discussion with a third reviewer*'.[20]

Quality assessment was performed for prospective cohort trials using the CRD Cohort Trial Checklist (reference 21 of the CS) and for randomised controlled trials using the guidance of the Centre for Review and Dissemination (reference 22 of the CS).

There were 12 trials included in the STC. Two single arm studies were identified for nivolumab; both were open label and single arm studies. The remainder trials were a mix of randomised controlled trials or single arm studies.

For the quality assessment of the randomised controlled trial the following domains were assessed: randomisation, allocation concealment, comparability of groups, blinding, drop out, selective reporting of outcomes and use of intention to treat analysis and appropriate methods for dealing with missing data (summarised in Table 14, D.1.5. of the CS)

Cohort studies are classed as a comparison of outcomes between a group of participants who have received an intervention and a group who have not. This is clearly not appropriate for a single arm study. For the quality assessment of cohort studies the following domains were assessed: comparability of groups, were the groups assessed at similar time points of disease progression, was the intervention reliably ascertained, comparable confounding variables, adequate adjustment of confounding variables, was a dose response relationship between intervention and outcome demonstrated, blinding, adequate follow-up, proportion of the cohort followed up, comparable drop-out rates. (Summarised in Table 13, D.1.5. of the CS). From this list it is clear that most questions are concerned with the comparability between groups, thereby illustrating that this risk of bias tool is not appropriate for the single arm studies

identified within the CS. Single arm studies are by definition low down in the hierarchy of study design and therefore the quality for these studies is low to start with and risk of bias tools have not been widely developed for this study design.

ERG comment: Study quality appeared to be appropriately assessed for randomised trials but not for the single arm studies (which include those for nivolumab). However, risk of bias has to be deemed to be high for all data used in the STC given that only single arms were used.

4.1.5 Evidence synthesis

According to the company, '*Data from CheckMate 275 and CheckMate 032 were pooled in the context of the STC presented in Section B.2.9 and Appendix D'* (CS, section B.2.8, page 59).[20] However, no methods are presented for the pooling of results, and results themselves have not been reported either. We asked the company to provide details of the statistical method(s) used for pooling the data from Checkmate 275 and CheckMate 032 and to explain which data were used (BIRC or investigator-assessed)**.** We also asked the company to conduct pooled analyses using data from each method separately.[21]

In the response to the clarification letter, the company did not state how the two nivolumab trials were pooled. They did clarify that the BIRC method was chosen for CheckMate 275 and only the investigator-assessed results were available for CheckMate 032.[7] They also stated the following on page 26 of the response: *'As agreed with the ERG on the preliminary teleconference to discuss the clarification questions, analyses using each method separately have not been provided.'* However, no such agreement was made. The ERG continues to believe that results derived from performing the STC twice using a) only BIRC or b) only investigator-led methods would provide valuable insight into the variability of the data. Given that the BIRC method was only available for CheckMate 275 this would imply using only the CheckMate 275 data for STC. This was suggested to the company during the teleconference but the analysis was not provided.

4.2 Critique of trials of the technology of interest, their analysis and interpretation (and any standard meta-analyses of these)

The company conducted a systematic literature review to identify relevant clinical evidence on the efficacy and safety of nivolumab for the treatment of unresectable or metastatic urothelial carcinoma. Two trials investigating nivolumab were found: CheckMate 275[8, 10] and CheckMate 032[9, 11].

An overview of CheckMate 275 and CheckMate 032 is provided in Table 4.2.

Table 4.2: Clinical effectiveness evidence for nivolumab

Study	CheckMate 275 (NCT02387996)	CheckMate 032 (NCT01928394)
Publications (primary reference in bold)	**Sharma *et al.* (2017)[8]** Clinical study report[10]	**Sharma *et al.* (2016)[9]** Clinical study report[11]
Study design	Multicentre, open-label, single-arm phase II study	Multicentre, open-label, two-stage, multi-arm, phase I/II[a]
Population	Patients with locally advanced unresectable or metastatic UC who had progressed or recurred after at least one previous line of platinum-containing chemotherapy (N=270)	Patients with locally advanced unresectable or metastatic UC who had progressed or recurred after treatment with at least one platinum-

				containing chemotherapy regimen (N=78)		
Intervention(s)	Nivolumab (IV 3 mg/kg Q2W)			Nivolumab (IV 3 mg/kg Q2W)		
Comparator(s)	N/A (single-arm)			N/A[a]		
Indicate if trial supports application for marketing authorisation	Yes	Indicate if trial used in the economic model	Yes	Yes	**Indicate if trial used in the economic model**	Yes
Reported outcomes specified in the decision problem	ORR OS PFS HRQoL via the European Organisation for Research and Treatment of Cancer (EORTC) general cancer module (QLQ-C30) and the EuroQoL-5 dimensions-3 levels (EQ-5D-3L) questionnaires Adverse events (AEs)			ORR OS PFS EQ-5D-3L AEs		
All other reported outcomes	Duration of response and additional safety outcomes			Duration of response and additional safety outcomes		

Source: CS, Table 4, pages 27-28

[a]CheckMate 032 investigated nivolumab or nivolumab combined with ipilimumab in patients with UC, triple-negative breast cancer, gastric cancer, pancreatic adenocarcinoma, small cell lung cancer, and ovarian cancer. Here, presentation of CheckMate 032 refers only to the nivolumab monotherapy UC cohort (n=86) of relevance to this submission.

BIRC = blinded independent review committee; CSR = clinical study report; EORTC QLQ-C30 = European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30; EQ-5D-3L = 3-level EuroQoL 5-Dimensions; HRQoL: health-related quality of life; IV= intravenous; N/A = not applicable; ORR = objective response rate; OS = overall survival; PD-L1 = programmed death ligand 1; PFS = progression-free survival; Q2W = every two weeks; UC = urothelial carcinoma.

4.2.1 Study design and methodology of the nivolumab studies

CheckMate 275

CheckMate 275 is an ongoing, phase II single-arm clinical trial investigating the efficacy and safety of nivolumab in patients with locally advanced unresectable or metastatic UC who had failed at least one previous line of therapy.[8]

Patients with histologically confirmed metastatic or surgically unresectable UC with disease progression or recurrence after at least one platinum-based chemotherapy were enrolled and assigned to a cohort according to tumour PD-L1 expression status (PD-L1 ≥5%, PD-L1 < 5%, or indeterminate). Enrolment in the trial continued until approximately 70 patients with confirmed PD-L1 expression of ≥5% were treated. Enrolment continued further in Japan until approximately 25 Japanese patients were treated, or until November 2015, whichever occurred sooner.

Enrolled patients were treated with IV nivolumab 3mg/kg Q2W until documented disease progression (based on RECIST v1.1 criteria) and clinical deterioration, unacceptable toxicity, or other protocol-defined reasons. Treatment beyond initial investigator-assessed RECIST v1.1-defined progression was

permitted if the patient had an investigator-assessed clinical benefit, did not have rapid disease progression, and was tolerating the study drug.

The primary endpoint of CheckMate 275 was objective response rate (ORR) based on Blinded Independent Review Committee (BIRC) assessment using RECIST v1.1 in the all-treated population, in patients with PD-L1 expression ≥1%, and in patients with PD-L1 expression ≥5%. Objective response was defined as the proportion of patients with best overall response of confirmed complete response (CR) or partial response (PR) assessed by the BIRC. Time to response and duration of response were estimated in patients with a confirmed CR or PR. Responses were confirmed at the second scan at least four weeks after criteria for objective response were met.

The trial consisted of three phases: screening, treatment, and follow-up. Treated patients were evaluated for response according to the RECIST v1.1 guidelines beginning eight weeks (±1 week) after the first dose of nivolumab and then every eight weeks (±1 week) thereafter up to 48 weeks, then every 12 weeks (±1 week) until disease progression (investigator-assessed RECIST v1.1-defined progression) or treatment discontinuation, whichever occurred later. Patients were followed for OS every three months until death, lost to follow-up, or withdrawal of study consent.

CheckMate 032

CheckMate 032 is an ongoing phase I/II multi-arm trial investigating the efficacy and safety of nivolumab monotherapy or nivolumab in combination with ipilimumab in patients with one of the following tumour types: UC, triple-negative breast cancer, gastric cancer, pancreatic adenocarcinoma, small cell lung cancer, and ovarian cancer.[9] The company used a subgroup of patients enrolled in this study in their analyses: the cohort of patients enrolled to receive nivolumab monotherapy for the treatment of locally advanced unresectable or metastatic UC who had progressed after at least one previous line of platinum-containing chemotherapy (n=86). Therefore, reference to CheckMate 032 in the CS refers only to this subgroup of UC patients.[9]

A total of 86 patients were enrolled in the nivolumab monotherapy treatment arm of CheckMate 032, of whom 78 patients received at least one dose of nivolumab. All 78 patients who received at least one dose of nivolumab were included in the safety and efficacy analyses. The subgroup of UC patients included in the company analyses (N=78) does include 18 patients who crossed-over to nivolumab in combination with ipilimumab.

Eligible patients with histologically or cytologically confirmed carcinoma of the renal pelvis, ureter, bladder, or urethra and disease progression after at least one previous platinum-based chemotherapy treatment were treated with IV nivolumab 3 mg/kg Q2W until documented disease progression (based on RECIST v1.1 criteria), unacceptable toxicity, or other protocol-defined reasons.

The primary endpoint of CheckMate 032 was the proportion of patients with a confirmed investigator-assessed objective response, defined as the number of patients with a best overall response of a CR or PR as per the RECIST v1.1 criteria divided by the number of treated patients. Patients were evaluated for response at baseline, six weeks after the first dose of nivolumab, continuing every six weeks for the first 24 weeks, and then every 12 weeks until disease progression or treatment discontinuation, whichever occurred later. Patients receiving nivolumab monotherapy could switch to nivolumab plus ipilimumab (nivolumab 1 mg/kg and ipilimumab 3 mg/kg or nivolumab 3 mg/kg and ipilimumab 1 mg/kg intravenously, every three weeks for four cycles) following disease progression if they met prespecified criteria.

For a CR or PR to be judged to be a best overall response, the assessment needed to be confirmed by a second scan no less than four weeks after the criteria for response was first met. Patients who did not meet response-evaluable criteria (i.e. at least one target lesion at baseline and at least one on-study assessment) were judged to be not assessable. Treatment beyond initial investigator-assessed RECIST v1.1-defined progression was permitted if the patient had an investigator-assessed clinical benefit and was tolerating the study drug.

A summary of the methodology and trial design of CheckMate 275 and CheckMate 032 is presented in Table 4.3. Further details of the methodology of CheckMate 275 and CheckMate 032, including the full eligibility criteria can be found in Appendix M of the CS.

ERG comment: The main problem with the design of the nivolumab trials is the absence of a comparator arm. No analysis can estimate the influence of bias in any outcome in these single arm trials in comparison to the outcomes of other comparator trials.

Table 4.3: Summary of CheckMate 275 and CheckMate 032 study methodology

Trial name	**CheckMate 275 (n=270)**	**CheckMate 032 (n=78)**
Location	International: 63 sites across 11 countries in North America (USA), Europe, Australia and Asia	International: 16 sites in 5 countries: Finland, Germany, Spain, UK and USA
Trial design	Multicentre, open-label, single-arm phase II study	Multicentre, open-label, multi-arm, phase I/II study[b]
Eligibility criteria for participants	Key inclusion criteria Males and females ≥18 years of age with an ECOG PS 0 or 1 Histologically or cytologically confirmed metastatic or surgically unresectable transitional cell carcinoma of the urothelium involving the bladder, urethra, ureter, or renal pelvis Measurable disease by CT or MRI per RECIST v1.1 criteria Progression or recurrence after treatment With at least 1 platinum-containing chemotherapy regimen for metastatic or surgically unresectable locally advanced urothelial cancer, or Within 12 months of peri-operative (neo-adjuvant or adjuvant) treatment with platinum agent in the setting of cystectomy for localised muscle-invasive urothelial cancer Patients that had received more than 2 prior lines of chemotherapy must not have had liver metastases Availability of tumour samples for PD-L1 expression analysis[a] Previous palliative radiotherapy must have been completed at least 2 weeks before administration of the study drug	Key inclusion criteria Males and females ≥18 years of age with an ECOG PS 0 or 1 Measurable disease by CT or MRI per RECIST v1.1 criteria Locally advanced or metastatic urothelial cell carcinoma Progression or recurrence After at least 1 previous platinum-containing chemotherapy treatment for metastatic or locally advanced unresectable urothelial cancer, or Recurrence within 1 year of completing previous platinum-based neoadjuvant or adjuvant treatment After previously refusing standard treatment with chemotherapy for the treatment of metastatic (stage IV) or locally advanced disease Key exclusion criteria Active brain metastases or leptomeningeal metastases Any serious or uncontrolled medical disorder History of or active, known or suspected autoimmune disease (vitiligo, type 1 diabetes mellitus, residual hypothyroidism caused by auto immune thyroiditis, and disorders not expected

Trial name	**CheckMate 275 (n=270)**	**CheckMate 032 (n=78)**
	Key exclusion criteria Active brain or leptomeningeal metastases Active, known or suspected autoimmune disease Previous malignancy active within the previous 3 years (except locally curable cancers that appeared to have been cured or carcinoma in situ) Any serious or uncontrolled medical disorder Autoimmune disease (vitiligo, type 1 diabetes mellitus, residual hypothyroidism due to an autoimmune condition only requiring hormone replacement, psoriasis not requiring systemic treatment, or conditions not expected to recur in the absence of an external trigger were permitted) Systemic treatment with either corticosteroids (>10 mg daily prednisone equivalents) or other immunosuppressive medications within 14 days of first study drug administration Prior treatment with an anti-PD-1, anti-PD-L1, anti-PD-L2, anti-CTLA-4 antibody, anti-CD137, or any other antibody or drug specifically targeting T-cell co-stimulation or immune checkpoint pathways Treatment with any chemotherapy, radiation therapy, biologics for cancer, or investigational therapy within 28 days of first study drug administration All toxicities attributed to previous anticancer therapy other than neuropathy, alopecia, and fatigue must have resolved to grade 1 or baseline before administration of study drug. A full list of inclusion and exclusion criteria is presented in Appendix M.	to recur in the absence of an external trigger were permitted) Need for immunosuppressive doses of systemic corticosteroids (>10 mg daily prednisone equivalents) for at least 2 weeks before study drug administration Prior treatment with experimental anti-tumour vaccines or any modulator of T-cell function or checkpoint pathway A full list of inclusion and exclusion criteria is presented in Appendix M.
Settings and locations where the data were collected	The study was conducted in a secondary care (hospital) setting at 63 sites across 11 countries worldwide The study was conducted in accordance with Good Clinical Practice guidelines by qualified investigators using a single protocol to promote consistency across sites	The study was conducted in a secondary care (hospital) setting at 16 sites across 5 countries worldwide The study was conducted in accordance with Good Clinical Practice guidelines by qualified investigators using a single protocol to promote consistency across sites

Trial name	**CheckMate 275 (n=270)**	**CheckMate 032 (n=78)**
Method of study drug administration	Nivolumab 3mg/kg Q2W via IV infusion over 60 minutes Treatment was continued until progression, unacceptable toxicity, or withdrawal of consent Patients were permitted to continue treatment beyond investigator-assessed RECIST v1.1-defined progression if they were experiencing a clinical benefit, as determined by the investigator, and were tolerating the study drug No dose modifications were allowed, but predefined dose delays were permitted for adverse events	Nivolumab 3mg/kg Q2W via IV infusion over 60 minutes Treatment was continued until progression, unacceptable toxicity, or withdrawal of consent. Patients were permitted to continue treatment beyond investigator-assessed RECIST v1.1-defined progression if they were experiencing a clinical benefit, as determined by the investigator, and were tolerating the study drug Patients could switch to nivolumab plus ipilimumab (nivolumab 1 mg/kg and ipilimumab 3 mg/kg or nivolumab 3 mg/kg and ipilimumab 1 mg/kg intravenously, every 3 weeks for four cycles) after progression if they met pre-specified criteria.
Permitted and disallowed concomitant medication	The following medications were prohibited during the study: Immunosuppressive agents (except to treat a drug-related adverse events) or systemic corticosteroids (>10 mg daily prednisone equivalent) within 14 days of study drug administration[b] Any antibody or drug specifically targeting T-cell co-stimulation or checkpoint pathways, or chemotherapy, radiation therapy, biologics for cancer, or investigational therapy within 28 days of first study drug administration	The following medications were prohibited during the study: Immunosuppressive agents (except to treat a drug-related adverse event) Systemic corticosteroids >10 mg daily prednisone equivalent[b] Any concurrent antineoplastic therapy (i.e. surgery, chemotherapy, hormonal therapy, immunotherapy, radiation therapy except for palliative radiation therapy described above or standard or investigational agents for treatment of cancer) Supportive care for disease-related symptoms was permitted to be offered to all patients on the trial. Palliative (limited-field) radiation therapy and palliative surgical resection were permitted if the certain protocol-defined criteria were met.
Primary endpoint	The primary endpoint of CheckMate 275 was BIRC-assessed ORR (as per RECIST v1.1) in the all-treated population, in patients with PD-L1 expression ≥1%, and in patients with PD-L1 expression ≥5% ORR was defined as the number of patients with a best overall response (BOR) of confirmed complete response (CR) or partial response (PR) divided by the number of all-treated patients, PD-L1 ≥1% patients or PD-L1 ≥5% subjects, respectively	The primary endpoint of CheckMate 032 was confirmed investigator-assessed ORR ORR was defined as the number of patients with a BOR of CR or PR as per RECIST v1.1 divided by the number of treated patients

Trial name	**CheckMate 275 (n=270)**	**CheckMate 032 (n=78)**
Secondary and exploratory endpoints	**Secondary endpoints:** BIRC-assessed PFS OS Investigator-assessed ORR (in the all-treated population, patients with PD-L1 expression ≥1%, and patients with PD-L1 expression ≥5%) **Exploratory endpoints:** Investigator-assessed PFS Safety HRQoL via the EORTC QLQ-C30 questionnaire General health status via the EQ-5D-3L questionnaire Pharmacokinetics and exploration of exposure-response relationships* Immunogenicity* Pharmacodynamic activity in the peripheral blood and tumour tissue as measured by flow cytometry, immunohistochemistry, soluble factor analysis, and gene expression (microarray technology, quantitative RT-PCR)* Association between biomarkers in the peripheral blood and tumour tissue with safety and efficacy* **Outcomes not considered relevant to present in this submission*	**Secondary endpoints:** Investigator-assessed PFS OS DOR Safety **Exploratory endpoints:** Assessed by PD-L1 expression (≥1% and <1%): ORR OS PFS HRQoL via the EQ-5D and EQ-VAS questionnaires
Timing of assessments	Tumour assessments were scheduled at 8 weeks from the date of first dose (±1 week), then every 8 weeks (±1 week) thereafter up to 48 weeks, then every 12 weeks (±1 week) until documented disease progression or treatment discontinuation (whichever occurred last). Assessments were performed using CT or MRI and included the pelvis, chest, abdomen and all known sites of disease Survival assessment was scheduled every 3 months until death, lost to follow-up or withdrawal of study consent AEs were assessed during treatment visits and were included in the safety analyses if they occurred within 30 days from the day of the last dose received	Treated subjects were evaluated for response by the investigator according to the RECIST v1.1 at baseline and then every 6 weeks (±1 week) from first dose for the first 24 weeks, then every 12 weeks (±1 week) until disease progression or treatment was discontinued (whichever occurred later) Assessments were performed using CT or MRI and included the pelvis, chest, abdomen and all known sites of disease AEs were assessed during treatment visits. Safety was defined as the incidence of treatment-related adverse events leading to drug discontinuation within the first 12 weeks of treatment in patients who had at least one dose of study drug HRQoL was assessed before study drug administration through Week 13, then

Trial name	**CheckMate 275 (n=270)**	**CheckMate 032 (n=78)**
	HRQoL and general health status were assessed before each dose at Week 1, then every 8 weeks up to 48 weeks, then every 12 weeks until disease progression or treatment discontinuation (whichever occurred later) Two follow-up visits and subsequent survival follow-up visits were also scheduled for AEs and HRQoL measures[c]	at the same time of subsequent tumour assessments, during Follow-Up Visit 1 and 2 and survival visits Two follow-up visits and subsequent survival follow-up visits were also scheduled (AEs and HRQoL)[c]
Pre-planned subgroups	A pre-planned analysis of the primary and secondary endpoints in patients with PD-L1 expression <1% and ≥1% was conducted Further subgroup analyses were conducted to assess the impact of pre-specified baseline characteristics, site of original tumour origin (bladder, renal pelvis/ureter), number of Bellmunt risk factors, and prior cancer therapy regimens (number of prior regimens in a metastatic setting, time from completion of most recent prior regimen to study treatment) on confirmed ORR per BIRC, PFS and OS	As part of the exploratory endpoints, ORR, OS and PFS were analysed in subgroups defined by PD-L1 expression (<1% and ≥1%). In addition, ad-hoc subgroup analyses were conducted to assess the impact several key baseline factors such as ECOG-PS, metastases, or haemoglobin on investigator-assessed ORR
Duration of study and follow-up	The first patient was treated on the 9th March 2015 and the trial is currently ongoing. The last patient last visit date for the primary database lock of the 30th May 2016, data from which are presented in this submission, was the 15th April 2016. The median follow-up for OS was 11.5 months. A further database lock took place on 2nd September 2016 and data from this are also presented in this submission.	The first patient was treated on the 5th June 2014 and the trial is currently ongoing. The last patient last visit date for the primary database lock of 24th March 2016 was the 11th February 2016, data from which are presented in this submission. The median follow-up for OS was 9.69 months.

Source: CS, Table 5, pages 30-35

[a]Patients were required to have an evaluable tumour tissue sample for PD-L1 expression testing at screening, but were not excluded based on PD-L1 status. [b]Several advanced or metastatic solid tumour types were studied in CheckMate 032, but only the urothelial carcinoma arm treated with nivolumab monotherapy is presented in this submission. [c]Patients were followed for at least 100 days after the last dose of study drug. Follow-up Visit 1 was scheduled for 35 days from the last dose ±7 days or coincided with the date of discontinuation (± 7 days) if date of discontinuation was >35 days after last dose. Follow-up Visit 2 was scheduled for 80 days (±7 days) from follow-up Visit 1. Survival follow-up visits were scheduled for every 3 months (± 7 days) from Follow-up Visit 2.

AEs = adverse events; BIRC = blinded independent review committee; BOR = best overall response; CR = complete response; CT = computer tomography; CTLA-4 = cytotoxic T-lymphocyte-associated protein 4; ECOG = Eastern Cooperative Oncology Group; EORTC QLQ-C30 = European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Core 30; EQ-5D-3L = 3-level EuroQoL 5-Dimensions; GCP = Good Clinical Practice; HRQoL = health-related quality of life; IV = intravenous; MRI = magnetic

Trial name	**CheckMate 275 (n=270)**	**CheckMate 032 (n=78)**
resonance imaging; ORR = objective response rate; OS = overall survival; PD-1 = programmed death 1; PD-L1 = programmed death ligand 1; PD-L2 = programmed death ligand 2; PFS = progression-free survival; PR = partial response; PROs = patient-reported outcomes; PS = performance status; RECIST = response evaluation criteria in solid tumours.		

4.2.2 Baseline characteristics of the nivolumab studies

Baseline demographics, disease characteristics and a summary of prior therapies of the patients included in CheckMate 275 and CheckMate 032 are presented in Table 4.4.

In CheckMate 275, median age was 66 years, the majority of patients were white and male, and over 70% were current or former smokers. The vast majority of patients (96.7%) had metastatic disease. Overall 71.5% of patients had received at least one prior regimen in the metastatic disease setting, and 29.3% had received two or more prior regimens for metastatic disease. Prior systemic cancer therapy was less common in the neoadjuvant and adjuvant settings, with 22.2% receiving at least one neoadjuvant regimen and 30.7% of patients receiving prior regimen(s) in the adjuvant setting.

The median age of the patient population in CheckMate 032 was 66 years; the majority were white (92.3%) and male (69.2%). The vast majority (91%) of patients had metastatic (stage IV) disease, and 75.6% of patients had at least two disease sites.

The company provided the following additional information based on feedback from clinical experts: *'Expert clinician feedback was that the patient populations of CheckMate 275 and CheckMate 032 were very similar, and could be considered generally representative of the patient population expected to receive nivolumab in UK clinical practice. Across both trials, expert clinician feedback was that the proportion of patients with PS 0 was perhaps slightly over-representative of the number of patients likely to have PS 0 in this setting, and that the median age of the patients in both trials may be slightly lower than the age of the average UC patient treated in the second-line setting in UK clinical practice. However, a recent chart review conducted in UK clinical practice of patients with locally advanced unresectable or metastatic UC initiating second-line therapy found that the mean patient age was in fact very similar, albeit slightly lower (mean of 62.8 years), than in both CheckMate trials.'*[5]

In response to the clarification request, the company stated that there were no UK sites in CheckMate 275 and in CheckMate 032, there were 6 patients (7.7%) treated in the study in the UK.[7]

ERG comment:
There are serious questions regarding the representativeness of the nivolumab trial patients to the UK population. Firstly, almost no patients in the UK were included and none in the largest trial (CheckMate 275).[10] Secondly, in response to the clarification request, the company confirmed that as few as 18.8% of patients in the UK might have and ECOG PS of 0, as opposed to over 50% in the two nivolumab trials.[7] Thirdly, there is a mismatch in terms of prior therapies, as confirmed in Table 8 of the response to clarification, which shows that, in a chart review, as many as over 75% of patients in the UK would have previously taken a gemcitabine platinum-based combination compared to fewer than 40% in the trials.[7] Finally, there is a question of the applicability to those with locally advanced unresectable as opposed to metastatic disease given the very small proportion of such patients in the trials. The company stated in the response to clarification that type of disease in these terms was not prognostic given no mention of this at their advisory board. However, lack of comment at the advisory board does not mean that clinical experts do not believe this to be the case.

Table 4.4: Baseline characteristics of patients in the all-treated population of CheckMate 275 and CheckMate 032

	CheckMate 275	CheckMate 032
Characteristic	**Total (n=270)**	**Total (n=78)**
Demographics		
Age, median years (range)	66 (38–90)	66 (31–85)
Age categorisation, n (%)		
<65	122 (45.2)	37 (47.4)
≥65 and <75	110 (40.7)	31 (39.7)
≥75 and <85	35 (13.0)	N/A
≥75	N/A	10 (12.8)
>85	3 (1.1)	N/A
Male, n %	211 (78.1)	54 (69.2)
Race, n %		
White	231 (85.6)	72 (92.3)
Asian	30 (11.1)	1 (1.3)
Black	2 (0.7)	4 (5.1)
Other	3 (1.1)	1 (1.3)
Not reported	4 (1.5)	N/A
Region, n (%)		
US	106 (39.3)	59 (75.6)
Japan	23 (8.5)	0 (0.0)
Rest of world	141 (52.2)	19 (24.4)
Tobacco use, n (%)		
Current/former smoker	194 (71.9)	48 (61.5)
Never smoked	67 (24.8)	29 (37.2)
Unknown	9 (3.3)	1 (1.3)
Disease characteristics		
ECOG PS, n (%)		
0	145 (53.7)	42 (53.8)
1	124 (45.9)	36 (46.2)
3	1 (0.3)	0
Bellmunt risk factors, n (%)		
0	98 (36.3)	27 (34.6)
1	111 (41.1)	39 (50.0)
2	46 (17.0)	8 (10.3)
3	15 (5.6)	4 (5.1)
Site of primary tumour, n (%)		
Urinary bladder	197 (73.0)	NR
Renal pelvis	46 (17.0)	NR

	CheckMate 275	CheckMate 032
Characteristic	**Total (n=270)**	**Total (n=78)**
Ureter	19 (7.0)	NR
Urethra	8 (3.0)	NR
Disease setting, n (%)		
Metastatic	261 (96.7)	71 (91.0)
Locally unresectable/non-metastatic	9 (3.3)	7 (9.0)
Baseline metastases, n (%)		
Any visceral involvement	227 (84.1)	61 (78.2)
Liver	75 (27.8)	20 (25.6)
Lymph node only	43 (15.9)	13 (16.7)
PD-L1 expression, n (%)		
Assessable	N/A	67 (85.9)
<1%	N/A	42 (53.8)
≥1%	124 (45.9)	25 (31.8)
<5%	N/A	53 (67.9)
≥5%	83 (30.7)	14 (17.9)
Number of sites with ≥1 lesion, n (%)		
1	85 (31.5)	19 (24.4)
2	94 (34.8)	30 (38.5)
3	51 (18.9)	24 (30.8)
4	29 (10.7)	3 (3.8)
≥5	11 (4.1)	2 (2.6)
Prior therapy		
Prior systemic therapy regimen setting, n (%)		
Metastatic	193 (71.5)	N/A
Adjuvant	83 (30.7)	33 (42.3)
Neo-adjuvant	60 (22.2)	14 (17.9)
Previous therapies in metastatic setting, n (%)		
0	77 (28.5)	N/A
1	114 (42.2)	26 (33.3)
2	57 (21.2)	N/A
2-3	N/A	42 (53.8)
>3	N/A	10 (12.8)
≥3	22 (8.1)	N/A
Prior surgery related to cancer, n (%)	250 (92.6)	71 (91.0)
Prior radiotherapy, n (%)	85 (31.5)	25 (32.1)

Source: CS, Table 5, pages 35-37
ECOG PS = Eastern Cooperative Oncology Group performance status; N/A = not applicable; NR = not reported; PD-L1 = programmed death ligand 1.

4.2.3 Statistical analyses in the nivolumab studies

The statistical analyses used for the primary and secondary endpoints alongside sample size calculations and methods for handling missing data are summarised in Table 4.5.

ERG comment: The ERG believes that the statistical methods used within the nivolumab studies were appropriate. The ERG notes that the primary design of CheckMate 275 was to evaluate ORR based on assessments of nivolumab monotherapy in patients with tumour expressing PD-L1 (membranous staining in ≥ 5% and ≥ 1% tumour cells) and overall patients. CheckMate 32 was primarily designed to evaluate the ORR of nivolumab monotherapy in patients with advanced or metastatic UC. Neither study design was appropriate for comparative analysis.

Table 4.5: Statistical methods for the primary analysis of CheckMate 275 and CheckMate 032

Trial name	CheckMate 275	CheckMate 032
Hypothesis objective	Treatment with nivolumab monotherapy would lead to clinical benefit in patients with metastatic or surgically unresectable UC who have progressed post platinum treatment as demonstrated by a clinically meaningful ORR	Treatment with nivolumab monotherapy will have clinical activity in subjects with advanced or metastatic tumours
Statistical analysis	ORRs (both BIRC- and investigator-assessed) were summarised by a binomial response rate and their corresponding two-sided 95% exact CIs using the Clopper-Pearson method.**[**CS REF 45**]** BOR was summarised by response category Median values of DOR were calculated along with two-sided 95% CI using Brookmeyer and Crowley method.**[**CS REF 46**]** TTR was summarised using descriptive summary statistics for the responders Time-to-event distributions were estimated using Kaplan-Meier techniques. This was done for PFS, OS and DOR (note that time to response was analysed using summary statistics such as mean, SD, median, min, max). Median survival time along with 95% CIs were constructed based on a log-log transformed CI for the survivor function S(t)**[**CS REF 46+47**]** Rates at fixed time points were derived from the Kaplan-Meier estimate and corresponding confidence interval were derived based on Greenwood formula**[**CS REF 48**]** for variance derivation and on log-log transformation applied on the survivor function S(t)**[**CS REF 49**]**	ORR was summarised by a binomial response rate and corresponding two-sided 95% exact CI using the Clopper-Pearson method. Time-to-event distributions (DOR, PFS and OS) were estimated using Kaplan-Meier techniques When appropriate, the median along with 95% CI was provided using Brookmeyer and Crowley methodology (using the log-log transformation for construction of CIs). Rates at fixed time points (e.g. OS at 12 months) were derived from the Kaplan Meier estimate along with their corresponding log-log transformed 95% CIs.
Sample size, power calculation	The primary objective was to estimate ORR as per BIRC assessment for: • All treated patients • Patients with PD-L1 expression ≥1% • Patients with PD-L1 expression ≥5%	The primary objective was to estimate investigator-assessed ORR An ORR of 10% or less was considered not of clinical value, and an ORR of 25% or greater was considered of strong clinical interest A sample size of 60–100 treated subjects would provide 90% to 97% power to reject the null hypothesis of 10% response rate if

Trial name	CheckMate 275	CheckMate 032
	For all treated patients, a sample size of 242 would provide 90% power to reject the null hypothesis that ORR was 10% at a two-sided 5% type I error if the true ORR in this population was 16.9%. Assuming ORR is 30%, 70 treated patients with PD-L1 expression ≥5% would provide 99.1% power at 5% type 1 error to reject the null hypothesis of a two-sided test that the true ORR was 10%, based on historical control data for single-agent chemotherapy,**[**CS REF 34, 35, 50**]** a threshold below which was considered not clinically meaningful in this population, and 90% power at 5% type I error to reject the null hypothesis of a two-sided test that the true ORR was 14.7%. Under the assumption of 32% prevalence rate of PD-L1 ≥5% among all PD-L1 evaluable patients, approximately up to 220 PD-L1 evaluable patients would be treated. Assuming an additional 10% of treated patients with PD-L1 indeterminate status, the total sample size was expected to be approximately 242. Under the assumption of 50% prevalence rate of PD-L1 ≥1% among all PD-L1 evaluable patients, approximately up to 110 patients with PD-L1 expression ≥1% would be treated. This would provide 90% power to reject the null hypothesis that ORR was 10% at a two-sided 5% type 1 error if the true ORR in this population was 20.6%.	the true response rate was 25% with a two-sided Type I error rate of 5%
Data management, patient withdrawals	The final analysis of the primary endpoint ORR (based on BIRC assessments) was to be performed six months after approximately 70 patients with PD-L1 expression of ≥5% had been treated (i.e. six months after last patient first treatment)	All 78 patients who received at least one dose of nivolumab were included in the safety and efficacy analyses

Source: CS, Table 5, pages 38-40

BOR = best overall response; CI = confidence interval; ORR = overall response rate; PD-L1 = programmed death ligand 1; TTR = time to response.

4.2.4 Quality assessment of the nivolumab studies

The company considered the CheckMate 275 and CheckMate 032 studies to be of satisfactory quality based on the CRD cohort study checklist.[22]

ERG comment: The ERG considers both studies as low-level evidence in the hierarchy of clinical study designs, and not suitable for comparisons with other interventions.

4.2.5 Results of the nivolumab studies

CheckMate 275

The primary endpoint in Checkmate 275 was ORR (based on BIRC assessments) and the primary database lock was 30 May 2016. The company responded to the clarification request by stating that the next database locks for CheckMate 275 and CheckMate 032 in [redacted] and [redacted], respectively.[7]

Treatment with nivolumab led to a confirmed objective response per blinded independent review committee (BIRC) in a total of 52 (19.6%) patients (95% CI: 15.0 to 24.9) and 6 (2.3%) patients achieved a complete response (CR) (see Table 4.6). Patients in the PD-L1≥1% cohort achieved an objective response rate (ORR) of 23.8% (95% CI: 16.5 to 32.3) and patients with <1% PD-L1 expression had a confirmed ORR of 16.1% (15.8% at the second database lock).

As reported in Sharma et al. (2017),[8] 177 high-quality gene expression profiles have been generated from patients' tumour tissues. Higher values of the 25-gene interferon-γ signature were associated with a greater proportion of responders to nivolumab and higher PD-L1 expression. Patients with high interferon-γ signature were more likely to respond to nivolumab than were those with low interferon-γ signature (p=0. 0003).

Time to response (TTR) and duration of response (DOR) were estimated in patients with a confirmed partial response (PR) or complete response (CR). Median TTR as per BIRC was 1.87 months (interquartile range (IQR): 1.81 to 1.97 months) and the majority of responders achieved their response at the time of first tumour assessment (Week 8).

At the time of the clinical database lock (30 May 2016), median DOR as per BIRC had not been reached in the efficacy-treated population and across the <1% and ≥1% PD-L1 subgroups. The majority of responders (76.9%) were still continuing to respond and [redacted] of patients had a DOR of at least three months.

Table 4.6: Primary efficacy results of CheckMate 275

Tumour response	Efficacy-treated population (n=265)	PD-L1 <1% (n=143)	PD-L1 ≥1% (n=122)
ORR, n (%) **95% CI**	52 (19.6) 95% CI: 15.0–24.9	23 (16.1) 95% CI: 10.5–23.1	29 (23.8) 95% CI: 16.5–32.3
BOR			
CR	6 (2.3)	1 (0.7)	5 (4.1)
PR	46 (17.4)	22 (15.4)	24 (19.7)
SD	60 (22.6)	25 (17.5)	35 (28.7)
PD	104 (39.2)	67 (46.9)	37 (30.3)
Unable to determine[a]	49 (18.5)	28 (19.6)	21 (17.2)
Median TTR (n=52), months; IQR	1.87	1.94	1.87

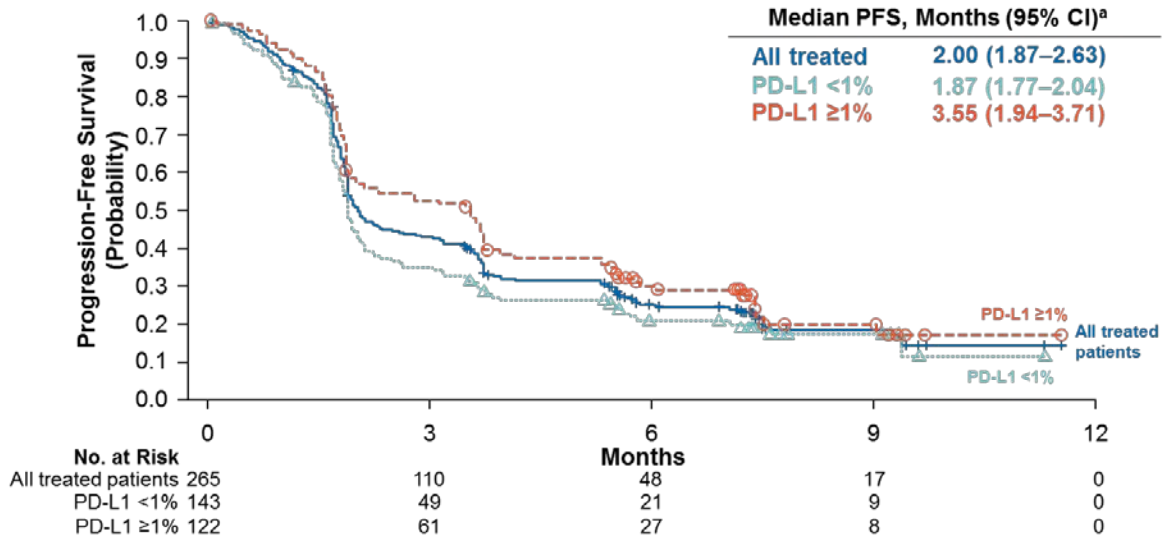
Tumour response	Efficacy-treated population (n=265)	PD-L1 <1% (n=143)	PD-L1 ≥1% (n=122)
	IQR: 1.81–1.97	IQR: 1.81–2.10	IQR: 1.81–1.97
Median DOR (n=52), months; 95% CI	NR 95% CI: 7.43–NR	NR 95% CI: 7.43–NR	NR 95% CI: 7.52–NR

Source: CS, Table 12, page 43-44
[a]BOR was reported as unable to determine in 49 patients (18.5%); main reasons were because the patient had died or started subsequent therapy before the first scan visit at Week 8.
BOR = best overall response; CI = confidence intervals; CR = complete response; DOR = duration of response; IQR = interquartile range; ORR = objective response rate; PD = progressive disease; PD-L1 = programmed death ligand 1; PR = partial response; SD = stable disease; TTR = time to response NR = not reached.

At the time of the primary clinical database lock (30 May 2016), 201 patients (75.8%) had experienced a PFS event. Median PFS in the efficacy-treated population was 2.00 months (95% CI: 1.87 to 2.63), and the PFS rates at three and six months were 43.1% (95% CI: 37.0 to 49.1) and 25.2% (95% CI: 20.0 to 30.8), respectively. Median PFS for patients in the PD-L1 ≥1% cohort was longer than in the all-treated population at 3.55 months (95% CI: 1.94 to 3.71), and in the PD-L1 <1%, median PFS was 1.87 months (95% CI: 1.77 to 2.04) (see Figure 4.1).

Results for investigator-assessed ORR were investigated as a secondary outcome and the results were consistent with BIRC-assessed ORR. A total of [redacted] patients ([redacted]) achieved an objective response of which [redacted] patients ([redacted]) achieved a CR.

Figure 4.1: Kaplan-Meier plot for progression-free survival in CheckMate 275

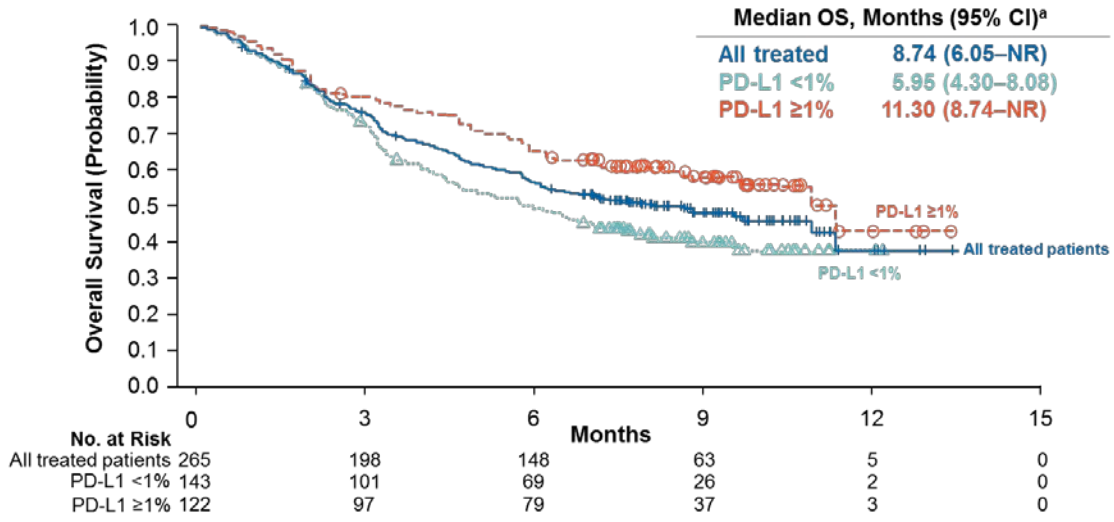


Source: CS, Figure 11, page 46
CI = confidence interval; PD-L1 = programmed death ligand 1; PFS = progression-free survival.

Median follow-up for overall survival (OS) (time between first dose and last known date alive or death) was 7.00 months (IQR: 2.96 to 8.77 months). At the primary analysis database lock (30 May 2016), 138 patients (51.1%) had died. Median OS in the efficacy-treated population was 8.74 months (95% CI: 6.05 to N/A); three-month and six-month OS rates were 75.8% (95% CI: 70.2 to 80.5) and 57.0% (95% CI: 50.7 to 62.7).

The Kaplan-Meier plot for OS is presented in Figure 4.2. Median OS for patients in the PD-L1 ≥1% cohort was longer than in the all-treated population at 11.30 months (95% CI: 8.74 to NR), and in the PD-L1 <1%, median OS was 5.95 months (95% CI: 4.30 to 8.08).

Figure 4.2: Kaplan-Meier plot for overall survival in CheckMate 275



Source: CS, Appendix E, Figure 26, page 146
CI = confidence interval; OS = overall survival; PD-L1 = programmed death ligand 1.

Results from the second database lock of CheckMate 275 (2 September 2016) were consistent with those from the primary analysis database lock in terms of ORR, PFS and OS. In total, 54 patients (20.0%) had achieved an ORR (95% CI: 15.4 to 25.3), and two more patients had achieved a CR. Median DOR was 10.35 months (95% CI: 7.52 to NR). A further six patients had died, taking the total to 154 (57%). A comparison of the main results between database locks and trials is shown in Table 11 of the CS and reproduced in Table 4.7. There also continued to be a statistically signification difference in median OS between PD-L1 <1% and PD-L1 >= 1% (5.95 months (95% CI: 4.37 to 8.08), and in the PD-L1 <1%, median OS was 11.63 months (95% CI: 9.10 to NA).

Table 4.7: Overview of clinical effectiveness results from CheckMate 275 and CheckMate 032

Outcome	CheckMate 275		CheckMate 032
	Initial database lock: 30 May 2016 n=265[c]	Latest database lock: 2 Sep 2016 n=270[c]	n=78
ORR, n (%), [95% CI]	52 (19.6), [15.0–24.9]	54 (20.0), [15.4–25.3][b]	19 (24.4) [15.3–35.4]
TTR, median (IQR), months	1.87 (1.81–1.97)[a]	1.94 (1.84–2.50)[b]	1.48 (1.25–4.14)
DOR, median (95% CI), months	NR (7.43–NR)[a]	10.35 (7.52–NR)[b]	NR (9.92–NR)
PFS, median (95% CI), months	2.00 (1.87–2.63)[a]	2.00 (1.87–2.63)[b]	2.78 (1.45–5.85)
OS, median (95% CI), months	8.74 (6.05–NR)[a]	8.57 (6.05–11.27)[b]	9.72 (7.26–16.16)
Source: CS, Table 11, page 43			

[a]Minimum follow-up of 6 months from the date of first dose. [b]Minimum follow-up of 8.3 months. [C]Follow-up for the latest database lock was sufficient to include 5 patients from Japan who were not included in efficacy analyses in the initial database lock.
CI = confidence intervals; DOR = duration of response; NR = not reached.ORR = objective response rate; OS = overall survival; PFS =progression free survival; TTR = time to response

Patient-reported outcomes data for the measurement of HRQoL was assessed via the EORTC QLQ-C30 questionnaire and the EQ-5D-3L questionnaire in CheckMate 275. Due to the limited study follow-up, interpretations of EORTC QLQ-30 results are limited to the first 41 weeks of follow-up for the all-treated population. Overall, patient HRQoL continued to increase or was maintained throughout the trial from baseline to Week 41.

The mean baseline EQ-5D VAS score was 60.2, and mean scores were higher at Week 9 on treatment (67.5). By Week 41, the average EQ-5D VAS was more than 80 points. However, by this time data was based on only n=24 patients.

A total of [redacted] of patients received ≥90% of the planned nivolumab dose intensity, and the median number of doses received was [redacted] (range: [redacted]). The median duration of therapy was [redacted] months. At the time of the 30 May 2016 database lock, 75.6% of patients had discontinued treatment with nivolumab. The most common reasons for discontinuation were disease progression (53.3%), AEs unrelated to nivolumab (12.6%), and nivolumab toxicity (5.2%).

A summary of the safety results from CheckMate 275 and CheckMate 032 is presented in Table 4.8. The majority of treated patients experienced at least one AE regardless of causality, during treatment with nivolumab or within 30 days of the last nivolumab dose. As of their respective clinical database locks, a total of 138 (51.5%) patients and 36 (46.2%) patients in the CheckMate 275 and CheckMate 032 trials had died, respectively. The proportion of deaths due to study drug toxicity was 1.1% and 3%, respectively. All-cause AEs leading to treatment discontinuation were reported in 20.7% and 7.7% of patients in CheckMate 275 and CheckMate 032, respectively.

Table 4.8: Summary of safety analysis in CheckMate 275 and CheckMate 032

Adverse event, n (%)	**CheckMate 275 (n=270)[a]**		**CheckMate 032 (n=78)[b]**	
Deaths	138 (51.1)		36 (46.2)	
Deaths due to study drug toxicity	3 (1.1)[c]		2 (2.6)[d]	
	Any grade	Grade 3-4	Any grade	Grade 3-4
All causality AEs	267 (98.9)	137 (50.7)	78 (100)	43 (55.1)
Drug-related AEs	174 (64.4)	48 (17.8)	65 (83.3)	18 (23.1)
All-causality serious AEs	147 (54.4)	99 (36.7)	36 (46.2)	23 (29.5)
Drug-related serious AEs	[redacted]	[redacted]	8 (10.3)	[redacted]
All-causality AEs leading to treatment discontinuation	56 (20.7)	42 (15.6)	6 (7.7)	4 (5.1)
Drug-related AEs leading to treatment discontinuation	13 (4.8)	8 (3.0)	2 (2.6)	2 (2.6)

Source: CS, Table 23, page 72-73
[a] AEs were coded using the MedDRA version 19.0 and were graded for severity according to the NCI CTCAE version 4.0. [b] AEs were coded using the MedDRA version 18.1 and were graded for severity according to the NCI CTCAE version 4.0. C Three deaths (Grade 5 pneumonitis, Grade 5 acute respiratory failure, and Grade

5 cardiovascular failure) were judged as study drug-related. [d] Two deaths (Grade 4 pneumonitis and Grade 4 thrombocytopenia) were assessed as study drug-related.
AEs = adverse events; MedDRA = Medical Dictionary for Regulatory Activities; NCI CTCAE = National Cancer Institute Common Terminology Criteria for Adverse Events; SAEs = serious adverse events.

Treatment-related adverse events occurred in 174 (64%) of 270 patients. The most common treatment-related adverse event of any grade was fatigue, which was noted in 45 patients (17%). Grade 3 or 4 treatment-related adverse events occurred in 48 patients (18%) – most commonly grade 3 fatigue and diarrhoea, each of which occurred in five patients (Table 4.9). Thirteen patients (5%) discontinued treatment because of nivolumab toxicity, including four (1%) from pneumonitis, two (1%) from pemphigoid, and one each (<1%) from dyspnoea, interstitial lung disease, maculopapular rash, pruritic rash, abdominal pain, diarrhoea, and circulatory collapse. The most common treatment-related select (immuno mediated) adverse events (any grade) were skin (47 [17%]) and endocrine (39 [14%]). Most select adverse events resolved and were manageable with immune-modulating drugs (mostly systemic corticosteroids; data not shown). Some drug-related endocrinopathies were not deemed to be resolved because of ongoing hormone replacement therapy.[8]

Of the 270 patients in the safety population, 138 deaths (51%) were reported, of which 121 (88%) were due to disease progression. Of the 53 patients who died within 30 days of their last nivolumab dose, 39 (74%) died of disease progression. Of the 14 deaths not related to disease progression, 11 were attributed to other reasons and three were attributed by investigators to treatment, all of which occurred in patients with metastatic disease. One patient died of pneumonitis, one of acute respiratory failure, and one of cardiovascular failure.[8]

Table 4.9: Drug-related adverse events in ≥5% patients in CheckMate 275 and CheckMate 032

Adverse event	CheckMate 275 (n=270)		CheckMate 032 (n=78)	
	Any grade	Grade 3-4	Any grade	Grade 3-4
Total patients with an event	174 (64.4)	48 (17.8)[a]	65 (83.3)	18 (23.1)[b]
General disorders and administration site conditions	80 (29.6)	10 (3.7)	29 (37.2)	2 (2.6)
Fatigue	45 (16.7)	5 (1.9)	28 (35.9)	2 (2.6)
Asthenia	16 (5.9)	4 (1.5)	N/A	N/A
Pyrexia	15 (5.6)	0 (0.0)	N/A	N/A
Gastrointestinal disorders	54 (20.0)	7 (2.6)	24 (30.8)	2 (2.6)
Diarrhoea	24 (8.9)	5 (1.9)	7 (9.0)	0 (0.0)
Nausea	19 (7.0)	1 (0.4)	10 (12.8)	1 (1.3)
Skin and subcutaneous tissue disorders	54 (20.0)	6 (2.2)	34 (43.6)	3 (3.8)
Pruritus	25 (9.3)	0 (0.0)	23 (29.5)	0 (0.0)
Rash	16 (5.9)	3 (1.1)	5 (6.4)	0 (0.0)
Rash maculo-papular	N/A	N/A	14 (7.9)	2 (2.6)
Dry skin	N/A	N/A	5 (6.4)	0 (0.0)
Investigations	N/A	N/A	26 (33.3)	8 (10.3)
Lipase increased	N/A	N/A	11 (14.1)	4 (5.1)
Amylase increased	N/A	N/A	7 (9.0)	3 (3.8)
Lymphocyte count decreased	N/A	N/A	5 (6.4)	2 (2.6)

Adverse event	CheckMate 275 (n=270)		CheckMate 032 (n=78)	
	Any grade	Grade 3-4	Any grade	Grade 3-4
Blood creatinine increased	N/A	N/A	4 (5.1)	0 (0.0)
Endocrine disorders	31 (11.5)	1 (0.4)	6 (7.7)	0 (0.0)
Hypothyroidism	21 (7.8)	0	4 (5.1)	0 (0.0)
Musculoskeletal and connective tissue disorders	N/A	N/A	13 (16.7)	1 (1.3)
Arthralgia	N/A	N/A	9 (11.5)	0 (0.0)
Metabolism and nutrition	27 (10.0)	3 (1.1)	10 (12.8)	2 (2.6)
Decreased appetite	22 (8.1)	0	5 (6.4)	0 (0.0)
Hyperglycaemia	N/A	N/A	5 (6.4)	1 (1.3)
Blood and lymphatic system disorders	N/A	N/A	11 (14.1)	1 (1.3)
Anaemia	N/A	N/A	8 (10.3)	0 (0.0)
Respiratory, thoracic and mediastinal disorders	N/A	N/A	11 (14.1)	1 (1.3)[b]
Dyspnoea	N/A	N/A	6 (7.7)	2 (2.6)
Nervous system disorders	N/A	N/A	7 (9.0)	0 (0.0)

Source: CS, Table 23, page 74-75

[a]Grade 5 events reported in 3 (1.1%) patients (1 death due to pneumonitis, 1 death due to acute respiratory failure, 1 death due to cardiovascular failure). [b] 1 (1.3%) Grade 5 drug-related AE (pneumonitis).

AEs = adverse events; N/A = not applicable.

Select AEs were defined as AEs of special clinical interest that are potentially associated with the use of nivolumab, and were identified based on the following principles:

- AEs that may differ in type, frequency, or severity from AEs caused by non-immunotherapies
- AEs that may require immunosuppression (e.g. Corticosteroids) as part of their management
- AEs whose early recognition and management may mitigate severe toxicity
- AEs for which multiple event terms may be used to describe a single type of AE, thereby necessitating the pooling of terms for full characterisation.

Considering the AEs already observed across other studies of nivolumab therapy, the AEs considered as select AEs were endocrinopathies, diarrhoea/colitis, hepatitis, pneumonitis, interstitial nephritis, rash and hypersensitivity/infusion reactions.

Most select AEs were considered drug-related by the investigator, with the exception of hepatic and renal events, where a lower proportion of select AEs were deemed to be drug-related. The most frequently reported any-grade drug-related select AE categories were skin (17.4%) and endocrine (14.4%) – see Table 4.10.

Table 4.10: Drug-related select adverse events in CheckMate 275 and CheckMate 032

Select adverse event, n (%)	CheckMate 275		CheckMate 032	
	Any grade	Grade 3–4	Any grade	Grade 3–4
Total patients with an event, by category				
Skin	47 (17.4)	4 (1.5)	33 (42.3)	2 (2.6)
Endocrine	39 (14.4)	1 (0.4)	6 (7.7)	0 (0.0)
Gastrointestinal	25 (9.3)	6 (2.2)	8 (10.3)	1 (1.3)
Hepatic	10 (3.7)	5 (1.9)	4 (5.1)	1 (1.3)
Pulmonary	11 (4.1)	3 (1.1)	2 (2.6)	0 (0.0)
Renal	3 (1.1)	1 (0.4)	7 (9.0)	1 (1.3)
Hypersensitivity/infusion reactions	3 (1.1)	1 (0.4)	2 (2.6)	0 (0.0)
Drug-related 'select' AEs, by category				
Skin				
Pruritis	25 (9.3)	0 (0.0)	23 (29.5)	0 (0.0)
Rash	16 (5.9)	3 (1.1)	5 (6.4)	0 (0.0)
Rash maculo-papular	4 (1.5)	1 (0.4)	14 (17.9)	2 (2.6)
Erythema	2 (0.7)	0 (0.0)	N/A	N/A
Pruritis generalised	2 (0.7)	1 (0.4)	N/A	N/A
Rash macular	2 (0.7)	0 (0.0)	N/A	N/A
Rash pruritic	2 (0.7)	1 (0.4)	N/A	N/A
Rash erythematous	N/A	N/A	2 (2.6)	0 (0.0)
Rash papular	N/A	N/A	1 (1.3)	0 (0.0)
Palmar-plantar erythrodysaesthesia syndrome	N/A	N/A	1 (1.3)	0 (0.0)
Blister	1 (0.4)	0 (0.0)	N/A	N/A
Dermatitis	1 (0.4)	0 (0.0)	N/A	N/A
Eczema	1 (0.4)	0 (0.0)	N/A	N/A
Rash generalised	1 (0.4)	0 (0.0)	N/A	N/A
Skin exfoliation	1 (0.4)	0 (0.0)	N/A	N/A
Skin irritation	N/A	N/A	1 (1.3)	0 (0.0)
Urticaria	1 (0.4)	0 (0.0)	N/A	N/A
Endocrine				
Thyroid disorder	35 (13.0)	0 (0.0)	6 (7.7)	0 (0.0)
Hypothyroidism	21 (7.8)	0 (0.0	4 (5.1)	0 (0.0)
Hyperthyroidism	11 (4.1)	0 (0.0	3 (3.8)	0 (0.0)
Blood thyroid stimulating hormone increased	10 (3.7)	0 (0.0	1 (1.3)	0 (0.0)
Blood thyroid stimulating hormone decreased	5 (1.9)	0 (0.0	N/A	N/A
Thyroiditis	2 (0.7)	0 (0.0	N/A	N/A
Thyroxine increased	2 (0.7)	0 (0.0	N/A	N/A

Select adverse event, n (%)	CheckMate 275		CheckMate 032	
	Any grade	Grade 3–4	Any grade	Grade 3–4
Autoimmune thyroiditis	1 (0.4)	0 (0.0	N/A	N/A
Thyroxine decreased	1 (0.4)	0 (0.0	N/A	N/A
Thyroxine free increased	1 (0.4)	0 (0.0	N/A	N/A
Adrenal disorder	2 (0.7)	0 (0.0)	N/A	N/A
Adrenal insufficiency	2 (0.7)	0 (0.0)	N/A	N/A
Pituitary disorder	2 (0.7)	1 (0.4)	N/A	N/A
Hypophysitis	2 (0.7)	1 (0.4)	N/A	N/A
Diabetes	1 (0.4)	0 (0.0)	N/A	N/A
Type I diabetes mellitus	1 (0.4)	0 (0.0)	N/A	N/A
Gastrointestinal				
Diarrhoea	24 (8.9)	5 (1.9)	7 (9.0)	0 (0.0)
Colitis	2 (0.7)	1 (0.4)	1 (1.3)	1 (1.3)
Hepatic				
Alanine aminotransferase increased	8 (3.0)	2 (0.7)	3 (3.8)	0 (0.0)
Aspartate aminotransferase increased	6 (2.2)	3 (1.1)	1 (1.3)	1 (1.3)
Blood alkaline phosphatase increased	3 (1.1)	2 (0.7)	1 (1.3)	0 (0.0)
Blood bilirubin increased	2 (0.7)	1 (0.4)	1 (1.3)	0 (0.0)
Liver function test increased	2 (0.7)	1 (0.4)	N/A	N/A
Transaminases increased	2 (0.7)	0 (0.0)	N/A	N/A
Hyperbilirubinaemia	1 (0.4)	0 (0.0)	N/A	N/A
Pulmonary				
Pneumonitis	10 (3.7)	2 (0.7)	2 (2.6)	0 (0.0)
Interstitial lung disease	1 (0.4)	1 (0.4)	N/A	N/A
Renal				
Acute kidney injury	1 (0.4)	0 (0.0)	1 (1.3)	1 (1.3)
Blood creatinine increased	1 (0.4)	1 (0.4)	4 (5.1)	0 (0.0)
Renal failure	1 (0.4)	0 (0.0)	N/A	N/A
Blood urea increased	N/A	N/A	3 (3.8)	0 (0.0)
Hypersensitivity/infusion reactions				
Infusion related reaction	2 (0.7)	1 (0.4)	1 (1.3)	0 (0.0)
Hypersensitivity	1 (0.4)	0 (0.0)	1 (1.3)	0 (0.0)

Source: CS, Table 26, page 76-78
Includes events reported between first dose and 30 days after last dose of study therapy.
AEs = adverse events; N/A = not applicable.

CheckMate 032

An overview of the primary efficacy results (primary database lock: 24 March 2016) from the UC cohort of CheckMate 032 is presented in Table 4.11. A confirmed investigator-assessed objective response was achieved in 19 (24.4%) patients (95% CI: 15.3 to 35.4) of 78 treated patients, with five patients (6%) achieving a CR and 14 patients (18%) achieving a PR.

Patients in the PD-L1≥1% cohort achieved an objective response rate (ORR) of 24.0% and patients with <1% PD-L1 expression had a confirmed ORR of 26.2%.

Table 4.11: Overview of clinical effectiveness results from CheckMate 032

Tumour response	Nivolumab (n=78)	PD-L1 <1% (n=42)	PD-L1 ≥1% (n=25)
ORR, n (%)	19 (24.4) [95% CI 15.3–35.4]	11 (26.2)	6 (24.0)
BOR, n (%)			
CR	5 (6.4)	1 (2.4)	4 (16.0)
PR	14 (17.9)	10 (23.8)	2 (8.0)
SD	22 (28.2)	11 (26.2)	8 (32.0)
PD	30 (38.5)	18 (42.9)	8 (32.0)
Unable to determine	7 (9.0)	2 (4.8)	3 (12.0)
Median TTR, months (IQR)	1.48 (1.25–4.14)	[redacted]	[redacted]
Median DOR, months (95% CI)	NR (9.92–NR)	[redacted]	[redacted]

Source: CS, Table 15, page 51; and CS, Appendix E, Table 56, page 148
BOR = best overall response; CI = confidence intervals; CR = complete response; DOR = duration of response; IQR = interquartile range; ORR = objective response rate; PD = progressive disease; PD-L1 = programmed death ligand 1; PR = partial response; SD = stable disease; TTR = time to response NR = not reached.

The Kaplan-Meier plots for PFS and OS in CheckMate 032 are presented in Figure 4.3 and Figure 4.4.

Median PFS was 2.78 months (95% CI 1.45 to 5.85) and 60 (77%) of 78 patients had disease progression or died by data cut-off. Of 18 (23.1%) censored patients, [redacted] had their PFS time censored on either the date of last on-study tumour assessment or date of last assessment prior to subsequent anti-cancer therapy. The most common reason for censoring among these patients was [redacted]. PFS rates (95% CI) were [redacted] at three months, [redacted] at six months and 20.8% (12.3 to 30.9) at 12 months.

Median PFS for patients in the PD-L1 ≥1% cohort was longer than in the all-treated population at 5.45 months (95% CI: 1.41–11.17), and in the PD-L1 <1%, median PFS was 2.76 months (95% CI: 1.41–6.51)

Figure 4.3: Kaplan-Meier plot for progression-free survival in subgroups of CheckMate 032

FIGURE REDACTED

Source: CS, Appendix E, Figure 27, page 148
CI = confidence interval; PD-L1 = programmed death ligand 1.

Median OS was 9.7 months (95% CI 7.3 to 16.2) and 46 (59%) of 78 patients had died at the time of data cut-off. OS rates (95% CI) were [redacted] at three months, [redacted] at six months, and 45.6% (34.2 to 56.3) at 12 months. Median follow-up for OS (time between dose date and last known date alive or death) for all nivolumab monotherapy treated UC patients was 9.69 months (range: 0.7 to 20.7 months).

Median OS for patients in the PD-L1 ≥1% cohort was longer than in the all-treated population at 16.16 months (95% CI: 7.59 to N.A.), and in the PD-L1 <1%, median OS was 9.89 months (95% CI: 7.03 to N.A.)

Figure 4.4: Overall survival in subgroups of CheckMate 032

FIGURE REDACTED

Source: CS, Appendix E, Figure 28, page 149
CI = confidence interval; PD-L1 = programmed death ligand 1.

Patient-reported outcomes data for the measurement of HRQoL was assessed via the EQ-5D-3L questionnaire in CheckMate 032. A total of 73 (93.5%) UC patients treated completed the EQ-5D VAS questionnaire at baseline and the mean baseline EQ-5D VAS score was 72.4 (SD 24.5). Overall, the mean EQ-5D VAS score increased over time. By Week 19, clinically meaningful improvements (>7-point change from baseline) were reported and the average EQ-5D VAS score was >80 points. The EQ-5D VAS continued to improve through Week 61. After week 61, the sample size was too small to interpret (<10).

In CheckMate 032, the majority ([redacted]) of patients received ≥90% of the planned nivolumab dose intensity; the median number of nivolumab doses received was 8.5 with [redacted] receiving >4 doses. The median duration of therapy was [redacted] months (95% CI: [redacted]). At the time of the 24 March 2016 database lock, 76.9% of patients in the UC cohort of CheckMate 032 had discontinued study treatment; the most common reason was disease progression (64.1%). Two (2.6%) patients discontinued due to study drug toxicity.

ERG comment: The outcomes for nivolumab in CheckMate 275 are generally worse than in the CheckMate 032 trial; given the low sample sizes of the studies this could be explained by sampling error. There appeared to be little change between the May and September database locks, although median OS did come down slightly. The company were asked to provide the most recent data in addition to those submitted in the CS, given that the survival data is from an analysis that is over a year old.[21] The company did not provide further data.[7] There was a statistically significant difference in OS between the PD-L1 < 1% and PD-L1 >= 1% subgroups. The company were requested to perform the

indirect treatment comparison (STC) for these subgroups, but they declined citing unavailability of PD-L1 status in the comparator trials as a reason.[7] The ERG would argue that, whilst PD-L1 status might be prognostic, it would be unlikely to affect the effectiveness of the comparator treatments given their different mode of action to nivolumab. Therefore the ERG considers PD-L1 status is unimportant for the comparator. Moreover, lack of information on other baseline characteristics did not preclude their inclusion in the prediction model for the STC (see Section 4.4.1 below) since such missing data was imputed by the company.

4.2.6 Meta-analyses of the nivolumab studies

According to the company, *'Data from CheckMate 275 and CheckMate 032 were pooled to perform the STC presented in Section B.2.9 and Appendix D'* (CS, section B.2.8, page 59).[2, 20] However, no methods or results are presented for the pooling of data.

ERG comment: The ERG asked the company to provide details of the statistical method(s) used for pooling the data from Checkmate 275 and CheckMate 032 and to explain which data were used (BIRC or investigator-assessed).[21] The ERG also asked the company to conduct pooled analyses using data from each method separately.

In the response to the clarification letter, the company did not state how the two nivolumab trials were pooled.[7] They did clarify that the BIRC method was chosen for CheckMate 275, but only the investigator-assessed results were available for CheckMate 032. They also stated the following on page 26 of the response: '*As agreed with the ERG on the preliminary teleconference to discuss the clarification questions, analyses using each method separately have not been provided.*'[7] However, no such agreement was made and the ERG continues to believe that the results of the STC using only BIRC or only investigator-led methods would provide valuable insight into the variability of those results. Given that the BIRC method was only available for CheckMate 275 this would imply a minimum of performing the STC using only the CheckMate 275 data. This additional analysis was suggested to the company during the teleconference (to which the company refer in the response to clarification) but was not performed.

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

The systematic literature review (SLR) identified no RCTs directly comparing the efficacy and safety of nivolumab in the patient population of interest versus any of the comparators relevant to this submission or placebo.

Three studies were excluded because the dose and/or treatment regimens did not correlate with current UK clinical practice.[5] The trials not considered for further assessment were Kim et al. (2016),[23] McCaffrey et al. (1997)[24] and Vaughn et al. (2002).[25]

Nine trials, including the two nivolumab trials, were considered eligible for STC.[8, 9, 13-16, 26-28] (See Table 4.12). Note that the single arm study design of the nivolumab studies prevented standard indirect comparison or mixed treatment comparisons since there was an incomplete network. To allow any comparison of nivolumab effectiveness to any eligible comparator the company performed an unanchored (no common comparator) STC. An unanchored STC relies on the major assumption that absolute outcomes can be predicted from a set of covariates; therefore it assumes that all effect modifiers and prognostic factors are accounted for.[1] In addition to the two nivolumab studies,[8, 9] a further seven studies were found to be used in the STC. The seven studies looked at paclitaxel,[15, 28] docetaxel,[16, 27] BSC,[26] and cisplatin plus gemcitabine.[13, 14]

Table 4.12: Summary of trials included in simulated treatment comparisons

Trial ID	Study design	Interventions (n patients assigned)	Treatment included in STC
Bellmunt 2009	RCT	Vinflunine + BSC (253) vs. BSC (117)	BSC
Choueiri 2012	RCT	Docetaxel + vandetanib (74) vs. Docetaxel +placebo (75)	Docetaxel
Gondo 2011	Single arm	Gemcitabine + cisplatin (33)	Gemcitabine + cisplatin
Joly 2009	Single arm	Paclitaxel (45)	Paclitaxel
Jones 2017	RCT	Pazopanib (66) vs. Paclitaxel (65)	Paclitaxel
Ozawa 2007	Single arm	gemcitabine + cisplatin (55)	Gemcitabine + cisplatin
Petrylak 2016	RCT	Docetaxel (49) vs. Docetaxel + ramucirumab (49) vs. Docetaxel + icrucumab (50)	Docetaxel
Sharma 2016	Single arm	Nivolumab (78)	Nivolumab
Sharma 2017	Single arm	Nivolumab (270)	Nivolumab

In the two trials identified for cisplatin plus gemcitabine (Gondo et al. (2011)[13] and Ozawa et al. (2007)[14]), all patients in Gondo et al. (2011)[13] had received MVAC in first-line treatment and are, according to the company, therefore not comparable to those receiving cisplatin plus gemcitabine retreatment in current UK clinical practice, as they are gemcitabine naïve.[5] The Ozawa et al. (2007)[14] trial included chemotherapy-naïve patients in addition to patients who had previously undergone first-line treatment. Although outcome data are reported separately for these two populations, patient baseline characteristic data are reported for the two populations combined. Therefore, it is not possible to determine baseline characteristics for patients who had only received first-line treatment, precluding an adjusted (STC) comparison with patients in other studies included in this analysis. Additionally, the two trials did not use the standard dosing regimen typically used for cisplatin plus gemcitabine in the UK. Furthermore, the study by Gondo et al. (2011)[13] provided no PFS data, and the study by Ozawa et al. (2007)[14] provided neither OS not PFS data. As the only identified evidence for cisplatin plus gemcitabine, these trials were taken forwards for the ITC, but the company used the comparison between nivolumab and cisplatin plus gemcitabine for the purposes of a scenario analysis only.

Only one of the studies was conducted exclusively in the UK,[15] one study included some patients from the UK (CheckMate 032: six out of 78),[7] one study was conducted in multiple countries, but it was unclear whether this included the UK[26] and the remaining six studies did not include UK patients.[8, 13, 14, 16, 27, 28].

All trials reported some inclusion criteria. All trials except Ozawa et al. (2007)[14] reported inclusion criteria relating to previous treatment. Six trials required patients to have shown evidence of recurrence or progression following first-line platinum therapy.[8, 9, 15, 26-28] One trial specified that the first-line treatment was MVAC.[13] Joly et al. (2009) did not name the type of first-line chemotherapy.[28] Ozawa et al. (2007) did not mention first-line treatment in their inclusion criteria.[14]

Although some of these studies are RCTs, the company used single arms only from each study. Therefore, all the advantages of comparability between groups in a RCT have been lost. The company

tried to use a STC to adjust for some of the differences between the included studies. As stated by the company, the network for nivolumab and its comparators is disconnected. Hence the indirect comparison was conducted using STC methodology. Ideally, for each outcome, the STC should adjust for all the effect modifiers and prognostic variables. However, this is rarely possible, as some effect modifiers and prognostic variables may not be reported by all of the trials or may not be known (for example, as yet undiscovered genetic markers). The company followed the recommendations in the NICE DSU TSD 18.[1] However, we reiterate an unanchored STC '…effectively assumes that absolute outcomes can be predicted from the covariates; that is, it assumes that all effect modifiers and prognostic factors are accounted for This assumption is very strong, and largely considered impossible to meet. Failure of this assumption leads to an unknown amount of bias in the unanchored estimate.'[1].

The nine studies included in the STC are described in Table 4.13. Details of prior chemotherapy received are reported in Table 4.14 (Patient characteristics). As can be seen in Table 4.14, patient populations in the studies differed in terms of Eastern Oncology Cooperative Group (ECOG) performance status at baseline, tumour location, presence and location of metastases, previous adjuvant treatment, prior chemotherapy treatments, prior radiotherapy, prior surgery and prior response to chemotherapy. In addition patient populations differ in BMI, ethnicity, smoking status, time since diagnosis, PDL-1 expression, haemoglobin level, platelet level, neutrophil level, CD8 count, and lactate dehydrogenase level. Baseline variables are available for some of the trials, but in many case cases no data are available.

The statistical analysis data for studies included in the STC are reported in Tables 22 and 23 of the CS (CS, Appendix D, pages 91-93).[2]

ERG comment:
There was much variability in patient populations between the included studies and so it is unlikely that they can be considered as comparable. The company did adjust for differences in performing the STC (see Section 4.4.). However, many characteristics were not reported for the comparator studies, thus leading to the likelihood of persistent imbalance in both prognostic factors and effect modifiers.[1] The majority of data for nivolumab or the eligible comparators did not come from UK patients.

Table 4.13: Single arms of studies included in the simulated treatment comparison

Trial ID	Population	Intervention	Reported outcomes specified in the decision problem	All other reported outcomes
Sharma et al. (2017)[8] CheckMate 275*	Histologically or cytologically confirmed metastatic or surgically unresectable transitional cell carcinoma of the urothelium involving the bladder, urethra, ureter, or renal pelvis, age ≥18 years, and ECOG PS of 0 or 1. Progression or recurrence after treatment either: o With at least 1 platinum-containing chemotherapy regimen for metastatic or surgically unresectable locally advanced urothelial cancer, or o Within 12 months of peri-operative (neo-adjuvant or adjuvant) treatment with platinum agent in the setting of cystectomy for localised muscle-invasive urothelial cancer.	Nivolumab 3 mg/kg Q2W	OS, PFS, ORR	BIRC-assessed PFS, OS and investigator-assessed ORR, PFS, safety, HRQoL (EORTC QLQ-C30 and EQ-5D-3L)
Sharma et al. (2016)[9] CheckMate 032*	Locally advanced or metastatic urothelial cell carcinoma, age ≥18 years, and ECOG PS of 0 or 1. Progression or recurrence either: o After at least 1 previous platinum-containing chemotherapy treatment for metastatic or locally advanced unresectable urothelial cancer, or o Recurrence within 1 year of completing previous platinum-based neoadjuvant or adjuvant treatment o After previously refusing standard treatment with chemotherapy for the treatment of metastatic (stage IV) or locally advanced disease	Nivolumab 3 mg/kg Q2W	OS, PFS, ORR	Investigator-assessed PFS, OS, DOR, safety, HRQoL (EQ-5D)

Trial ID	Population	Intervention	Reported outcomes specified in the decision problem	All other reported outcomes
Bellmunt et al. (2009)[26]	Patients with histologically confirmed locally advanced or metastatic TCC of urothelial tract, documented progression after first-line platinum-containing chemotherapy, age ≥18 years, and ECOG PS of 0 or 1.	BSC (including palliative radiotherapy, antibiotics, analgesics, corticosteroids and/or transfusions); 3-week cycle;	OS, ORR	Disease control rate, clinical benefit, QoL
Choueiri et al. (2012)[27]	Eligible patients required histologically or cytologically confirmed locally advanced or metastatic UC, progression of disease documented by the investigator after platinum-containing chemotherapy, age ≥18 years, and ECOG PS of 0 or 1.	Docetaxel (75mg/m^2 D1) + Placebo (100mg daily); 21-day cycle;	PFS, ORR	Safety and disease control rate
Jones et al. (2017)[15]	Histologically confirmed TCC of the bladder, renal pelvis, ureter or urethra which was locally advanced or metastatic; Progressive disease during or after one prior platinum-based chemotherapy regimen for advanced disease	Paclitaxel (80mg/m^2 IV over 1 hour, D1, D8, D15); 28 day course;	OS, PFS, Grade 3 and Grade overall AEs	PR, SD, QoL, toxicity
Petrylak et al. (2016)[16]	Patient had histologically or cytologically confirmed TCC of the bladder, urethra, ureter, or renal pelvis, locally advanced or metastatic and unresectable TCC of the bladder, urethra, ureter, or renal pelvis and had received treatment with a platinum-containing regimen.	Docetaxel (75 mg/m^2 IV; D1); 3-week cycle,	OS, PFS, ORR	DoR, safety, pharmacokinetics, pharmacodynamics and immunogenicity
Gondo et al. (2011)[13]	Patients with histologically confirmed advanced and metastatic UC. All patients had evidence of disease progression, relapse or no response after MVAC chemotherapy as first-line treatment.	Gemcitabine (1,000 mg/m2; D1, D8, D15); Cisplatin (35 mg/m2; D1, D2); 28 day-cycle;	OS, ORR	Toxicity
Joly et al. (2009)[28]	Patients had urothelial carcinoma of the bladder, or urothelial tract, with a progressive measurable disease after previous line of chemotherapy for	Paclitaxel (80mg/m2 IV over 1 hour, D1, D8, D15); 28 day course;	ORR	CR, PR, SD

Trial ID	Population	Intervention	Reported outcomes specified in the decision problem	All other reported outcomes
	advanced disease (neoadjuvant, adjuvant, or metastatic therapy), life expectancy ≥3 months, WHO performance status of 0-2			
Ozawa et al. (2007)[14]	Patients had histological or cytological proof of UC, at least one bi-dimensionally measurable lesion according to WHO criteria, and a WHO performance status <2	Gemcitabine (1000mg/m2 D1, D8, D15) Cisplatin (70mg/m2 D2); Every 28 days	ORR	Toxicity

Source: CS, Appendix B, Tables 16 and 17, pages 67-68 and *response to clarification letter.
AE = adverse event; BSC = best supportive care; CR = complete response; D = day; DoR = duration of response; ECOG = Eastern Cooperative Oncology Group; IV = intravenous; MVAC = methotrexate, vinblastine; adriamycin (doxorubicin) and cisplatin; NA = not applicable; NICE = The National Institute for Health and Care Excellence; ORR = objective response rate; OS = overall survival; PFS = progression-free survival; PR = partial response; PS = performance status; QoL = quality of life; SD = stable disease; TCC = transition cell carcinoma; TTR = time to response; UC = urothelial carcinoma; WHO = World Health Organization.

Table 4.14: Patients' characteristics in studies included in the simulated treatment comparison

Trial, treatment arm and population size	Age (years)	Males n (%)	ECOG status n (%)	Location of urothelial cancer n (%)	Presence and location of metastasis n (%)	Prior neo-adjuvant or adjuvant treatment	Prior type of chemotherapy n (%)	Prior radio-therapy n (%)	Prior surgery n (%)	Response to prior chemo-therapy n (%)
Sharma et al. (2017)[8] CheckMate 275*	Median 66 (38-90)	211 (78.1)	0: 145 (53.7) 1: 124 (45.9) 3: 1 (0.3)	Urinary bladder: 197 (73.0) Renal pelvis: 46 (17.0) Ureter: 19 (7.0) Urethra: 8 (3.0)	Visceral: 227 (84.1) Liver: 75 (27.8) Lymph node only: 43 (15.9)	Adjuvant: 83 (30.7) Neo-adjuvant: 60 (22.2)	Cisplatin and gemcitabine: 87 (32.2) Carboplatin and gemcitabine: 54 (20.0) MVAC: 16 (5.9) Vinflunine 20 (7.4) Paclitaxel 18 (6.7) *Therapies used in ≥5% patients in metastatic setting listed*	85 (31.5)	250 (92.6)	CR: 23 (8.6) PR: 44 (16.4) SD: 51 (19.0) PD: 88 (32.7) N/A, UtD, NR: 63 (23.3)[a] *Percentage based on prior platinum containing regiment associated with recurrence/regression (n=72)*
Sharma et al. (2016)[9] CheckMate 032*	Median 66 (31-85)	54 (69.2)	0: 42 (53.8) 1: 36 (46.2)	NR	Visceral: 61 (78.2) Liver: 20 (25.6) Lymph node only: 13 (16.7)	Adjuvant: 33 (42.3) Neo-adjuvant: 14 (17.9)	Cisplatin and gemcitabine: 23 (29.5) Carboplatin and gemcitabine: 15 (19.2) MVAC: 7 (9.0)	25 (32.1)	71 (91.0)	CR: 2 (2.8) PR: 15 (20.8) SD: (19 (26.4)

Trial, treatment arm and population size	Age (years)	Males n (%)	ECOG status n (%)	Location of urothelial cancer n (%)	Presence and location of metastasis n (%)	Prior neo-adjuvant or adjuvant treatment	Prior type of chemotherapy n (%)	Prior radio-therapy n (%)	Prior surgery n (%)	Response to prior chemo-therapy n (%)
							Carboplatin and paclitaxel: 5 (6.4) Vinflunine: 4 (5.1) *Therapies used in ≥5% patients in metastatic setting listed*			PD: 24 (33.3) N/A, UtD: 12 (16.7)[a] *Percentage based on prior platinum containing regiment associated with recurrence/regression (n=72)*
Bellmunt et al. (2009)[26] BSC n=117	65+: n=57 (48.7%)	NR	Grade 0: 45 (38.5); Grade 1: 72 (61.5); Grade 2: 0; Grade 3: 0	NR	Visceral involvement: 87 (74.4)	NR	Cisplatin and no other platinum: 85 (7.26) Carboplatin and no other platinum: 12(19.7) Other platinum combination: 9(7.7)	NR (22)	NR	NR
Choueiri et al. (2012)[27] Docetaxel and placebo n=72	≥65: n=33 (45.8%)	49 (68.1)	Grade 0: NR; Grade 1: 38 (52.8); Grade 2: NR;	NR	Visceral: 46 (63.9); Liver: 27 (37.5)	NR	Previous treatment with platinum-based chemotherapy was a requirement of the eligibility criteria.	15 (21)	Cystectomy: 36 (50)	NR

Trial, treatment arm and population size	Age (years)	Males n (%)	ECOG status n (%)	Location of urothelial cancer n (%)	Presence and location of metastasis n (%)	Prior neo-adjuvant or adjuvant treatment	Prior type of chemotherapy n (%)	Prior radio-therapy n (%)	Prior surgery n (%)	Response to prior chemo-therapy n (%)
			Grade 3: NR				Prior paclitaxel: 8 (11.1).			
Jones et al. (2017)[15] Paclitaxel n=65	Median 70 (IQR: 63-77)	49[a] (75)	Grade 0: (39); Grade 1: (52); Grade 2: (9); Grade 3: (0)	Bladder primary: NR (66)	Nodal: NR (45); Liver: NR (29) Visceral (non-lymph node): 49 (75.4)[b]	NR	Platinum based: 65 (100)	NR	NR	NR
Petrylak et al. (2016)[16] Docetaxel n=45	Median 69 (IQR: 29-84)	35 (78)	Grade 0: 17 (38); Grade 1: 26 (58); Grade 2: 1 (2.2); Grade 3: 0; Missing: 1 (2.2)	NR	Visceral: 29 (64); Liver: 12 (NR)	NR	Platinum-based therapy (cisplatin or carboplatin): 45 (100); Gemcitabine: 42 (93); Cisplatin: 31 (69); Carboplatin: 20 (44); Doxorubicin: 4 (9); Methotrexate: 4 (9); Vinblastine: 4 (9); Investigational drug: 1 (2); Paclitaxel: 4 (9); Capecitabine: 0; Fluorouracil: 1 (2); Ifosfamide: 1 (2); Mitomycin: 0; Pemetrexed: 1 (2).	5 (11)	40 (89)	44 (98)

Trial, treatment arm and population size	Age (years)	Males n (%)	ECOG status n (%)	Location of urothelial cancer n (%)	Presence and location of metastasis n (%)	Prior neo-adjuvant or adjuvant treatment	Prior type of chemotherapy n (%)	Prior radio-therapy n (%)	Prior surgery n (%)	Response to prior chemo-therapy n (%)
Gondo et al. (2011)[13] Gemcitabine and Cisplatin n=33	Median 66 (40-82)	26 (78.8)	Inclusion criteria: ECOG PS <1 n: NR	Bladder alone: 19 (57.6); Ureter: 7 (21.2); Renal pelvis: 7 (21.2).	Bone: 5 (15.2); Bone only: 1 (3) Lymph nodes only: 10 (30.3); Lymph nodes and lung: 5 (15.2); Lymph nodes and local recurrence: 4 (12.1); Lymph nodes and liver: 2 (6.1); Lymph nodes and bone: 1 (3.0); Evaluable lymph nodes: 24 (72.7) Lung only: 3 (9.1); Evaluable lung: 11 (33.3); Lung and local recurrence: 2 (6.1) Liver: 5 (15.2); Liver and peritoneum: 1 (3.0); Visceral lesions: 23; Other: 10 (30.3).	Adjuvant: 14 (42)	MVAC. Number of courses: 1: 2 (6.1); 2: 10 (30.3); 3: 10 (30.3); 4: 14 (12.1); ≥5: 7 (21.2).	NR	32 (97)	NR

Trial, treatment arm and population size	Age (years)	Males n (%)	ECOG status n (%)	Location of urothelial cancer n (%)	Presence and location of metastasis n (%)	Prior neo-adjuvant or adjuvant treatment	Prior type of chemotherapy n (%)	Prior radio-therapy n (%)	Prior surgery n (%)	Response to prior chemo-therapy n (%)
Joly et al. (2009)[28] Paclitaxel n=45	Mean 64 (47-79)	36 (80a)	NR	Bladder alone: 38 (84); Non-bladder cancer reported as other: 7 (16a)	Bone: 14 (33); Visceral: 26 (58); Nodes: 23 (55); Pulmonary: 22 (52); Liver: 16 (38); Other: 11	Adjuvant: 32 (71)	Gemcitabine and Cisplatin: 40(89) MVAC: 5(11) Paclitaxel with cisplatin: 1; Paclitaxel with cisplatin and gemcitabine: 1 first-line adjuvant: 32 (71) first-line for metastasis: 13 (29).	16 (36)	Total: 39 (87); Radical surgery: 28 (NR); Transurethral resection of the bladder: 7 (NR)	NR (62)
Ozawa et al. (2007)[14] Gemcitabine n=55	Median 71 (32-84)	44 (80)	NR	Bladder alone: 28 (50.9); Ureter: 16 (29.1); Renal pelvis: 11 (20)	Lymph nodes: 23; Lymph node and lung: 6; Lymph node and liver: 3; Lymph node and bone: 4; Lymph node, lung and liver: 1; Lymph node, lung, liver and bone: 1; Lung: 5; Lung and liver: 1; Lung and bone: 1;	NR	20/47 patients with metastatic disease received prior chemo MVAC: 14 (25a); MEC: 5 (9a); Low dose cisplatin: 1 (2a)	NR	NR	NR

Trial, treatment arm and population size	Age (years)	Males n (%)	ECOG status n (%)	Location of urothelial cancer n (%)	Presence and location of metastasis n (%)	Prior neo-adjuvant or adjuvant treatment	Prior type of chemotherapy n (%)	Prior radio-therapy n (%)	Prior surgery n (%)	Response to prior chemo-therapy n (%)
					Lung, liver and bone: 2					

Source: CS, Appendix D, Tables 20 and 21, pages 84-87 and response to clarification letter*
[a]Reviewer-calculated value, [b]Data provided by study author on request.
ECOG: Eastern Cooperative Oncology Group; IQR: interquartile range; MEC: methotrexate, epirubicin and cisplatin; MVAC: methotrexate, vinblastine, adriamycin (doxorubicin) and cisplatin; NR: not reported; PD: progressive disease.

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

4.4.1 Methodology of the simulated treatment comparison

The company used a population-adjusted method (STC) to conduct comparisons between nivolumab and eligible comparators with respect to OS, PFS and ORR outcomes.[1]

The STC was informed by individual patient data (IPD) from the two nivolumab studies[8, 9] and published data from the other seven studies of comparator treatments.[13-16, 26-28]

The methods followed the recommendations of the NICE DSU TSD 18: Methods for Population-Adjusted Indirect Comparisons in Submissions to NICE.[1]

For each outcome, the key steps of the STC approach were:

1. Use the nivolumab IPD to develop a model that predicts how patients respond to treatment based on key baseline patient characteristics.
2. For each comparator trial in the network, use the baseline characteristics from the comparator trial to predict how patients in the comparator trial might have responded to nivolumab. Compare the real data from the comparator, to the predicted data for nivolumab.
3. Use a meta-analysis to synthesise the results across all of the comparator trials.

Details of each of the steps are shown in Appendix D of the CS.[20]

For Step 1, prognostic factors and effect modifiers were identified via a targeted literature search and via discussion with clinicians at the advisory board meeting.[5] The Prediction models were estimated on the pooled CheckMate 275 and CheckMate 032 data. It was reported that stepwise model selection suggested that the best Cox Proportional hazards (PH) model for OS is based on ECOG PS, haemoglobin level, visceral metastases and liver metastases. Note that this model includes all three of the key prognostic factors identified by Bellmunt et al. (2010)[26] (ECOG PS, haemoglobin level and liver metastases). For PFS the same approach showed the best model is based on ECOG PS, age, visceral metastases and liver metastases. Stepwise model selection suggested that the best logistic regression model for objective response is based on age and visceral metastases. The basis of selection was reported to be parsimony as indicated by the Akaike information criterion (AIC). No models other than the final and presumably most parsimonious models (no more than four covariates) were presented despite the consideration of eleven possible covariates.

For Step 2, because not all of these baseline characteristics were reported for all comparator trials, for each comparator trial, any baseline characteristics that were in the final prediction models, but not reported for the comparator trial, were then predicted using the correlations between baseline characteristics in the nivolumab trials.

This method essentially adjusts the outcomes estimated from the nivolumab trials to attempt to simulate how they might be observed in each of the comparator trials. Therefore, there is one adjusted value (for nivolumab) for each outcome, e.g. ORR, for each comparator trial. This means that there can be more than one adjusted value for nivolumab per comparator. For example, as shown in Table 4.17, ORR is estimated for docetaxel from two trials, Choueiri et al. (2012)[27] and Petrylak et al. (2016).[16] Therefore, there will be two adjusted values of ORR for nivolumab to compare to these trials and to estimate the treatment effect in terms of a relative risk. For OS and PFS adjusted hazards are predicted with one for

each of a set of four-weekly time intervals. As with ORR, there are two trials for docetaxel and so this means two sets of adjusted hazards, each one of which goes into the meta-analysis model in Step 3.

For Step 3, OS and PFS were evaluated using a fractional polynomial approach, which permits the estimation of hazard ratios (HRs) that vary over time. ORR was evaluated using an evidence synthesis model for binomial outcomes.[29] For all outcomes, both fixed effect and random effects models were applied. For the survival outcomes, different types of fractional polynomial model (according to variation in two parameters that determine the shape of the survival curves) were also explored. The deviance information criterion (DIC) was used to evaluate model fit and guide the best choice of model. For the survival outcomes, clinical plausibility of the extrapolated HRs was also considered based on expert clinical feedback elicited via an advisory board and further clinician interviews.

In addition the company stated that they conducted naïve indirect comparisons alongside STCs as recommended by the DSU.[1] Although not explicitly stated, one can presume that this means that the hazards for nivolumab were not adjusted using Steps 1 and 2 above.

In order to investigate how well the STC method performed the company also compared the docetaxel versus docetaxel plus vandetanib results from Choueiri et al. (2012) [27] to the results of an STC using data from this trial.

For STCs, the NICE DSU TSD 18 recommends estimating the residual bias.[1] This is the bias due to effect modifiers or prognostic variables that are not accounted for in the prediction models because they are not available for either the nivolumab and/or the comparator studies. The NICE DSU TSD 18 emphasises that there are no standard methods for estimating the residual bias and that this is a key area for further research. The NICE DSU TSD 18 suggests two general options for evaluating residuals bias: 'in-sample' methods, which use the same data that was used to develop the prediction model, and 'out-of-sample' methods which incorporate additional data.

ERG comment: As stated above the DSU report mentions that an unanchored STC '*effectively assumes that absolute outcomes can be predicted from the covariates; that is, it assumes that all effect modifiers and prognostic factors are accounted for. This assumption is very strong, and largely considered impossible to meet. Failure of this assumption leads to an unknown amount of bias in the unanchored estimate.*'[1] The ERG believes the STC was limited by the following issues:

1. The method used for the prediction models lacked transparency; the results at each stage of the stepwise selection process were not provided. In particular, it is not clear that the most parsimonious model is the best model. It would have been useful to see an STC that was based on prediction models with more covariates including all 11 considered.
2. There was a lack of information from the comparator studies on possible effect modifiers or prognostic variables, which led to the company imputing the missing values in Step 2.
3. The company pooled the two nivolumab trials despite each one using different methods of outcome assessment, CheckMate 275 using BIRC and CheckMate 032 using investigator-assessed.
4. In an ideal scenario, the results of the STC using only BIRC or only investigator-led methods would have provided valuable insights into the variability of the results. Given that the BIRC method was only available for CheckMate 275 at a minimum it would have been useful to perform the STC using only the CheckMate 275 data. This was suggested to the company during the teleconference but was not performed.

An attempt was made by the company to validate the STC. It is the understanding of the ERG that the data from the trial by Choueiri et al. (2012)[27] was used to compare docetaxel or docetaxel plus vandetanib to nivolumab using unadjusted meta-analysis and using STC. However, this comparison is bound to produce almost identical results because in both the STC and the non-STC meta-analysis the data to inform the comparison was the same i.e. from this trial. The only difference between the STC and the direct method is that in the STC data on other trials was entered, but none of this data informs the comparison between docetaxel and docetaxel plus vandetanib. Therefore, this is essentially a spurious test of validity.

The company performed an 'in-sample' method to evaluate the residual bias. However, this method is likely to underestimate the residual bias.[1] Hence, the use of an 'out-of-sample' method is strongly recommended in NICE DSU TSD 18. This relies on the idea that, if the STC has accounted for all prognostic variables then the variance of the predictions (in this case based on the model estimated from the nivolumab data and combined with the comparator baseline characteristics) should be the same as that observed in the trial data. Unfortunately, the company concluded that the 'out-of-sample' method described in NICE DSU TSD 18 would not provide an accurate estimate of the residual bias. In the clarification letter, the company was asked to perform this analysis and in response, the company stated that in this appraisal the data was too limited to estimate the between-study variability.[7] They also argued that the fractional polynomial model constrains the between-studies variance.[29] However, they did perform the analysis and it did show much lower variance in the STC model predictions. Whether this is due to the lack of data or a limitation of the fractional polynomial model it does illustrate the point made in TSD 18 that: *'…the accuracy of the resulting estimates is entirely unknown, because there is no analysis of the potential magnitude of residual bias, and hence no idea of the degree of error in the unanchored estimates. It is, of course, most unlikely that systematic error has been eliminated.'*[1]

4.4.2 Results of the simulated treatment comparison

All studies reported data for at least one outcome. Outcome data was considered eligible for the STC analysis if a Kaplan-Meier curve was provided in addition to numerical data.

OS was reported by seven studies, including five for the four comparators with two for docetaxel.[8, 9, 13, 15, 26, 27] All of the studies except Bellmunt et al. (2009) reported a definition of OS.[26] Median survival was reported in all of the studies except Gondo et al. (2011), which reported a mean OS of 10.5 months.[13] Median OS ranged from 4.6 months in response to BSC[26] to 9.7 months in response to nivolumab.[9]

As well as in the CheckMate 032 and CheckMate 275 trials, PFS was reported by three comparators studies, for docetaxel and paclitaxel.[8, 9, 15, 27] Jones et al. (2017)[15] did not report a definition for PFS. The median PFS ranged from 1.58 months[27] in response to docetaxel and placebo to 4.1 months in response to paclitaxel.[15]

Eight studies reported ORR, including six for the four comparators.[8, 9, 13, 14, 16, 26-28] Only one study of paclitaxel by Jones et al. (2017) did not.[15] Four comparator studies did not report a definition of ORR.[8, 9, 13, 26] The ORR ranged from 0% in response to BSC[26] to 40% in response to gemcitabine and cisplatin.

The individual results of the comparator trials included in the STC are given in tables 4.15 to 4.17. The pooled results for nivolumab were not reported and were not provided in the response to the clarification letter.[2, 7]The results for the individual nivolumab trials were added to tables 4.15 to 4.17 to provide a comparison, in the absence of the pooled data.

Table 4.15: Overall survival in studies included in the simulated treatment comparison

Trial ID	Treatment arm	Population assessed (n)	Survival definition	Survival median (CI)
Sharma et al. (2017)[8] CheckMate 275	Nivolumab	265	From first dose and last known date alive or death	8.74 (95%CI 6.05 to NR)
Sharma et al. (2016)[9] CheckMate 032	Nivolumab	78	From first dose and last known date alive or death	9.7 (95% CI 7.3 to 16.2)
Bellmunt et al. (2009)[26]	BSC	117	NR	4.6 (95% CI 4.1 to 6.6)
Choueiri et al. (2012)[27]	Docetaxel and placebo	72	From date of random assignment until date of death	7.03 (95% CI 5.19 to 10.41)
Jones et al. (2017)[15]	Paclitaxel	65	From the date of randomisation	8 (80% CI 6.9 to 9.7)
Petrylak et al. (2016)[16]	Docetaxel	45	The time from random assignment to death resulting from any cause	9.2 (95% CI 5.7 to 11.7)
Gondo et al. (2011)[13]	Gemcitabine and cisplatin	33	OS was measured from the start of the gemcitabine-cisplatin regimen until the date of death or the last follow-up.	10.5 (95% CI 3 to 22.9)
Joly et al. (2009)[28]	Paclitaxel	Outcome not reported		
Ozawa et al. (2007)[14]	Gemcitabine and cisplatin	Outcome not reported		

Source: Tables 24 and 27 of CS Appendix D
BSC = best supportive care; CI = confidence interval; NR = not reported; OS = overall survival

Table 4.16: Progression-free survival in studies included in the simulated treatment comparison

Trial ID	Treatment arm	Population assessed (n)	PFS definition	PFS median (CI)
Sharma et al. (2017)[8] CheckMate 275	Nivolumab	265	Time from first dosing date to the date of the first documented tumour progression, based on BIRC assessments (per RECIST 1.1), or death due to any cause.	2.00 (95% CI 1.87 to 2.63)
Sharma et al. (2016)[9] CheckMate 032	Nivolumab	78	Time from treatment assignment to the date of the first documented tumour progression, as determined by the investigator (per RECIST 1.1), or death due to any cause.	2.78 (95% CI 1.45 to 5.85)
Bellmunt et al. (2009)[26]	BSC	Outcome not reported		
Choueiri et al. (2012)[27]	Docetaxel and placebo	72	Time between random assignment and documented progression per RECIST criteria or death.	1.58 (95% CI 1.48 to 3.09)
Jones et al. (2017)[15]	Paclitaxel	65	NR	4.1 (80% CI 3 to 5.6)
Petrylak et al. (2016)[16]	Docetaxel	45	The time from random assignment until the first radiographic documentation of objective progression defined by RECIST v1.1 or death resulting from any cause	2.8 (95% CI 1.9 to 3.6)
Gondo et al. (2011)[13]	Gemcitabine and cisplatin	Outcome not reported		
Joly et al. (2009)[28]	Paclitaxel	Outcome not reported		
Ozawa et al. (2007)[14]	Gemcitabine and cisplatin	Outcome not reported		

Source: Table 25 of CS Appendix D
BSC = best supportive care; CI = confidence interval; NR = not reported; PFS = survival

Table 4.17: Objective response rate in studies included in the simulated treatment comparison

Trial ID	Treatment arm	Population assessed (n)	OR definition	Observed cases, n (%) (CI)
Sharma et al. (2017)[8] CheckMate 275	Nivolumab	265	The best response designation, as determined by BIRC, recorded between the date of first dose and the date of objectively documented progression per RECIST v1.1 or the date of subsequent therapy.	52 (19.6) (95% CI 15.0 to 24.9)
Sharma et al. (2016)[9] CheckMate 032	Nivolumab	78	Best overall response (BOR) of complete response (CR) or partial response (PR) divided by the number of treated subjects, as determined by the investigator. Assessment of ORR in accordance with RECIST 1.1. Recorded between the date of treatment assignment and documented progression or the start date of subsequent anti-cancer therapy.	19 (24.2) (95% CI 15.3 to 35.4)
Bellmunt et al. (2009)[26]	BSC	85	NR	0 (NR)
Choueiri et al. (2012)[27]	Docetaxel and placebo	72	The percentage of participants who achieved a confirmed overall PR or CR using RECIST criteria on treatment. Patients without measurable disease only at baseline are included, based on status of non-target lesions.	8 (7.1) (NR)
Jones et al. (2017)[15]	Paclitaxel	Outcome not reported		
Petrylak et al. (2016)[16]	Docetaxel	45	Objective response: defined as the proportion of patients with a best overall response of complete or partial.	4 (8.9) (95% CI 2.5 to 21.2)
Gondo et al. (2011)[13]	Gemcitabine and cisplatin	33	NR	13 (39.4) (NR)

Trial ID	Treatment arm	Population assessed (n)	OR definition	Observed cases, n (%) (CI)
Joly et al. (2009)[28]	Paclitaxel	45	Overall ORR – not further defined	4 (9) (95% CI 2 to 21)
Ozawa et al. (2007)[14]	Gemcitabine and cisplatin	20	Objective response – not further defined	8 (40) (NR)
Source: Tables 24 and 27 of CS Appendix D BSC = best supportive care; CI = confidence interval; CR = complete response; NR = not reported; ORR = objective response rate; PR = partial response				

For each comparator trial, and each outcome, the response to nivolumab was estimated by applying the final prediction model to the baseline characteristics in the trial in order to produce adjusted values of the outcome. Tables (see Tables 4.18 and 4.19) of hazard ratios simulated as the adjusted hazard of nivolumab in each of the trials compared to the unadjusted hazard of nivolumab in the Checkmate trials were provided by the company.[20]

Table 4.18: Overall survival. Simulated hazard ratios for response to nivolumab in each of the comparator trials versus response to nivolumab in CheckMate 275 and CheckMate 032

• Trial	Mean HR[a]	Mean log HR	SD log HR
Bellmunt *et al.* (2009)	1.04	0.043	0.608
Choueiri *et al*. (2012)	0.99	-0.010	0.635
Gondo *et al.* (2011)	0.85	-0.162	0.624
Jones *et al.* (2017)	1.04	0.043	0.609
Petrylak *et al.* (2016),	0.98	-0.025	0.618
Source: Table 35 of CS Appendix D [a] Mean HR, back-transformed from the log scale. Abbreviations: HR: hazard ratio; SD: standard deviation			

Table 4.19: Progression-free survival. Simulated hazard ratios for response to nivolumab in each of the comparator trials versus response to nivolumab in CheckMate 275 and CheckMate 032

• Trial	Mean HR[a]	Mean log HR	SD log HR
Choueiri *et al.* (2012)[30]	0.96	-0.045	0.421
Jones *et al.* (2017)[31]	0.95	-0.056	0.391
Petrylak *et al.* (2016)[32]	0.88	-0.128	0.405
Source: Table 35 of CS Appendix D [a] Mean HR, back-transformed from the log scale. Abbreviations: HR: hazard ratio; SD: standard deviation			

In terms of OS, these data suggested that patients in Choueri et al. (2012) (docetaxel and placebo), Petrylak et al. (2016) (docetaxel) and Gondo et al. (2011) (Gemcitabine and cisplatin) would have had on average a better response to nivolumab than patients in the nivolumab trials.[13, 16, 27] However patients in Bellmunt et al. (2009) (BSC) and Jones et al. (2017) (paclitaxel) would have had on average a poorer response.[15, 26]

In all three studies evaluating PFS (Choueri et al. (2012), Jones et al. (2017) (paclitaxel) and Petrylak et al. (2016)) patients would have had a better response to nivolumab than patients in the nivolumab trials[15, 16, 27].

The simulation suggested that patients in each of the six comparator trials evaluating objective response would have had a better response to nivolumab than in the nivolumab trials.

For OS, the company stated that the second order (P1=0, P2=0) fixed effect model was used in the base case in the cost effectiveness model analysis because it provided the most clinically plausible extrapolations out of the three best fitting models. Therefore we present in Table 4.20 the results of this model as in the main company submission. It should be noted that HRs greater than 1 favour nivolumab.

Table 4.20: Overall survival: STC results (second order (P1=0, P2=0) fixed effect model): HRs and 95% credible intervals for each of the comparators versus nivolumab for selected time intervals

Comparison	Time Interval (weeks)	HR (95% CrI)
Paclitaxel versus nivolumab	0-4	0.13 (0.02–0.64)
	8-12	0.69 (0.36–1.26)
	20-24	1.43 (0.86–2.31)
	44-48	2.27 (1.41–3.56)
	68-72	2.63 (1.17–5.52)
	92-96	2.75 (0.82–8.52)
Docetaxel versus nivolumab	0-4	0.31 (0.09–0.84)
	8-12	1.15 (0.75–1.72)
	20-24	1.81 (1.25–2.62)
	44-48	2.11 (1.46–3.00)
	68-72	2.01 (1.14–3.37)
	92-96	1.83 (0.8–3.87)
BSC versus nivolumab	0-4	0.81 (0.33–1.79)
	8-12	2.05 (1.36–3.08)
	20-24	2.51 (1.69–3.72)
	44-48	2.27 (1.57–3.25)
	68-72	1.86 (1.17–2.85)
	92-96	1.51 (0.82–2.66)
Cisplatin plus gemcitabine versus nivolumab (scenario analysis only)	0-4	0.06 (0.00–0.70)
	8-12	0.61 (0.21–1.37)
	20-24	1.33 (0.66–2.49)
	44-48	1.75 (0.96–2.99)
	68-72	1.61 (0.68–3.31)
	92-96	1.36 (0.37–4.05)
Source: Table 18 of CS BSC = best supportive care; CrI = credible interval; HR = hazard ratio		

For PFS, the second order (P1=0, P2=0) fixed effect model was taken forward for the base case analysis in the cost effectiveness model because it had clinical plausibility and the lowest DIC. No PFS data were available for cisplatin plus gemcitabine or BSC. Therefore we present the results of this model as in the main company submission. It should be noted that HRs greater than 1 favour nivolumab (Table 4.21).

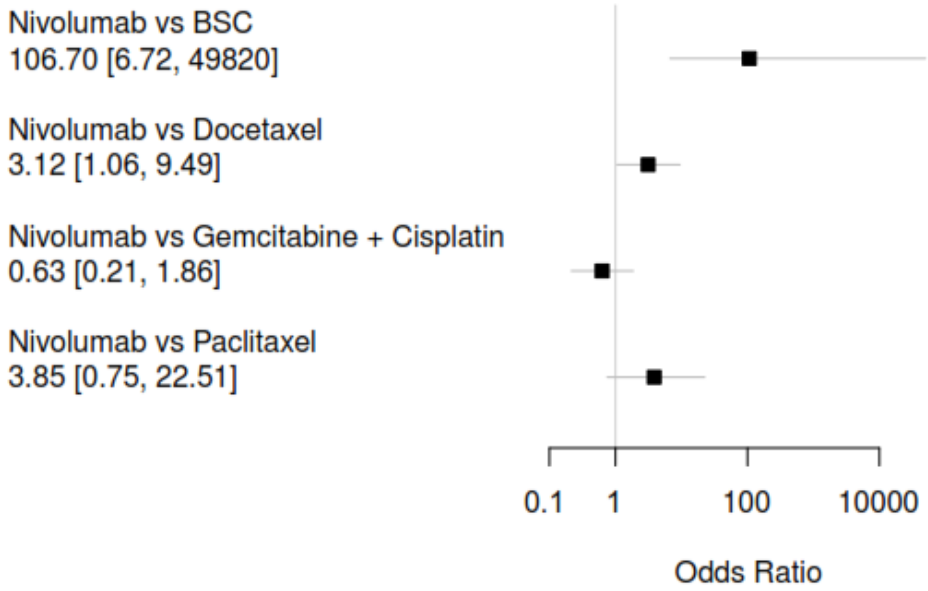
Table 4.21: Progression-free survival: STC results (fixed effect second order (P1=0, P2=0) model): HRs and 95% credible intervals for each of the comparators versus nivolumab for selected time intervals

Comparison	Time Interval (weeks)	HR (95% CrI)
Paclitaxel versus nivolumab	0-4	0.07 (0.01, 0.36)
	8-12	0.53 (0.30, 0.90)
	20-24	1.63 (1.04, 2.52)
	44-48	4.36 (1.84, 9.08)
	68-72	7.26 (1.40, 28.85)
	92-96	10.21 (0.91, 76.04)
Docetaxel versus nivolumab	0-4	1.24 (0.61, 2.42)
	8-12	1.72 (1.18, 2.49)
	20-24	1.36 (0.78, 2.20)
	44-48	0.75 (0.16, 3.19)
	68-72	0.45 (0.04, 4.82)
	92-96	0.29 (0.01, 6.93)

Source: Table 20 of the CS:
CrI = credible interval; HR = hazard ratio

For ORR the fixed effect model was used in the base case analysis so network meta-analysis results for this model are presented here (Figure 4.5, Table 4.22). However the random effects model results are also presented (Figure 4.6, Table 4.23).

Figure 4.5: Objective response rate: STC results (fixed effect model): Odds ratios for nivolumab versus each of the comparators

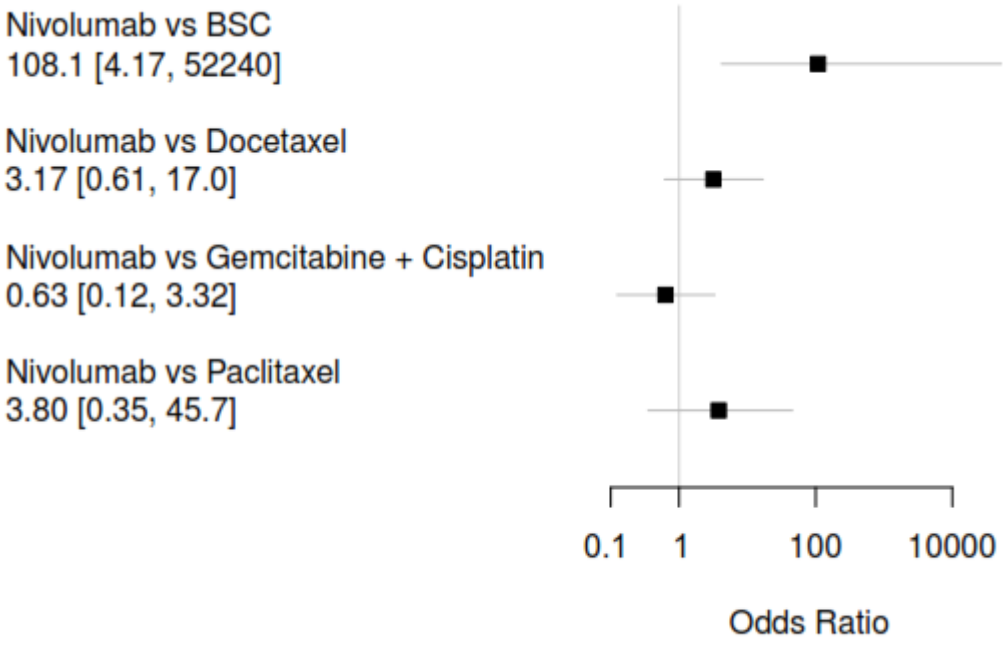


Source: Figure 30 of the CS

Table 4.22: Objective response rate: STC results (fixed effect model): Odds ratios and 95% credible intervals for each pairwise comparison

	Nivolumab	BSC	Docetaxel	Cisplatin plus gemcitabine
BSC	106.7 (6.72, 49820)			
Docetaxel	3.12 (1.06, 9.49)	0.03 (0.00, 0.59)		
Paclitaxel	3.85 (0.75, 22.5)	0.03 (0.00, 1.00)	1.23 (0.17, 9.74)	6.15 (0.87, 48.4)
Cisplatin plus gemcitabine	0.63 (0.21, 1.86)	0.01 (0.00, 0.12)	0.20 (0.04, 0.93)	
Source: Table 22 of the CS BSC = best supportive care				

Figure 4.6: Objective response rate: STC results (random effects model): Odds ratios for nivolumab versus each of the comparators



Source: Figure 19 of the CS appendices

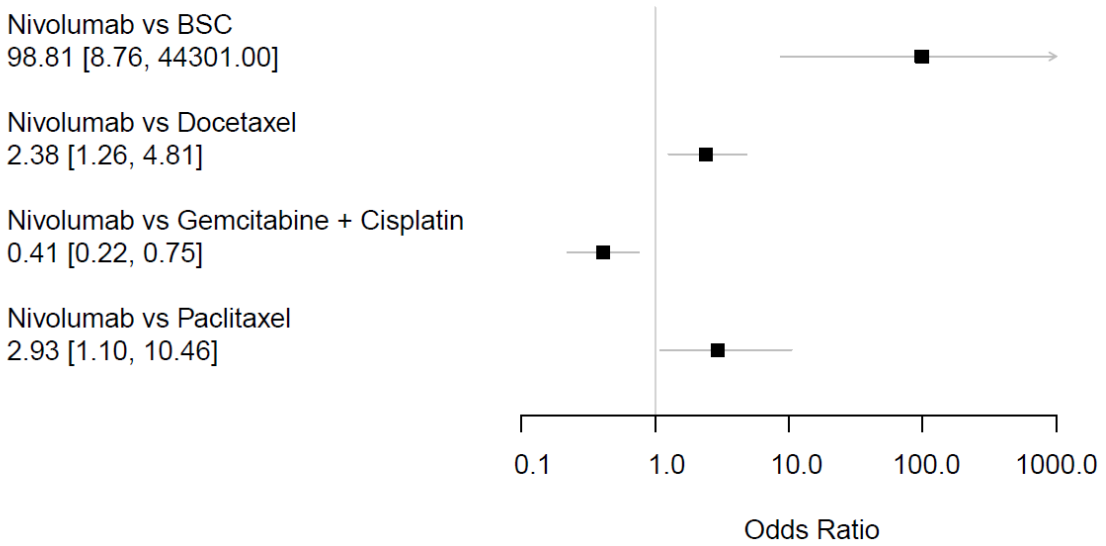
Table 4.23: Objective response rate: STC results (random effects model): Odds ratios and 95% credible intervals for each pairwise comparison

	Nivolumab	BSC	Docetaxel	Gemcitabine + cisplatin
BSC	108.1 (4.17, 52240)			
Docetaxel	3.17 (0.61, 17.0)	0.03 (0.00, 1.16)		
Gemcitabine + cisplatin	0.63 (0.12, 3.32)	0.01 (0.00, 0.23)	0.20 (0.02, 2.04)	
Paclitaxel	3.80 (0.35, 45.7)	0.03 (0.00, 2.17)	1.20 (0.07, 23.3)	6.02 (0.32, 118.1)

Source: Table 45 of the CS appendices
BSC = best supportive care

Finally, the results of a naïve indirect comparison conducted by the company in a sensitivity analysis for the outcome of objective response are presented below (both fixed effect and random effects models). Results for OS and PFS were not reported for the naïve indirect comparison: only model fit statistics were presented in the CS.[2] The results for ORR are presented in Tables 4.24 and 4.25 and Figures 4.7 and 4.8.

Figure 4.7: Naïve indirect comparison forest plot with the estimated odds ratio and its 95% credible interval, for the fixed effect model of objective response of nivolumab versus comparator treatments



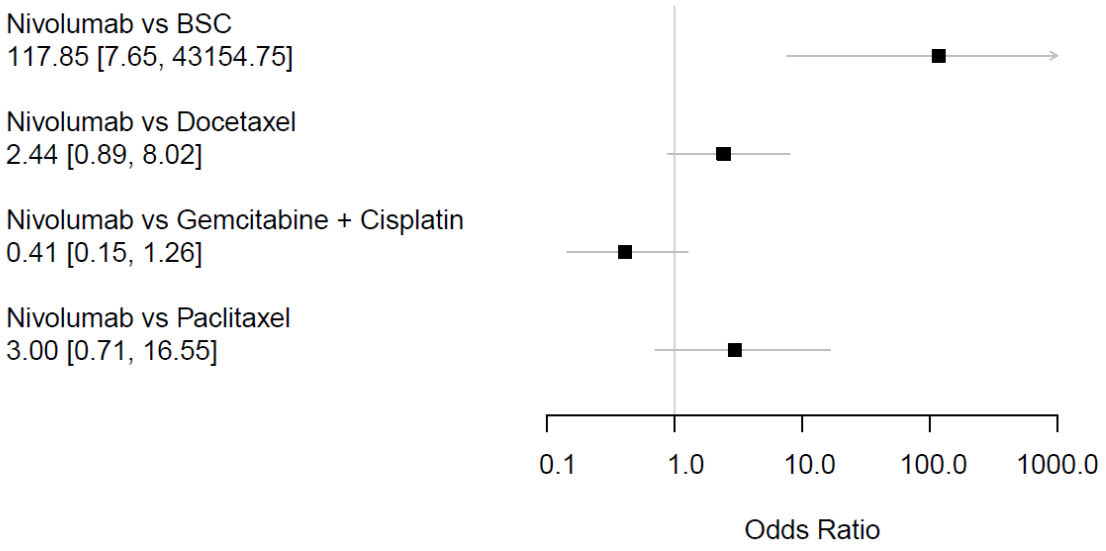
Source: Figure 20 of the CS appendices

Table 4.24: Naïve indirect comparison estimated odds ratio and 95% credible interval of the fixed effect model for the pairwise comparison of objective response between treatments

	Nivolumab	BSC	Docetaxel	Gemcitabine + Cisplatin
BSC	98.8 (8.76, 44301.00)			
Docetaxel	2.38 (1.26, 4.81)	0.02 (0.00, 0.31)		
Gemcitabine + Cisplatin	0.41 (0.22, 0.75)	0.00 (0.00, 0.05)	0.17 (0.07, 0.38)	
Paclitaxel	2.93 (1.10, 10.5)	0.03 (0.00, 0.47)	1.24 (0.39, 4.87)	7.26 (2.43, 27.8)

Source: Table 46 of the CS appendices
BSC = best supportive care

Figure 4.8: Naïve indirect comparison forest plot with the estimated odds ratio and its 95% credible interval, for the random effects model of objective response of nivolumab versus comparator treatments



Source: Figure 21 of the CS appendices

Table 4.25: Naïve indirect comparison estimated odds ratio and 95% credible interval of the random effects model for the pairwise comparison of objective response between treatments

	Nivolumab	BSC	Docetaxel	Cisplatin plus gemcitabine
BSC	117.8 (7.65, 43154.75)			
Docetaxel	2.44 (0.89, 8.02)	0.02 (0.00, 0.37)		
Paclitaxel	3.00 (0.71, 16.55)	0.03 (0.00, 0.57)	1.21 (0.26, 6.77)	7.38 (1.56, 43.0)
Cisplatin plus gemcitabine (scenario analysis only)	0.41 (0.15, 1.26)	0.00 (0.00, 0.05)	0.17 (0.05, 0.55)	
Source: Table 47 of the CS appendices BSC = best supportive care				

4.4.3 Adverse events

No formal comparison was made of AEs between the comparators. However, three studies reported overall adverse events.[8, 9, 15] Jones et al. (2017)[15] reported 27% of patients had Grade 3 or higher adverse events based on the Common Terminology Criteria for Adverse Events (CTCAE) v4.02 criteria. CheckMate 032[9] and CheckMate 275[8] both used the National Cancer Institute - Common Toxicity Criteria (NCI-CTC) v4.0 and reported the number of overall adverse events separately for Grade 3 and Grade 4. In CheckMate 032[9] and in CheckMate 275[8] the number of Grade 3 adverse event was 17 (22%) and 44 (16%) respectively. No Grade 4 adverse events were reported by CheckMate 032.[9] CheckMate 275 reported 4 (1%) Grade 4 adverse events.[8] In response to the request for clarification the company provided some more details of AEs in the comparator trials, as shown in Table 4.26.[7] A comparison can be made between these results and those reported for the CheckMate 032 and CheckMate 275 trials shown in Table 4.9. However, the AEs incorporated in the CEA and thus probably of most importance were summarised in the CS in the cost effectiveness section and reproduced in Table 5.7 below.[2] This shows that the rate of neutropaenia was generally lower than for most comparators, the exception being BSC, and much lower than for cisplatin and gemcitabine. The rate for anaemia was a little lower except for being much lower than BSC and even lower again in comparison to cisplatin and gemcitabine. For leaukopaenia the rate was comparable i.e. 0% between all comparators where it was reported except again cisplatin plus gemcitabine. The rate of asthaenia was also lower than all comparators except cisplatin plus gemcitabine.

Table 4.26: Comparator adverse events

Study	Treatment	Safety population n	Neutropenia n (%)	Febrile Neutropenia n	Anaemia n (%)	Thrombocytopenia n (%)	Asthenia n (%)	Nausea n (%)	Vomiting n (%)	Diarrhoea n (%)	Pruritus n (%)	Pneumonia n (%)	Lung infiltration n (%)	ALT increase n (%)	Hepatitis n (%)	Abdominal pain with	Fever n (%)	Leukopenia n (%)	Constipation n (%)
Bellmunt et al. (2009)[26]	Vinflunine and BSC	248 at base line	123 (50)	15 (6)	47 (19.1)	14 (5.7)	48 (19.3)	6 (2.4)	7 (2.8)	NR	NR	NR	NR	NR	NR	NR	NR	NR	40 (16.1)
	BSC	117 at base line	1 (0.9)	0 (0)	9 (8.1)	1 (0.9)	21 (17.9)	1 (0.9)	0 (0)	NR	NR	NR	NR	NR	NR	NR	NR	NR	1 (0.9)
Choueiri *et al*. (2012)[27]	Docetaxel and Vandetanib	142	10 (14)	NR	1 (1)	NR	4 (6)	NR	NR	0 (0)	NR	NR	NR	NR	NR	NR	NR	NR	NR
Jones *et al.* (2017)[15]	Paclitaxel	129	Grade 3>: (6)	NR	NR	Grade 3≥ 0 (0)	Grade 3≥: NR (5)	Grade 3≥: 0 (0)	NR	Grade 3≥: NR (2)	NR	NR	NR	Grade 3>: NR (2)	NR	NR	NR	NR	NR
Petrylak *et al*. (2016)[16]	Docetaxel	140	Grade 3>: 16 (36)	Grade 3≥: 6 (13)	Grade 3≥: 3 (6.7)	Grade 3≥: 0	Grade 3≥: 6 (13)	Grade 3≥: 0 (0)	Grade 3≥: 0 (0)	Grade 3>: 1 (2.2)	NR	Grade 3>: 4 (8.9)	NR	NR	NR	NR	NR	Grade 3>: 6 (13)	NR
Gondo *et al*. (2011)[13]	Gemcitabine and cisplatin	33	Grade 3:	NR	Grade 3:	Grade 3: 5	Grade 3: 0	Grade 3: 0	Grade 3: 0	NR	NR	NR	NR	NR	NR	NR	Grade 3: 0	Grade 3:	NR

Study	Treatment	Safety population n	Neutropenia n (%)	Febrile Neutropenia n	Anaemia n (%)	Thrombocytopenia n (%)	Asthenia n (%)	Nausea n (%)	Vomiting n (%)	Diarrhoea n (%)	Pruritus n (%)	Pneumonia n (%)	Lung infiltration n (%)	ALT increase n (%)	Hepatitis n (%)	Abdominal pain with	Fever n (%)	Leukopenia n (%)	Constipation n (%)
			19 (57.6); Grade 4: 3 (9.1)		12 (36.4); Grade 4: 2 (6.1)	(15.2); Grade 4: 6 (18.2)	(0); Grade 4: 0 (0)	(0); Grade 4: 0 (0)	(0); Grade 4: 0 (0)								(0); Grade 4 0 (0)	14 (42.4); Grade 4: 1 (3).	
Joly *et al*. (2009)[28]	Paclitaxel	44	Grade 3: 1 (2); Grade 4: 2 (4)	NR	Grade 3: 3 (7); Grade 4: 2 (4).	NR	Grade 3: 6 (14); Grade 4: 0 (0)	Grade 3: 1; Grade 4: 0.	Grade 3: 1 (2); Grade 4: 0 (0).	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
Ozawa *et al*. (2007)[14]	Gemcitabine and cisplatin	55	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR

Source: response to clarification

BSC: Best supportive care; NR: not reported

ERG comment:
In terms of ORR the main analysis using the fixed effect model presented finds that nivolumab is significantly better than BSC and docetaxel. No significant differences were found for nivolumab paclitaxel and gemcitabine. In the random effects model nivolumab is only statistically significantly superior to BSC. In the naïve indirect comparison nivolumab is superior to all three comparators in the fixed effect model but only to BSC in the random effects model. The results of the STC show that for OS and PFS nivolumab is superior to all comparators at most time points. However, the credible intervals for the HRs are quite wide, crossing 1 in many cases. The results of the naïve indirect comparison i.e. with the fractional polynomial model, but without the STC, were not reported. Results for other functional forms of the fractional polynomial model were presented in Appendix D, but of many functional forms, the results of only two more were presented.[20] The company was also asked to provide the results assuming proportional hazards i.e. one HR (fixed with respect to time) per comparator.[21] In response, the company provide the results of both random and fixed effects models. The method described appeared to be ad hoc. They first estimated so-called 'naïve' HRs using a proportional hazards model, but not using adjusted data i.e. apparently using the CheckMate trial data. They then adjusted these HRs to produce those intended to be as a result of the STC by the following method:

1) For each patient they calculated an adjusted HR by multiplying this 'naïve' HR by a factor calculated as the ratio of the hazard predicted by the prediction model (given the patient's characteristics) and the hazard of a patient with characteristics at the average CheckMate values

2) They then took the average of the log of this adjusted HR to get the mean adjusted log HR for each trial i.e. five values, which was then entered in the meta-analysis model.

No formal comparison was made of AEs and perhaps the most important AE data was reported in the cost effectiveness section of the CS.[2] However, it appears that the rates for nivolumab were either lower or comparable to those for the comparators.

4.5 Additional work on clinical effectiveness undertaken by the ERG

The company did not show the unadjusted hazards (estimated directly from the CheckMate 032 and CheckMate 275 trials), but they did state that they used a proportional hazard model, which suggests that hazards at all time points would be increased by the same amount, as indicated by the HRs in Tables 4.18 and 4.19. In order to check the reproducibility of the STC the data and code for running the models was requested by the ERG.[21] In response the company supplied this as an R script, with a Winbugs script embedded. However, the ERG could not run this without it generating errors and so requested it purely as a Winbugs script i.e. with the data incorporated in Winbugs format. The ERG has been able to run the meta-analyses and reproduced results only different by an amount that could be attributed to random error. The ERG can also verify that the data for OS and PFS includes the adjusted log hazards for nivolumab i.e. as a result of the STC. Because the company failed to show the unadjusted values i.e. those estimated directly from the CheckMate 032 and CheckMate 275 trials the ERG sought a method of estimating these. For OS, it was found that there were 110 values of the log hazard in five sets of 22 (corresponding to 22 four–weekly time intervals), one set for each of the five comparator trials shown in Table 4.18 (Table 35 in the CS). It was shown that by re-adjusting each of the five sets of the log hazards by the mean log HRs in Table 4.18, a single set of 22 hazards could be obtained. This verified the proportional hazard assumption since only one log HR per set was required to obtain the same original set of hazards. This single set, by definition, must be those without adjustment by the STC and which can thus be considered as having been estimated directly from the CheckMate 032 and CheckMate 275 trials.

The ERG was also able to perform the same analysis for PFS as for OS described above. In this case there were 36 nivolumab log hazards in three sets corresponding to the three PFS studies, as shown in Table 4.19 and used for only two comparators, paclitaxel and docetaxel.

The ERG was also able to check the last stage i.e. the evidence synthesis by which the fixed HRs were estimated, which revealed that this was essentially pointless in that the HRs that acted as inputs ended up being identical to the outputs, except for that versus docetaxel. This is because there was only one input per comparator, except for docetaxel for which there were two i.e. from two trials, Choueiri et al. (2012)[27] and Petrylak et al. (2016).[16] The ERG was also not convinced that the method prior to this final stage i.e. adjusting the naïve HRs was valid. Instead, for OS, the ERG performed the method advocated by Jansen, which sets the time dependent parameters in the fractional polynomial model to zero, thus allowing only a difference in the time-independent hazard.[29] This should then allow the estimation of fixed HRs. Following this method produced HRs that were quite dissimilar to those reported in the response to clarification.

4.6 Conclusions of the clinical effectiveness section

Ideally, in order to determine the relative benefits of nivolumab and its comparators there would be a series of randomised controlled trials comparing nivolumab and its comparators. Failing this, a network meta-analysis of RCTs using a set of common comparators would be the preferred approach. This would be the clearest way of determining if there was a gain in PFS or OS. However the submission relies on two single arm studies of nivolumab, one of which is small, which are then entered into a STC together with the single arms from some RCTs. Comparisons based on single arms from RCTs and studies are by their nature far less reliable than those made using the difference between arms from RCTs; in effect a comparison of observational data. The methods used by the company to conduct the STC largely follow those described in NICE DSU TSD 18, but, as stated in the same TSD, given no comparative data (unanchored analysis) the results obtained should be treated with caution.[1] As TSD 18 makes clear, unless all baseline characteristics that might be prognostic variables and effect modifiers are incorporated in any model to adjust for bias, it is unclear what the size of any bias might be. The ERG found the following limitations in the STC analysis:

1. Although the company stated that they had tested the fit of prediction models with various sets of baseline characteristics, it is not entirely clear how this was done: the final model had far fewer covariates than originally considered and no models with more covariates were presented or incorporated in the STC as part of a sensitivity analysis.
2. Many baseline characteristics were not available across all comparator trials and had to be imputed
3. The only external test of validity of the STC i.e. the ‘out-of-sample’ method seemed to either show insufficient reduction in bias or be inapplicable given the use of the fractional polynomial model that was used for survival analysis.
4. To compound the uncertainty, the numbers of actual patients are small for all comparisons and not all studies provided data for all outcomes.
5. The survival data are not fully mature in the nivolumab trials. The latest database lock provided updated OS data with a median follow-up time of 11.5 months, and at this point, only 57% of patients had died. The ERG did ask for an analysis based on more recent data, but none was provided.[7]
6. Not all study outcomes are based on independent review. An analysis based only on BIRC derived data from the nivolumab trials was also requested.[21] However, in the response to the clarification letter, the company declined to do this.[7] They also stated the following on page 26 of the response: ‘As agreed with the ERG on the preliminary teleconference to discuss the

clarification questions, analyses using each method separately have not been provided.' However, no such agreement was made. Given that the BIRC method was only available for CheckMate 275 the best analysis would use only the CheckMate 275 data. This was suggested to the company during the teleconference to which the company refer in the response to clarification.

7. The company also stated that a naïve indirect comparison was performed, which the ERG understands to be without the STC, but still using the fractional polynomial model for OS and PFS. Given the ERG's opinion that the fractional polynomial model was probably appropriate, and there is doubt as to the validity of the STC, the ERG considers that the results of this naïve indirect comparison should be presented. The ERG did attempt this, but only by back-calculation and with no estimate of uncertainty.
8. The ERG would accept that the polynomial fraction model appears to be a valid and highly flexible approach to estimating HRs. However, the results of very few functional forms were presented, leaving some doubt as to the most appropriate. Also, one legitimate form is to assume proportional hazards i.e. a fixed HR with respect to time. The company did attempt this, but the methods are questionable and the method, which uses the same model as that with time-dependent HRs was not employed. Its employment by the ERG, at least for OS, seemed to produce quite different results.

Although the pooled nivolumab trial data that was used for the STC was not presented in the CS, one can compare at least crudely (without any adjustment for baseline characteristics) the outcomes of the nivolumab trials (in Tables 4.15 to 4.17) with those of the comparator trials. In particular, OS and PFS do appear to be superior for nivolumab than for BSC. However, there appears to be almost complete overlap in the 95% CIs for PFS and OS between CheckMate 275 and the docetaxel trial.[10, 16] Of course, this is without any adjustment, but even the STC, which includes the CheckMate 032 trial, which is more favourable to nivolumab, shows considerable uncertainty.[11] It is also the belief of the ERG that the comparison with gemcitabine plus cisplatin is legitimate despite the differences to the scope identified by the company in the treatment history of the patients in these trials.[13, 14] The main reason for this is that it appears to the ERG that these differences affect comparability in the same way as in all of the other comparator trials and which the company has attempted to adjust for using the STC.

It should also be highlighted that no evidence synthesis of AEs or HRQoL was performed, although the rates for nivolumab did appear to be similar or lower than for the comparators.

In conclusion, it is difficult to be sure what the effectiveness of nivolumab is in comparison to the comparators in the scope. There is evidence from directly examining the single arms of the trial data that there is little difference between the outcomes measured from the nivolumab and comparator studies. Of course, naïve comparison of single arms clearly carries a high risk of bias. However, there is also no clear evidence that this risk of bias would be reduced by the STC analysis. Multiple limitations in the STC were identified and a judgment of the influence of the adjustment due to the STC cannot be evaluated because the company did not present an unadjusted (naïve) analysis. The ERG was able to estimate the unadjusted hazards, but not with estimates of uncertainty. The effect of an analysis based on a different prediction model remains unknown. As stated on page 56 of TSD 18, and used by the company for the basis of the STC: '*The size of this systematic error can certainly be reduced, and probably substantially, by appropriate use of MAIC or STC. Much of the literature on unanchored MAIC and STC acknowledges the possibility of residual bias due to unobserved prognostic variables and effect modifiers; however, it is not made clear that the accuracy of the resulting estimates is entirely unknown, because there is no analysis of the potential magnitude of residual bias, and hence no idea of the degree of error in the unanchored estimates. It is, of course, most unlikely that systematic error*

has been eliminated. Hoaglin,[72, 73] in a series of letters critiquing an unanchored comparison by Di Lorenzo et al.[78] based upon a matching approach similar to MAIC, remarked that, without providing evidence that the adjustment compensates for the missing common comparator arms and the resulting systematic error, the ensuing results 'are not worthy of consideration'.'[1]

5. COST EFFECTIVENESS

5.1 ERG comment on company's review of cost effectiveness evidence

This section pertains mainly to the review of cost effectiveness analysis studies. However, the search section (5.1.1) also contains summaries and critiques of other searches related to cost effectiveness presented in the CS. Therefore, the following section includes searches for identifying economic evaluations; studies reporting utility values and; studies reporting cost/resource use data.

5.1.1 Objective and searches performed for cost effectiveness section

The following paragraphs contain summaries and critiques of all searches related to cost effectiveness evidence presented in the CS.

Objective of cost effectiveness analysis search and review

The company performed an SLR with the objective to identify evidence to support the development of a cost effectiveness model for nivolumab as a treatment for locally unresectable or metastatic UC. With a single review, the company aimed to identify relevant UC studies in terms of published:

1. economic evaluations;
2. studies reporting utility values and;
3. studies reporting cost/resource use data.

The CS reported that searches were carried out in December 2016. Searches were not limited by date or by language. A single review was performed to identify relevant studies in UC that included published economic evaluations, studies reporting cost/resource use data, and studies reporting utility values

Searches were carried out on the following databases: Embase, MEDLINE, MEDLINE In-Process, MEDLINE Epub Ahead of Print, HTA and NHS EED via the Cochrane Library and EconLit. Searches were carried out in line with the NICE 2013 guide to the methods of technology appraisal sections 5.2.2 and 5.2.4.[33]

Supplementary searches of the following conference proceedings for 2014-2016 were reported: American Society of Clinical Oncology (ASCO), European Association of Urology (EAU), European Multidisciplinary Meeting on Urological Cancers (EMUC), European Society for Medical Oncology (ESMO), International Society for Pharmacoeconomics and Outcomes Research (ISPOR - Europe and International). The CS also reported searches of the following resources: NICE, SMC and NCPE websites, Cost-Effectiveness Analysis (CEA) Registry, University of Sheffield Health Utilities Database (ScHARRHUD) and EQ-5D Publications Database.

Bibliographies of identified systematic reviews, meta-analyses and HTA submissions were searched for relevant articles.

ERG comment:

- The searches in Appendix G were clearly structured, documented and reproducible, using a wide range of resources to identify published and unpublished literature. Database hosts and dates of searches were all reported. Most database searches used combinations of indexing terms appropriate to the resource searched, free text and a number of synonyms for the condition. Language limits were not applied.
- The EconLit strategy was limited, however due to the database content this is unlikely to have resulted in missed relevant studies.

- Study design filters were applied to the Embase and MEDLINE searches, and although these do not appear to be published validated filters, they contain a wide range of search terms and are therefore unlikely to have missed any relevant studies.
- Search strategies were missing from the CS for NHS EED, the HTA database and EconLit, and for the conference and website searches, however these were supplied in full by the company following a request for clarification.

5.1.2 Inclusion/exclusion criteria used in the study selection

Full details regarding the inclusion/exclusion criteria are provided in Appendix G of the CS (Table 60). In summary the following criteria were used:

- **Patient:** Patients with advanced, metastatic or unresectable UC (mixed populations were excluded unless results were presented separately for those with advanced, metastatic or unresectable)
- **Intervention and comparator:** any intervention or comparator except non-pharmacological interventions, which were excluded
- **Outcomes:** 1) LYs, quality adjusted life years (QALYs) or costs (UK perspective); 2) original health state utility data or; 3) original costs or resource use data relevant to the UK NHS or social work in Scotland or the Health Service Executive in Ireland
- **Study design:** original research or SLR
- **Other:** English language only

5.1.3 Included/excluded studies in the cost effectiveness review

In total 676 references were identified in the SLR. Duplicates (n=100) were excluded, resulting in 576 references for the title and abstract screening. During this process 539 references were excluded (22 due to reference not being in English/not in human participants). After full-text screening of the remaining 37 references, another 31 references were excluded (see Appendix G of the CS (Table 61) for the reason for exclusion per study). After including three references identified by hand search, nine references (seven unique studies) were included, including three economic evaluations.[34-36] See Appendix G of the CS (Figure 29) for the PRISMA diagram. The included studies are summarised in Appendices G.2.1, G.2.2, H and I of the CS.

5.1.4 Conclusions of the cost effectiveness review

Although economic evaluations were identified with populations that matched the population described in the final scope of this appraisal, these did not consider the cost effectiveness of nivolumab and therefore a de novo health economic analysis was conducted for the purposes of this appraisal.

In the vast majority of the studies that report original health-state utility data, no EQ-5D health state descriptions were used, and the studies did not report full details of the elicitation and valuation methods. Therefore, none of the included utility studies were deemed consistent with the NICE reference case for use in the health economic model. To inform the utility values for the economic model, the company used EQ-5D-3L data collected from the CheckMate 275 trial. Additionally, the disutilities for Grade 3 and 4 AEs were derived from the literature (CS Table 35). However, it was unclear how these studies were identified (as these studies were not retrieved from the SLR).

One of the identified resource use and cost studies was used to retrieve the AE costs for leukopoenia (CS Table 41). Although other literature sources were used (e.g. for terminal care costs and costs for other AEs), it was unclear how these studies were identified (as these studies were not retrieved from the SLR).

ERG comment: Since the identified cost effectiveness studies were not performed using the intervention of interest, the ERG agrees that conducting a de novo health economic analysis was necessary. Relevant health-state utility, as well as resource use and cost studies were identified by the company. It was however unclear why the company used literature sources not identified in the SLR to inform the model and not for instance TA272 (the only other NICE submission in this indication), which was identified in the SLR. Additionally, it was unclear how these alternative literature sources were identified.

5.2 Summary and critique of company's submitted economic evaluation by the ERG

Table 5.1: Summary of the company's economic evaluation (with signposts to CS)

	Approach	Source / Justification	Signpost (location in CS)
Model	A cohort-based partitioned survival model was implemented in Excel	To capture the progressive nature of UC disease and to provide consistency with previous NICE submissions relating to metastatic cancers.	Section B.3.2.2
States and events	Health states: - Progression-free state - Progressed disease state - Death	To be in line with previous NICE submissions relating to metastatic cancers, including the only previous submission in this indication (TA272, 2013)[34]	Section B.3.2.2
Comparators	- Paclitaxel - Docetaxel - Best supportive care - Cisplatin + gemcitabine (only in scenario analysis)	Paclitaxel, docetaxel and BSC were included to be consistent with the scope. The scope also specified cisplatin + gemcitabine as a comparator but this was only included in scenario analysis because of limited evidence on cisplatin + gemcitabine for retreatment with first-line platinum-based chemotherapy.	Section B.3.2.3
Population	Patients with metastatic or unresectable UC who have progressed following first-line platinum-based chemotherapy.	This is consistent with the population of the CheckMate 275 and 032 trials, as well as the final scope issued by NICE.	Section B.3.2.1
Treatment effectiveness	Treatment effectiveness was estimated in terms of gains in OS and PFS that nivolumab could provide over the comparators. Estimates were informed by the CheckMate 275 and 032 single-arm studies, using response-based survival analysis	A response-based modelling approach to estimate OS and PFS was adopted in order to reflect the mechanism of action of nivolumab and that the nivolumab survival curve changes over time as the hazard changes. According	Sections B.3, B.3.3.1 and B.3.3.2

	Approach	Source / Justification	Signpost (location in CS)
	implemented using landmark analysis, where responders and non-responders were modelled separately from the chosen 8-week landmark. A simulated treatment comparison informed time-varying hazard ratios for nivolumab versus each comparator.	to the company, standard parametric models were deemed unlikely to be flexible enough to characterise this change in the hazard. To overcome immortal time bias, landmark analysis was used. It was necessary to generate time-varying hazard ratios as the proportional hazard assumption did not hold for these comparators given the unique mechanism of action of nivolumab.	
Adverse events	Resource use, costs and utility decrements were considered for Grade 3 and 4 AEs.	To represent those AEs that are more likely to have an effect on quality of life.	Sections B.3.4.4, B.3.4.5 and B.3.5.1
Health related QoL	The HRQL data used in the cost effectiveness analysis for the progression-free and the progressed disease state were derived from EQ-5D-3L data collected in CheckMate 275 and analyses using a mixed model. Disutilities for AEs were also included; these were derived from the literature.	None of the studies identified through the SLR were deemed to be consistent with the NICE reference case.	Section B.3.4
Resource utilisation and costs	Resource use and costs in the model consisted of drug acquisition costs and drug dosing, drug administration and monitoring, costs associated with best supportive care, treatment discontinuation, terminal care and AEs. These were based on information from CheckMate 275, the BNF, EMIT, published sources identified in the SLR and expert clinician feedback.	CheckMate and published sources were used when they provided estimates of resource use and costs. In the absence of such estimates, assumptions were made and validated through discussions with clinicians.	Section B.3.5
Discount rates	Discount rate of 3.5% for utilities and costs	As per NICE reference case	Table 42
Sub groups	None	As per NICE scope	Section B.3.9

	Approach	Source / Justification	Signpost (location in CS)
Sensitivity analysis	Both DSA and PSA were performed as well as scenario analyses	The PSA excluded key parameters.	Sections B.3.8

Source: CS
Abbreviations: AE, adverse events; BNF, British National Formulary; BSC, best supportive care; CS, company submission; DSA, deterministic sensitivity analysis; EMIT, electronic market information tool; HRQL, health-related quality of life; OS, overall survival; PFS, progression-free survival; PSA, probabilistic sensitivity analysis; SLR, systematic literature review; UC, urothelial cancer

5.2.1 NICE reference case checklist

Table 5.2: NICE reference case checklist

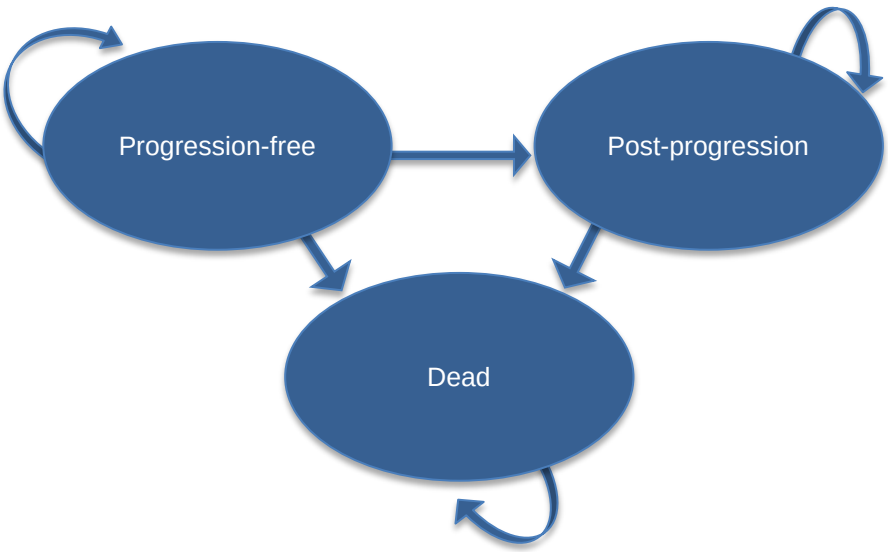
Elements of the economic evaluation	Reference Case	Included in submission	Comment on whether *de novo* evaluation meets requirements of NICE reference case
Population	As per NICE scope	Yes	
Comparator(s)	Therapies routinely used in the National Health Service (NHS), including technologies regarded as current best practice	Partly	Comparator cisplatin + gemcitabine was identified in NICE scope but only included in scenario analysis.
Type of economic evaluation	Cost effectiveness analysis	Yes	
Perspective on costs	NHS and Personal Social Services (PSS)	Yes	
Perspective on outcomes	All health effects on individuals	Yes	
Time horizon	Sufficient to capture differences in costs and outcomes	Yes	
Synthesis of evidence in outcomes	Systematic review	Yes	
Measure of health effects	Quality adjusted life years (QALYs)	Yes	
Source of data for measurement HRQoL	Described using a standardised and validated instrument	Yes	
Source of preference data for valuation of changes in HRQoL	Time-trade off or standard gamble	Yes	
Discount rate	An annual rate of 3.5% on both costs and health effects	Yes	

Elements of the economic evaluation	Reference Case	Included in submission	Comment on whether *de novo* evaluation meets requirements of NICE reference case
Equity weighting	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Yes	
Sensitivity analysis	Probabilistic modelling	No	The PSA does not incorporate all relevant parameters (the HRs, a key parameter in the model, are not reflected in the PSA).
Source: CS Abbreviations: HR, hazard ratio; NHS, National Health Service; PSA, probabilistic sensitivity analysis; PSS, Personal Social Services; QALY, quality-adjusted life year			

5.2.2 Model structure

The company developed a de novo economic model using a cohort-based partitioned survival model. The model consists of three mutually exclusive health states: progression-free (PF) and post-progression (PP) disease states and death. Patients enter the model in the PF state and are treated with nivolumab or one of its comparators. Patients remain in the PF state until disease progression or death. The proportion of patients in each health state changes over time and is determined by the OS and PFS curves, which are treatment dependent. Patients cannot move from the PP state back to the PF state. This model structure was chosen to capture the progressive nature of UC disease and to be consistent with previous submissions to NICE relating to metastatic cancers, including the previous submission in this indication (TA272, 2013)[34]. The model structure is depicted in Figure 5.1.

Figure 5.1: Partitioned survival model structure



ERG comment: The ERG's comments include (1) a critique of the choice of partitioned survival analysis for this decision problem and (2) the use of response-based analysis without reflecting responder and non-responder states in the model structure.

(1) The recent TSD 19 critiques partitioned survival analysis modelling in cancer appraisals.[37] It is stated that it is the most commonly used decision modelling approach in advanced or metastatic cancer. Limitations of the method include that (1) survival functions are modelled independently even though there are dependencies such as that progression is a prognostic factor for mortality, (2) transition probabilities are not estimated for each possible transition between health states. These limitations are especially evident in the extrapolation beyond trial data (before that, dependencies are reflected in the data) and can lead to inappropriate extrapolation [37]. This can, for example, be caused by mortality hazards being extrapolated independently of progression, whilst the mix of progressed and non-progressed patients changes over time (at a certain time all patients will have progressed), or by inappropriate reflection of the treatment effect mechanism in the estimated long-term hazards. Alternatives include other types of transition models, as well as a hybrid modelling approach, by which patients were first allocated to a treatment response category using a decision tree, and second a partitioned survival analysis approach was used. The company, in response to clarification questions, stated that other model structures were not explored.[7] Based on TSD 19, the ERG considers that alternative model structures should and will be considered more frequently in the future, but the company's approach is consistent with past technology appraisals.

(2) The company used a response-based approach to modelling overall survival (OS) and progression-free survival (PFS), but does not reflect the resulting responder and non-responder groups in their model structure. The combination of these groups introduces a superfluous assumption, which is that the proportions of responders and non-responders remain the same throughout the model time horizon. This assumption is unrealistic given that responders are likely to survive longer compared to non-responders, resulting in an increase in the proportion of responders over time. Had the company kept these two groups separate by allowing for differential responder and non-responder health states, the change in responder and non-responder proportions over time would have been reflected automatically. The company argued in their response to clarification questions that it was not possible to keep these two groups separate because the STC required a larger sample size to estimate HRs for responders and non-responders separately.[7] The ERG wishes to highlight that it is not necessary to estimate separate HRs for the two groups and that this was explained in detail at the preliminary teleconference to discuss the clarification questions. The same HR could have been applied to both groups, as is done in the model currently.

5.2.3 Population

The model includes patients with metastatic or unresectable UC who have progressed following first-line platinum-based chemotherapy. Patient characteristics included in the model were age, gender, weight and body surface area (BSA). These were based on the CheckMate 275 study[10].

ERG comment: This patient group is consistent with the population of the CheckMate 275 and CheckMate 032 trials, as well as the final scope issued by NICE for this appraisal. Age and gender estimates are relevant for the calculation of background mortality and are further discussed in Section 5.2.6. Weight and BSA influence the calculation of dose and there is a discussion about this in Section 5.2.9.

5.2.4 Interventions and comparators

Nivolumab is implemented as per its EMA Summary of Product Characteristics (SmPC) posology and method of administration for second-line UC (i.e. 3mg/kg Q2W).

The company considered the following comparators in their base-case:

- Paclitaxel: 80mg/m^2 Q3W of a four-week cycle
- Docetaxel: 75mg/m2 Q3W
- Best supportive care (BSC)

The company also presented a scenario analysis, in which cisplatin + gemcitabine was added as a comparator. The company justified this deviation from the scope (i.e. not including cisplatin + gemcitabine in its base-case) by stating that there was limited evidence for retreatment with first-line platinum-based chemotherapy regimens for patients with locally advanced unresectable or metastatic UC. The SLR had not identified any relevant trials for this comparator. The only available data stemmed from a trial in which cisplatin + gemcitabine was used in re-challenge[13], assuming a gemcitabine-naïve patient population. The company argued that this study was non-generalisable to the UK, where it is standard clinical practice that patients would receive cisplatin plus gemcitabine as first-line treatment, and where different dosing schedules from the ones in the study are used.

ERG comment: The ERG requested that the company provide the comparison of nivolumab with cisplatin plus gemcitabine in the base-case, but the company did not provide this analysis within the base-case analysis. The company justified this in their response to clarification question A15[7] citing expert opinion stating that the population in the Gondo (2011) study[13] differed from the UK population in that the study population received MVAC in first line instead of cisplatin plus gemcitabine. The ERG challenges the position of the company in that patients in the Gondo (2011) study[13] would have had exposure to platinum-based therapy (part of MVAC is cisplatin) and that the precise combination of first-line treatment or naivety to gemcitabine might therefore be irrelevant. Furthermore, a relevant comparator should not be excluded based on issues with the data. Indeed, if that was a valid argument, the other comparisons could not be performed either because no RCTs were available. The company could have adjusted the available data based on expert opinion. It is the ERG's view that the company did not present valid arguments to exclude cisplatin plus gemcitabine as a comparator and the ERG will therefore include this comparison in its base-case based on the data from Gondo (2011)[13].

5.2.5 Perspective, time horizon and discounting

The model adopts the perspective of the NHS and Personal and Social Services (PSS) in England and Wales. The cycle length is four weeks to account for the length of treatment cycles. A lifetime time horizon was adopted to capture all relevant costs and health-related utilities. All costs and utilities were discounted at a rate of 3.5% per year.

ERG comment: The ERG considers the adopted perspective, time horizon and discounting to be appropriate for this appraisal.

5.2.6 Treatment effectiveness and extrapolation

Parametric time-to-event models were used to estimate OS, PFS and TTD in the company's cost effectiveness model. A response-based approach was adopted to estimate OS and PFS, but not for TTD in the company's base-case.

5.2.6.1 OS and PFS of nivolumab

The parametric time-to-event models representing OS and PFS of nivolumab were informed by the CheckMate 032 and CheckMate 275 trials, which are both single arm trials.[10, 11] The time-to-event data of both trials were combined (pooling method not stated) to perform the survival analyses described in the following sections.

Response-based and landmark analyses

The company implemented a response-based analysis to estimate OS and PFS of the nivolumab arm because it claimed that standard survival modelling approaches would not appropriately characterise the novel mechanism of action of nivolumab, i.e. responders may have long and durable response to treatment leading to extended survival. Therefore the company suggested that standard parametric time-to-event models were not deemed flexible enough to characterise the change in hazard over time resulting from having (long-term) responders, and non-responders (no supporting evidence provided).[2]

The company used a landmark analysis to prevent the occurrence of the immortal-time bias. In this landmark analysis, OS and PFS of both groups (responders and non-responders) were estimated together until a specified landmark point after which different survival curves were fitted separately for each group. For the base-case analysis, the company chose an eight-week landmark point, which corresponds to the median time to response in both CheckMate 032 and CheckMate 275 trials (1.87 and 1.48 months in CheckMate 275 and CheckMate 032, respectively). Before this eight-week landmark point, the Kaplan-Meier estimates for the whole group were used to estimate OS and PFS. After the landmark point, parametric time-to-event models were fitted to the responders' and non-responders' survival data for the remainder of the time horizon, and adjusted for background mortality.

A sensitivity analysis explored the impact of using a 26-week landmark point, with the justification that, at that time point, '*all patients had responded while leaving a sufficiently long observational period for further extrapolation.*'[2]

ERG comment: The main concerns of the ERG were (1) the method used for pooling both CheckMate 032 and CheckMate 275 trials, (2) the use of response-based analysis, (3) the use of landmark analysis to model PFS and OS of nivolumab, and (4) the use of KM estimates up to the chosen landmark.

(1) The CS reported that data from both CheckMate 032 and CheckMate 275 studies were pooled without stating which method was used to pool the data. Upon request from the ERG, the company explained that OS and PFS data from both studies were combined without adjustments because there was no evidence of differences between the studies based on a Wald test. Hence, the pooled CheckMate studies dataset contained 348 patients (78 patients from CheckMate 032 and 270 patients from CheckMate 275).[7] Concerns with pooling from CheckMate 032 and CheckMate 275 studies were outlined in Section 4.2.6.

(2) The company justified the use of a response-based approach stating that standard parametric time-to-event models were not flexible enough to characterise the change in hazard over time due to possible sustained and long-term response to treatment. However, the ERG noted that most standard parametric time-to-event models include changing hazards over time; some standard parametric time-to-event models allow for non-monotonic changing hazard functions over time (i.e. log-logistic, log-normal and generalised gamma distributions). The company did not provide any mathematical reasoning to support their argument that a different response cannot be accurately described by standard parametric survival models. The ERG considers that based on visual inspection of the not response-based, conventional survival analysis alone, the case for response-based analysis might not be supported, as the parametric time-to-event model fitted to OS made a good fit and the model for PFS could be regarded as providing a reasonable fit (see Figures 5.2 and 5.3).

The company's second argument in favour of the landmark analysis was that it was implemented to address concerns from previous appraisals of nivolumab in which standard parametric time-to-event models were not deemed suitable to model survival with nivolumab treatment.[7] The company argued in response to clarification questions that landmark analysis '*allows for a more flexible shape to the*

nivolumab survival curve whilst adhering to the Committee's previous preference of using the trial data for a proportion of the survival curves.'[7] The ERG considers that a standard approach should be shown to be inappropriate in the particular decision problem at hand before discarding it and the company failed to do so, as described in the previous paragraph.

The ERG requested that the company justify whether alternative methods (e.g. spline models, mixture cure models) were considered instead of the landmark analysis because spline models are suggested in the NICE DSU TSD 14 as a flexible alternative to standard parametric time-to-event models (while the landmark approach is not mentioned).[38] The company responded that spline models were generally not accepted in previous appraisals of nivolumab and that the acceptability of mixture cure models for HTA bodies is yet unknown. The ERG considers that this is not a valid argument given that spline models and mixture cure models are recommended in the TSD.

In conclusion, the company (a) did not provide sufficient evidence to demonstrate that conventional parametric time-to-event models failed to describe nivolumab survival, (b) did not provide evidence to support that the committee's criticisms on previous nivolumab appraisals applied to the current appraisal, and (c) did not provide evidence to demonstrate that the landmark analysis provided more valid results than standard survival modelling analyses or alternative methods recommended in NICE DSU TSD 14 (for example, no expert opinion was used to validate the resulting curves).

(3) The ERG's third concern is the choice of the eight-week landmark. The choice of the eight-week landmark was based on the collected evidence while it is advised to determine the landmark point a priori to the analysis in order *'to safeguard the analysis against the danger of a data-driven decision'*.[39] Therefore, the ERG asked the company to investigate the influence of a 12- or 20-week landmark point on the results but these analyses were not provided by the company due to time constraints. As demonstrated in a previous nivolumab appraisal, the choice of the landmark point may not have a linear relationship with the ICER.[40] Hence, the influence of this assumption, i.e. the arbitrarily post-hoc selected landmark point, on the results is highly unpredictable.

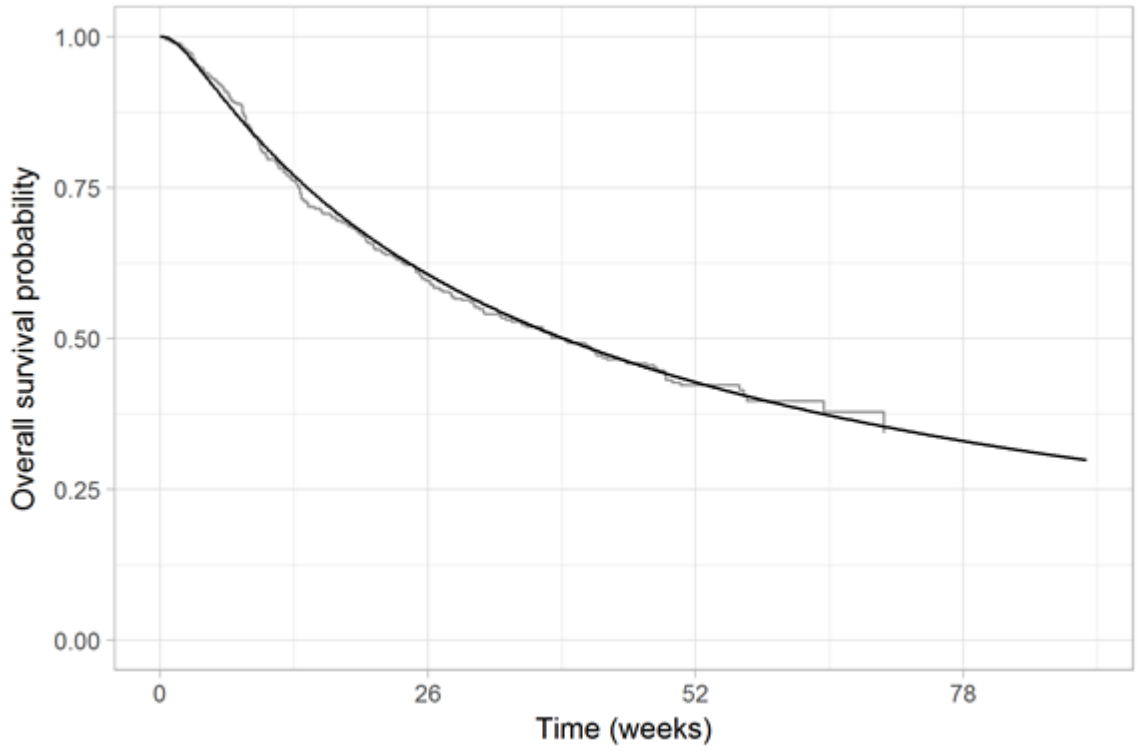
(4) The ERG asked the company to justify why the Kaplan-Meier estimates were used until the landmark point instead of a parametric time-to-event model, and to provide the results of an analysis using a parametric time-to-event model until the landmark point. The company did not provide the results of such analysis and responded that using the Kaplan-Meier estimates until the landmark point reflected the *'Committee's previous preference of using the trial data for a proportion of the survival curves.'*, not clearly referring to a specific technology appraisal.[7] According to the company, using a parametric time-to-event model would also add unnecessary complexity to the model. The ERG does not consider these arguments to be valid: a previous precedent does not relieve the company from demonstrating appropriateness of their method, and fitting a distribution to the data up to the landmark does not present more complexity than making Kaplan-Meier estimates probabilistic. The ERG therefore prefers the use of a parametric time-to-event model to estimate survival until the landmark point to avoid the problem of overfitting when using Kaplan-Meier estimates. The possibility for this analysis was, however, not included in the company's model.

In conclusion, the company deviated from the NICE TSD recommendations by using a response-based analysis. However, the company did not demonstrate (1) that conventional modelling approaches of survival failed to correctly characterise the OS and PFS of nivolumab, and (2) that the response-based approach resulted in estimates that could be considered more realistic than the standard approach. The uncertainty about whether this approach more accurately reflects prognosis for patients treated with nivolumab was exacerbated by additional assumptions required for response-based analysis, such as, most crucially, the choice of the landmark point, which has an unpredictable effect on results. Fitting

parametric models to the responder and non-responder groups also results in larger uncertainty about these fitted curves: the sample size used is significantly smaller, a) because of the splitting up of the study population into two groups and b) because only the available data after the landmark is used. The fact that responder and non-responder groups had to be combined for the indirect comparison casts further doubt over whether the response-based analysis has any benefits (hazard ratios are derived from the overall population and are then applied in a combined responder and non-responder population, as described below in the section on relative treatment effectiveness). It should also be noted that response-based and conventional approaches result in vast differences in the predicted life-years for nivolumab, with a predicted mean of 2.80 life years in the response-based analysis and 1.84 life years in the conventional, not response-based, approach (deterministic estimates). No explanation for this deviation was provided, and these estimates were not validated using expert opinion.

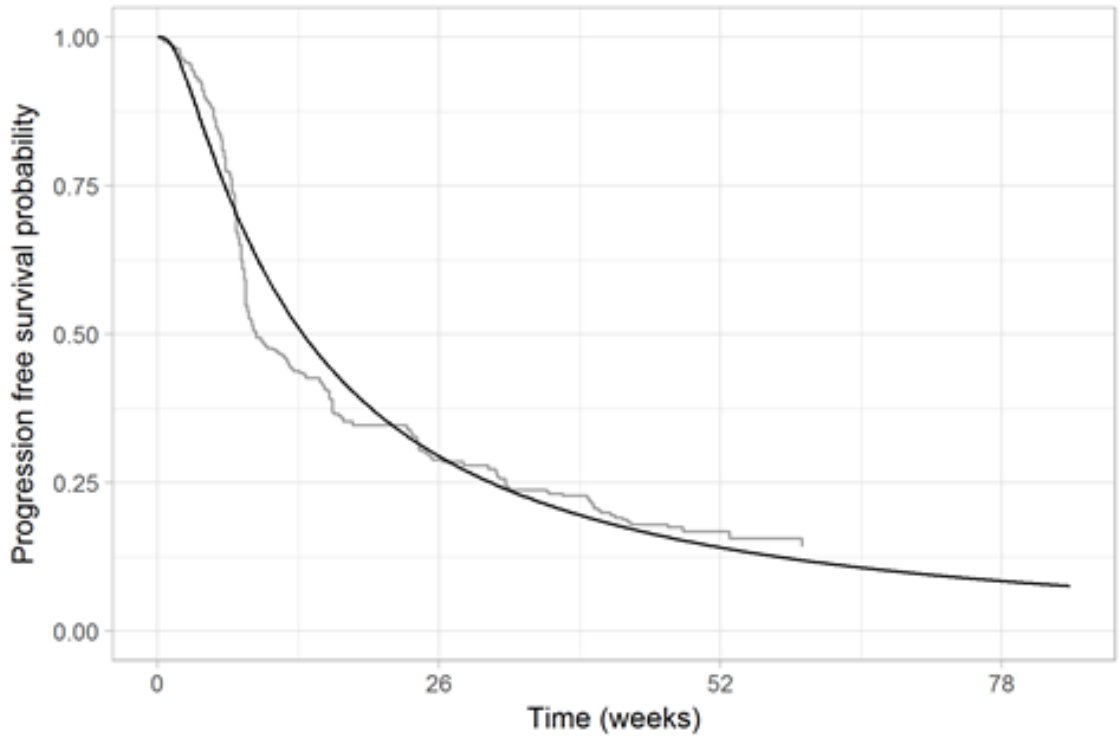
For the aforementioned reasons, and in line with the TSD recommendations, the ERG used the conventional approach of fitting parametric time-to-event models to the overall population in its base-case analysis. Based on statistical fit and visual inspection, the ERG considers the distributions preferred by the company (i.e. the generalised gamma for both OS and PFS) to be the most plausible in its base-case analysis (Figures 5.2 and 5.3). Alternative distributions are explored in scenario analyses. The ERG also explored the use of a response-based analysis in scenario analyses.

Figure 5.2: Standard parametric time-to-event model for overall survival (generalised gamma distribution)



Source: Appendix L, figure 114

Figure 5.3: Standard parametric time-to-event model for progression-free survival (generalised gamma distribution)



Source: Appendix L, figure 120

Time-to-event models selection for OS and PFS estimations of nivolumab

Parametric time-to-event models were fitted separately to the OS and PFS data of the responder and non-responder groups (without investigating the proportional hazard assumption through log-cumulative hazard plots). The company stated that the following six parametric distributions were fitted to the OS and PFS data as recommended by the NICE Decision Support Unit Technical Support Document 14[38]:

- Exponential
- Weibull
- Gompertz
- Lognormal
- Log-logistic
- Generalised gamma

The parametric time-to-event models used to estimate OS and PFS were selected based on statistical fit (Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC)) and visual inspection. Table 5.3 provides an overview of the statistical fit of the different distributions for OS and PFS in the responder and non-responder groups.

The company considered the model selection for OS and PFS (in both responders and non-responders groups) simultaneously and selected the generalised gamma distribution to represent OS and PFS of both responder and non-responder groups. The generalised gamma distribution was selected because 1) it was the best fitting distribution based on 3 out of 8 criteria (see numbers printed in bold in Table 5.3), and 2) the Weibull distribution (which was the best fitting distribution based on 4 out of 8 criteria) provided a poor fit to the responders' OS and PFS (unclear how this was determined). Hence, the

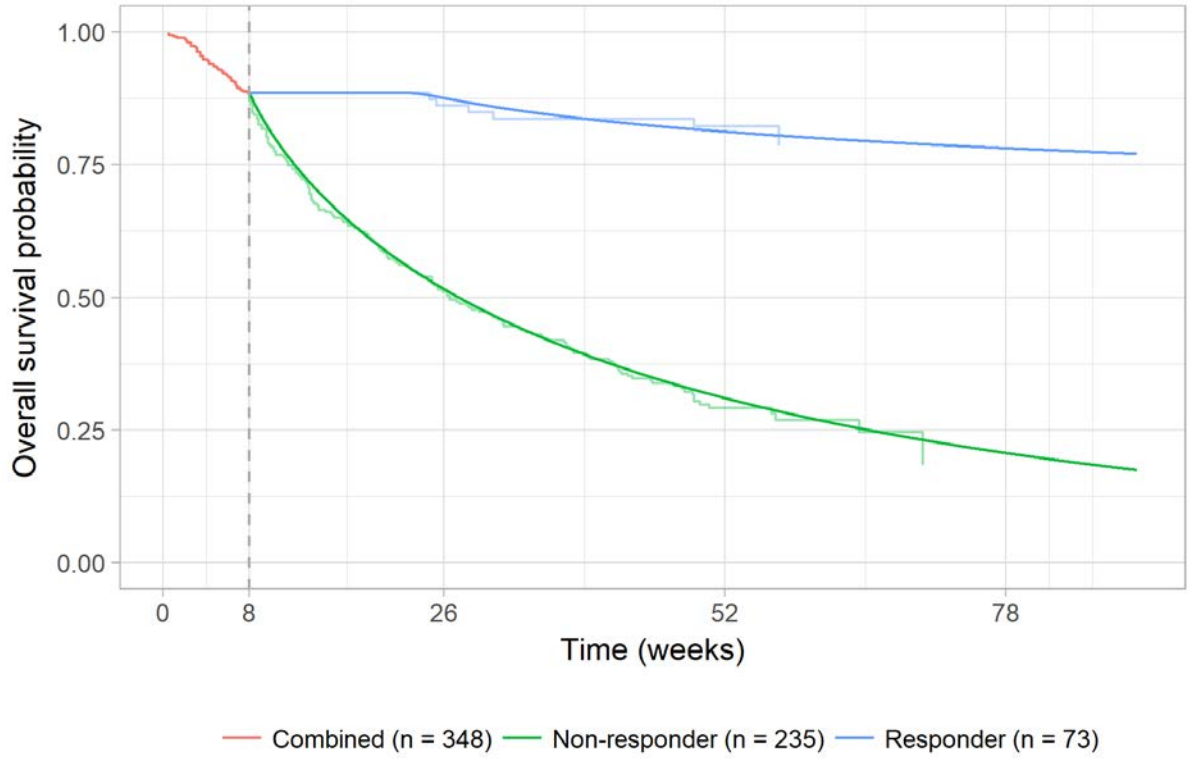
company concluded that the generalised gamma provided the best fit overall. Experts were not consulted to support the selection of the parametric time-to-event models applied to the responder and non-responder groups. Figures 5.4 and 5.5 present the landmark analyses for OS and PFS based on responders' status.

Table 5.3: Statistical fit measures of the distributions representing OS and PFS in the responder and non-responder groups at the eight-week landmark

Distribution	OS				PFS			
	Responders		Non-responders		Responders		Non-responders	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
Exponential	90.1	**92.4**	1402.7	1406.1	276.9	279.2	787.8	790.6
Weibull	91.1	95.7	**1393.3**	**1400.2**	266.9	271.5	**763.4**	**769.2**
Gompertz	91.9	96.4	1395.4	1402.3	273.1	277.7	780.7	786.4
Lognormal	90.4	95.0	1397.4	1404.3	262.4	267.0	773.1	778.8
Log-logistic	91.0	95.6	1394.4	1401.3	264.6	269.2	776.7	782.4
Generalised gamma	**87.9**	94.8	1394.5	1404.8	**256.6**	**263.5**	765.0	773.5

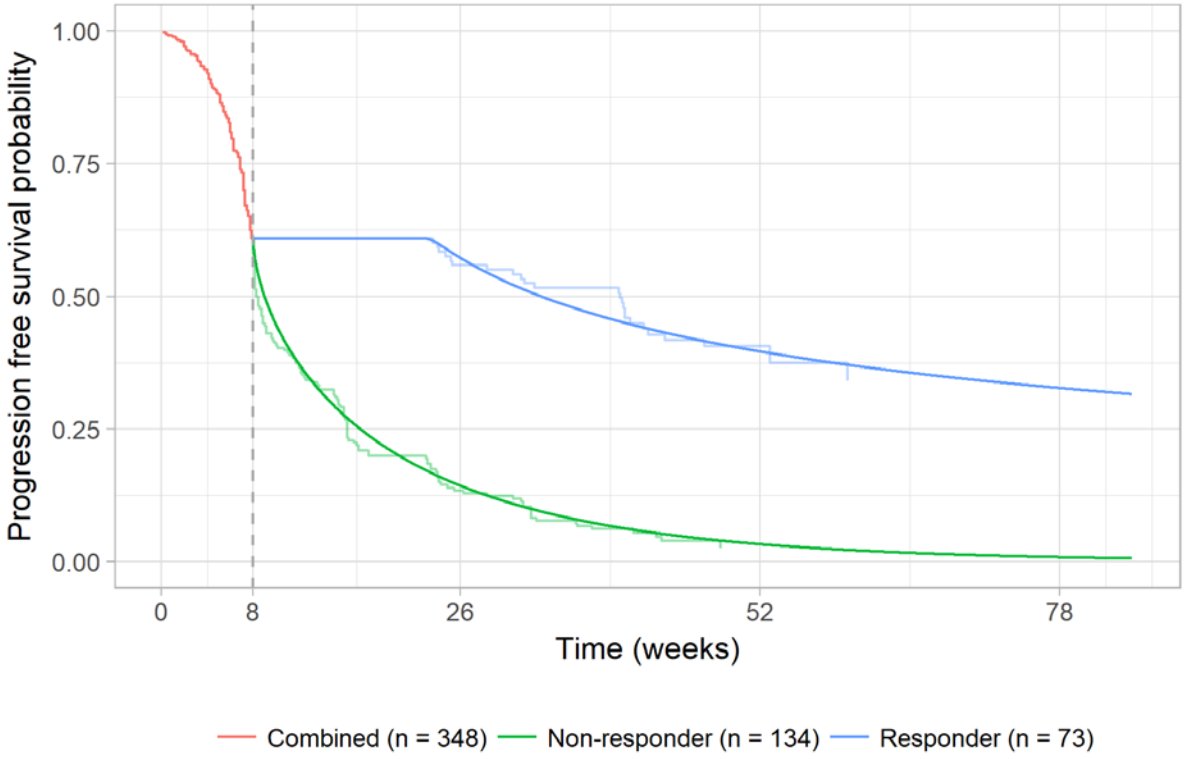
Source: Adapted from Table 29 of the CS[2]
Abbreviations: AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; OS, overall survival; PFS, progression-free survival.
Bold printed values represent the distributions with the lowest AIC or BIC (i.e. the 'best fitting' time-to-event models)

Figure 5.4: Week 8 landmark – overall survival with generalised gamma[a]



Source: Response to clarification letter, Figure 34[7]
[a] The ERG requested corrected figures because the number of responder was incorrect in the original CS

Figure 5.5: Week 8 landmark – progression-free survival with generalised gamma[a]



Source: Response to clarification letter, Figure 35[7]

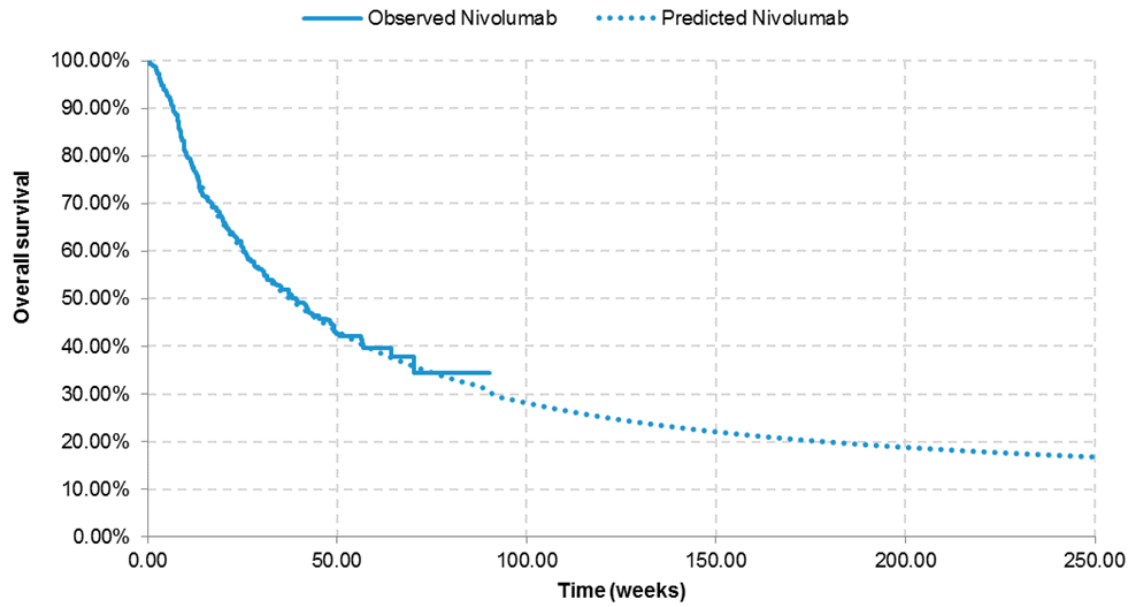
[a] The ERG requested corrected figures because the number of responder was incorrect in the original CS

In order to implement the parametric time-to-event models in the cost effectiveness model, OS and PFS estimates obtained from the parametric time-to-event models estimated for responders and non-responders separately were combined by using a weighted average). This weighting was based on the proportion of responders in patients being progression-free and alive at the eight-week landmark point (based on both CheckMate 032 and CheckMate 275 trials[10, 11], and was assumed to stay constant for the remainder of the time horizon.

Figures 5.6 and 5.7 present the survival curves as used in the base-case analysis for OS and PFS, respectively, compared to the observed OS and PFS obtained with nivolumab. These curves are the result of the weighted average of the responders' and non-responders' OS and PFS estimates, and are compared to the OS and PFS Kaplan-Meier estimates of the pooled CheckMate studies dataset.

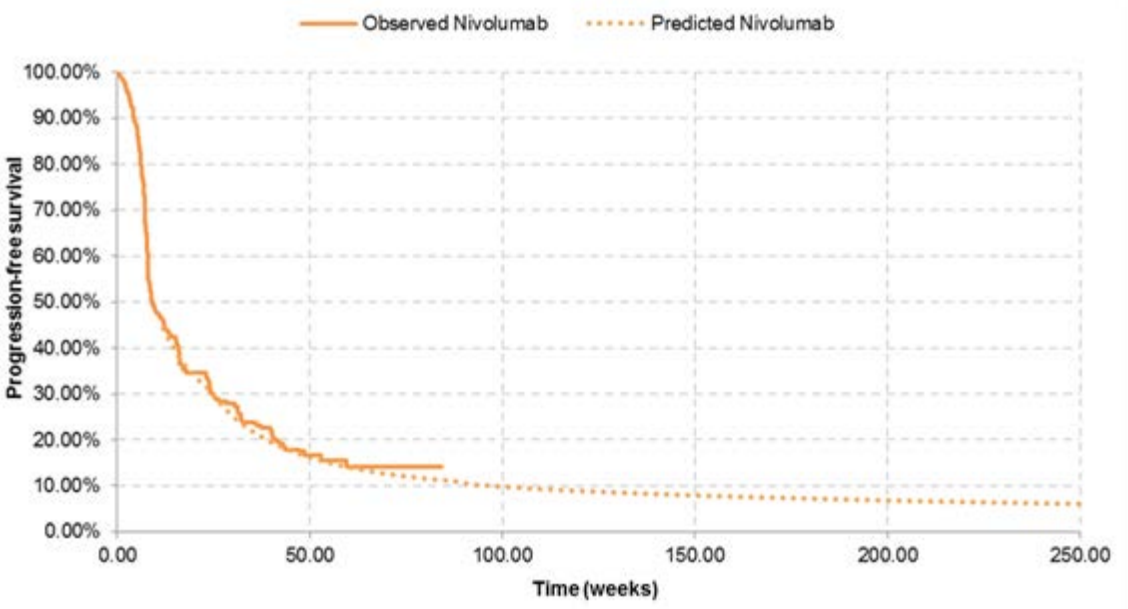
Figure 5.6: Kaplan-Meier estimate of OS with nivolumab, based on the pooled CheckMate 032 and CheckMate 275 trials dataset ('Observed Nivolumab') compared to the predicted values

based on the landmark and response-based analysis (generalised gamma distribution) ('Predicted Nivolumab')



Source: Figure 37 of the CS[2]

Figure 5.7: Kaplan-Meier estimate of PFS with nivolumab, based on the pooled CheckMate 032 and CheckMate 275 trials dataset ('Observed Nivolumab') compared to the predicted values based on the landmark and response-based analysis (generalised gamma distribution) ('Predicted Nivolumab')



Source: Figure 36 of the CS[2]

ERG comment: The main issues concerning the selection of the parametric time-to-event models are (1) the rejection of the proportional hazard assumption between responders and non-responders, (2) the simultaneous selection of the parametric time-to-event models, (3) the lack of expert consultation, and

(4) the combination of the responders' and non-responders' curves at a weight which stays constant over time.

(1) The company assumed in its base-case analysis that the proportional hazard assumption did not hold between responders and non-responders, but did not provide log-cumulative hazard plots to support this assumption. Upon the ERG's request, the company provided the log-cumulative hazard plots and concluded that the proportional hazard assumption could potentially be valid for OS but not for PFS. However, the company did not assume proportional hazards *'as this meant there was no requirement to assume the same distribution to be appropriate for both responder and non-responder curves'*.[7] The ERG does not agree with this argument since the proportional hazard assumption seemed to hold for OS, and could potentially also hold for PFS, based on the examination of the log-cumulative hazard plots. No additional evidence was provided to discard the proportional hazard assumption based on clinical implausibility of the assumption. The influence of assuming proportionality of hazards and using a hazard ratio on one of the curves on the results was not investigated by the company.

(2) In the base-case model, the company selected the same distributions (generalised gamma) for responder and non-responder groups without justifying why. This contradicts the company's argument that there was *'no requirement to assume the same distribution to be appropriate for both responder and non-responder curves'*.[7] This decreased the flexibility allowed by the different parametric time-to-event models. In response to the clarification questions, an updated model was provided by the company, which allowed the selection of different parametric time-to-event models for responders and non-responders.

(3) The NICE DSU TSD 14 recommends to consult clinical experts to support the choice of the parametric time-to-event models besides using statistical fit and visual inspection.[40] According to the CS and response to clarification questions,[7] clinical experts were only consulted during an advisory board. The survival curves presented during this advisory board were fitted to the CheckMate 275 trial only and did not include response-based analysis.[5] The final parametric time-to-event models were therefore not validated using expert opinion.

(4) The parametric time-to-event models were fitted separately to responders and non-responders and were weighted based on the proportions of responders and non-responders at the landmark point. This inflated the proportion of non-responders in later periods because the proportion of responders is expected to increase over time compared to the proportion of non-responders. This assumption is likely to be conservative but it is not clear, and, as described in Section 5.2.6.1, using different landmark points may have an unpredictable influence on the results.

In conclusion, most issues identified in the selection of parametric time-to-event models are avoided by using conventional analysis, as opposed to response-based analysis. These issues include the pooling of responder and non-responder groups, making assumptions about proportional hazards between the two groups and the potential for using differential curves for responders and non-responders. Therefore, the ERG used the conventional approach in its base-case analysis using the company's base-case and alternative parametric time-to-event models. As mentioned before, the influence of using a response-based analysis will be explored in the ERG's scenario analyses, using the company's base-case and alternative parametric time-to-event models.

Background mortality

After 88 weeks, general population mortality estimates were used to adjust OS and PFS estimations. This was implemented in order to '*appropriately characterise the relationship between age and*

increasing risk of death.'[2] To avoid double-counting, general population mortality estimates were applied from the 88th week onwards, which represented the end of the CheckMate 032 and CheckMate 275 studies' follow-up. This adjustment was implemented by multiplying the survival estimates obtained from the parametric time-to-event model estimating OS (described in previous sections) by the probability of being alive according to age-adjusted UK life tables.

ERG comment: The ERG's comments relate to (1) an error in the calculation of background mortality, (2) the use of an age distribution to calculate background mortality, and (3) the implementation of adjusting OS and PFS by background mortality.

(1) When reviewing the cost effectiveness model, the ERG noted that the mortality rates implemented in the model did not match the values reported by the Office of National Statistics UK life tables. The ERG therefore used the correct age-adjusted background mortality rates and fixed the conversion of the background mortality rate into a probability.

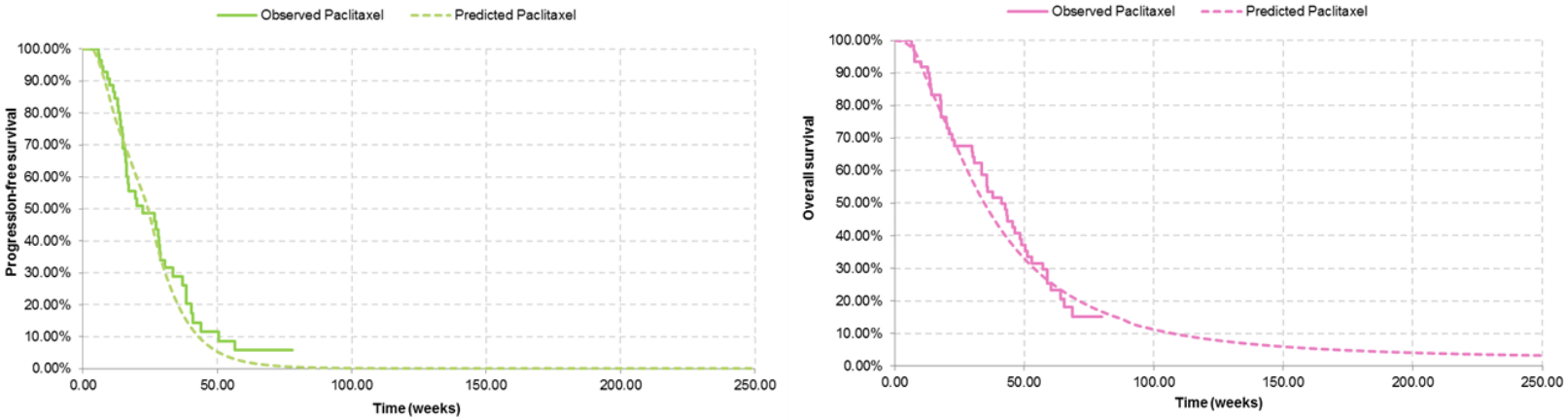
(2) Not in line with conventional methods of incorporating background mortality in parametric survival models, the company used a distribution of age instead of a fixed mean age, to reflect patient heterogeneity. This resulted in slightly higher background mortality compared to standard background mortality estimates. Despite this being unconventional in cohort models, the ERG considers that it is appropriate to reflect patient heterogeneity in the calculation of background mortality.

(3) The conventional approach seen in many technology appraisals is to implement a maximum function to incorporate general UK population mortality data in the cost effectiveness model, to ensure that the probability of dying does not become lower than the probability of dying based on the age-adjusted UK life tables. However, the company's approach of implementing this background mortality by multiplying OS by the probability of being alive based on the age-adjusted UK life tables, was viewed as appropriate. Lastly, any adjustment for background mortality should be applied to responder and non-responder groups separately, if response-based analysis is used. However, the company applied it to the combined responder and non-responder groups, which, due to the different prognoses in both groups, is inappropriate. This issue becomes redundant with a conventional, not response-based analysis. PFS was not directly adjusted using the general population mortality data but a minimum function was implemented to ensure that PFS did not become higher than OS.

5.2.6.2 Relative effectiveness of nivolumab

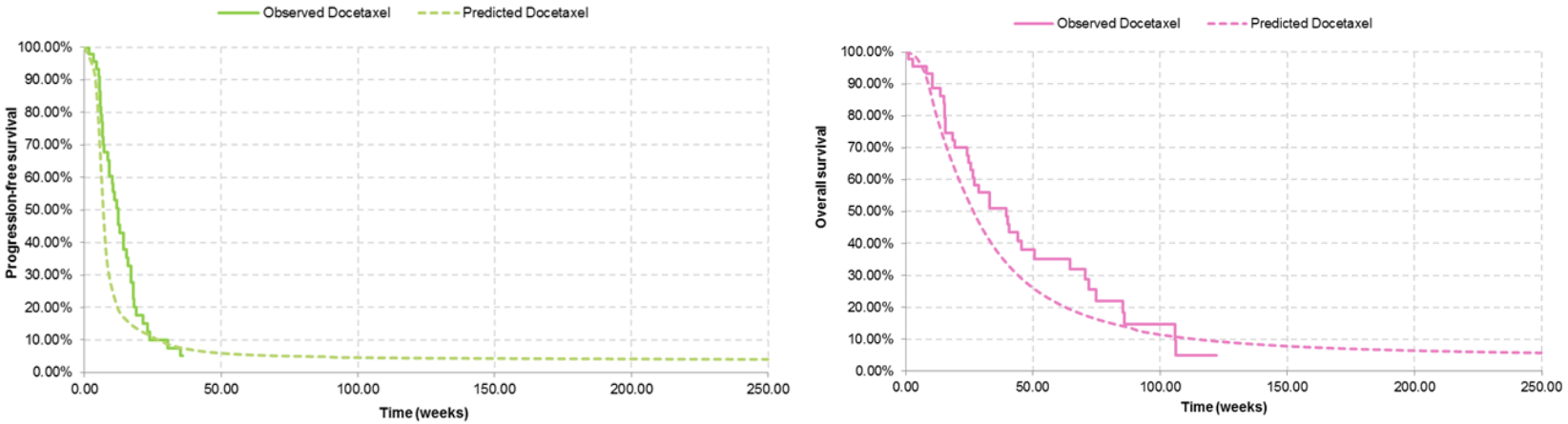
The relative effectiveness of nivolumab versus the comparators was modelled through time-varying hazard ratios (HRs) because the '*proportional hazard assumption did not hold for these comparisons given the unique mechanism of action for nivolumab*'.[2] No evidence was provided to support the violation of the proportional hazard assumption. A STC was performed to obtain these time-varying HRs. More detail about this methodology is provided in Section 4.4.1. The STC was performed based on the pooled CheckMate 032 and CheckMate 275 trials dataset, in which response status was not taken into account. The HRs obtained from the STC were then applied to the combined parametric time-to-event models of nivolumab which took response status into account. Figures 5.8 to 5.9 present the survival curves estimating OS and PFS of each comparator, obtained by applying the time-varying HRs to the combined survival curves of nivolumab (Figures 5.10 and 5.11), compared to the Kaplan-Meier estimates observed in the comparator studies. The company explained that the predicted OS and PFS of the comparators were mostly lower than the observed OS and PFS, especially for docetaxel, because of the differences in patient characteristics between the comparator trials and the CheckMate 032 and CheckMate 275 studies.

Figure 5.8: Progression-free survival and overall survival with paclitaxel – observed and predicted values with the generalised gamma distribution



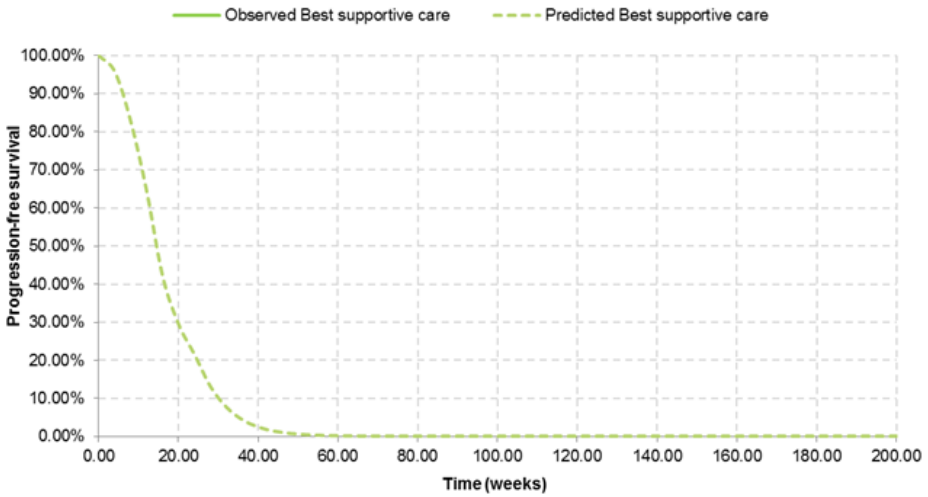
Source: Figure 40 of the CS

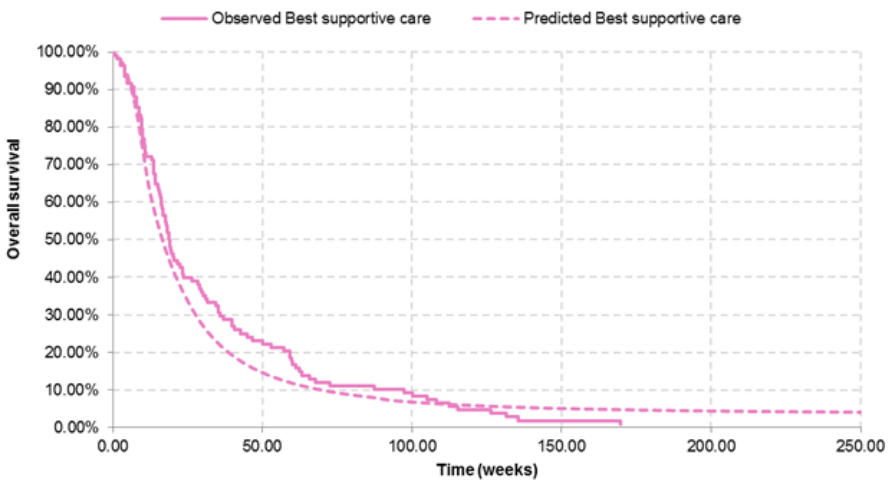
Figure 5.9: Progression-free survival and overall survival with docetaxel – observed and predicted values with the generalised gamma distribution



Source: Figure 40 of the CS

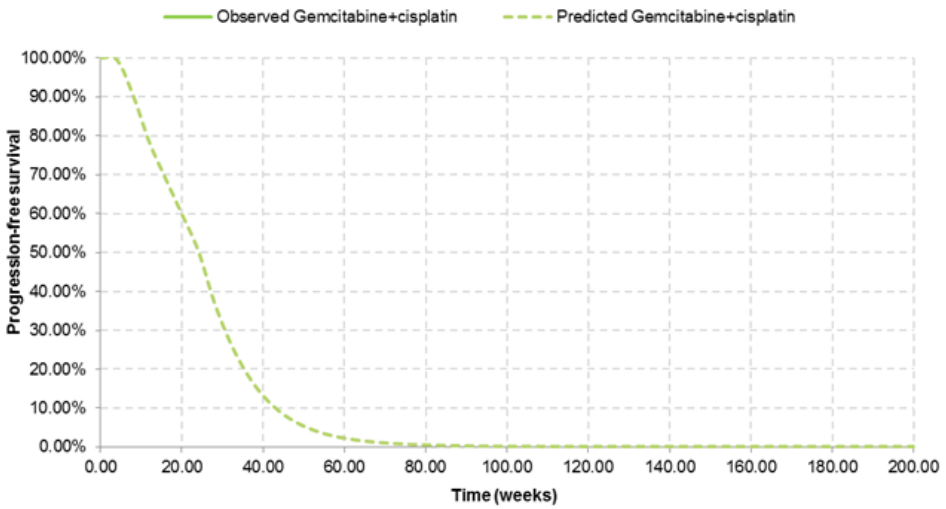
Figure 5.10-free survival and overall survival with best supportive care – observed and predicted values with the generalised gamma distribution[a]

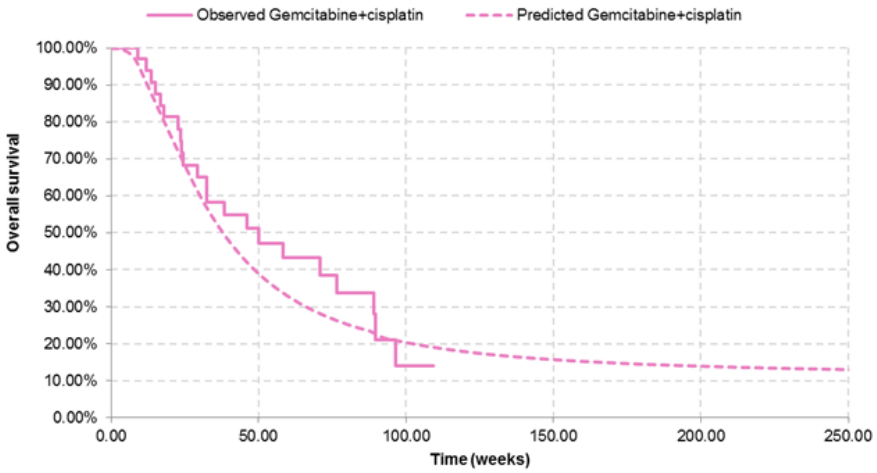




[a] No observed progression-free survival data were identified for best supportive care
Source: Figure 42 of the CS

Figure 5.11: Progression-free survival and overall survival with cisplatin plus gemcitabine – observed and predicted values with the generalised gamma distribution[a]





[a] Cisplatin plus gemcitabine treatment was analysed as a scenario analysis. No observed progression-free survival data were identified for cisplatin plus gemcitabine
Source: Figure 43 of the CS

Best supportive care (BSC) was not included in the STC for PFS due to a lack of relevant PFS data identified in the clinical SLR (Section 4.3). Therefore, the company assumed that the HR for BSC versus paclitaxel (1.47) was equivalent to that of BSC versus vinflunine for second-line UC patients (Bellmunt et al. (2009)[26]). The company assumed that this HR could be applied to the paclitaxel PFS curve to estimate the PFS of BSC, due to the similarities in terms of outcomes between vinflunine and paclitaxel/docetaxel. This HR was held constant during the time horizon of the cost effectiveness model, due to the absence of alternative data. No evidence was provided to support these assumptions.

Cisplatin plus gemcitabine was not included in the STC for PFS due to a lack of relevant PFS data identified in the clinical SLR (Section 4.4). The HR of paclitaxel versus nivolumab was applied to estimate the PFS of cisplatin plus gemcitabine because the company expected that paclitaxel and cisplatin plus gemcitabine would provide similar PFS results since they are all chemotherapy agents. No evidence was provided to support this assumption.[20]

ERG comment: The ERG's concerns include (1) the uncertainty and bias induced by comparing single-arm studies, (2) the discrepancy in populations in which relative effectiveness estimates are derived and applied, (3) the need for and effect of applying time-dependent HRs instead of time-independent HRs, (4) the estimation of HRs for PFS of BSC and cisplatin plus gemcitabine, and (5) the large impact of the parameter values used for the fractional polynomial NMA model.

(1) As described in Section 4.6, the STC and NMA performed by the company to obtain time-dependent HRs were associated with considerable uncertainty and the introduced bias associated with the STC was not quantified. For these reasons, the cost effectiveness analysis performed by the company suffers from significant uncertainty and potential bias. As stated in NICE DSU TSD 18 for STC's incorporating one-arm studies only, the accuracy of the resulting estimates is entirely unknown and without any evidence that the STC reduces the systematic error, the results *'are not worthy of consideration'*.[1]

(2) An additional concern is that the time-dependent HRs were obtained based on a comparison using the pooled CheckMate 032 and CheckMate 275 trials dataset, which did not take response status into account. Instead, the HRs for all patients (regardless of response status) were applied to the combined parametric time-to-event models, which accounted for response status. More specifically, the same time-dependent HRs were applied to the combined survival curves based on the weighted average of the responders and non-responders time-to-event models. Hence, there is a discrepancy between the a priori population on which the relative effectiveness is based on the a posteriori population in which the HRs are applied. The potential bias introduced by this methodology was not investigated by the company, despite a request in the clarification questions.[7] The ERG notes that applying HRs to the combined survival curves may underestimate the relative effectiveness in the responders group, but overestimate the relative effectiveness in the non-responders group. The ERG would have preferred to apply separate HRs to responders and non-responders, however, these were not provided by the company. This concern is redundant when using the conventional, not response-based, approach. The ERG further noticed that the code supplied to estimate the time-dependent HRs only estimated them up to a time horizon of 256 weeks, ending much before the end of the model time horizon. It is not clear where the time-dependent HRs implemented after 260 weeks were sourced from.

(3) The company applied time-dependent HRs to model the relative effectiveness of nivolumab versus the comparators because it assumed that the proportional hazard assumption did not hold. The company did not consult log-cumulative hazard plots to support this assumption, as recommended by the NICE DSU TSD 14[38]. Upon the ERG's request, the company provided the log-cumulative hazard plots of nivolumab versus the comparators. Based on these plots, the company confirmed that the proportional hazard assumption did not hold. The ERG considers that the proportionality of hazards could not be

ruled out based on the company's analyses because both CheckMate 032 and CheckMate 275 trials were presented separately in these plots, while the HRs were derived based on the pooled CheckMate 032 and CheckMate 275 trials dataset. Therefore, these plots did not allow investigation as to whether the proportional hazard assumption held for the analysis performed by the company. Because the company did not provide sufficient evidence to support the violation of the proportional hazard assumption and to support the need for time-dependent HRs, the ERG requested scenario analyses using time-independent HRs (i.e. fixed for the entire time horizon) to estimate the relative effectiveness of nivolumab versus the comparators. The company provided a network meta-analysis using fixed and random effects to estimate time-independent HRs in its response to the clarification letter.[7] These time-independent HRs were still in favour of nivolumab, except for cisplatin plus gemcitabine, which became more effective than nivolumab. The use of these time-independent HRs increased all cost effectiveness estimates (Section 5.2.10). The company did not consider these scenario analyses to be appropriate for decision making because a) the survival estimates for the comparator arm were considered to be implausible overestimations and b) the proportional hazard assumption was violated. The ERG considers that these claims were not strongly supported by the evidence submitted by the company. In response to a), the company only presented a single parametric time-to-event model to illustrate the overestimation of survival in the comparator arms but different parametric time-to-event models could lead to different results and a better fit with the data. In addition, the ERG notes that using time-independent HRs has the advantage of preventing over-parameterisation which might occur when estimating time-dependent HRs with the relatively little amount of data submitted by the company. In response to b), as stated above, the violation of the proportional hazard assumption was not demonstrated sufficiently by the company.

(4) Finally, the HRs used to estimate PFS of BSC and cisplatin plus gemcitabine were not obtained through the STC but were based on assumptions, which were not supported by clinical evidence (i.e. same HRs for BSC vs paclitaxel as for BSC vs vinflunine and same HRs for cisplatin plus gemcitabine versus nivolumab as for paclitaxel versus nivolumab).[2, 7] The assumption that PFS when treated with cisplatin plus gemcitabine is the same as when treated with paclitaxel is likely non-conservative. The ERG performed scenario analyses to investigate the influence of alternative time-dependent HRs for BSC and cisplatin plus gemcitabine PFS on the cost effectiveness results. In these scenario analyses, the time-dependent HRs obtained for OS of BSC and cisplatin plus gemcitabine were used. These time-dependent HRs were selected because they were based on evidence concerning the drug of interest instead of being based on assumptions lacking supporting evidence. However, the ERG is aware that the relative effectiveness of a treatment compared to another may change across different outcomes.

(5) The use of the fractional polynomial model introduces some uncertainty into the cost effectiveness analysis. The company showed the effects of a set of alternative p1 and p2 values on the ICERs, showing that their base-case ICERs increased significantly. In response to clarification questions the company enabled in the model 10 different p1 an p2 values, resulting in 100 possible combinations. It is the ERG's concern that these different combinations could have an unpredictable effect on model outcomes and the ERG therefore explored the range of ICERs that could be obtained through a 'mini-PSA', in which 10,000 draws from different combinations of these parameter values are used in the model. Whilst implementing this, the ERG noted that certain combinations of parameter values result in extreme hazard ratios and survival estimates above 100%, showing that not all of these are plausible candidates. The ERG adjusted survival estimates to prevent this problem from occurring in their mini-PSA.

5.2.6.3 Time to treatment discontinuation

Treatment with nivolumab should continue *'as long as clinical benefit is observed or treatment is no longer tolerated by the patient.'*[2] Time-to-treatment discontinuation (TTD) was estimated through a parametric time-to-event model. The same (six) distributions as for OS and PFS were fitted to the pooled CheckMate 032 and CheckMate 275 studies' dataset and statistical fit of the different curves was assessed through the AIC and BIC (Table 5.4). In the CS, TTD was estimated independent of response status.

The generalised gamma distribution was selected to estimate TTD in the base-case analysis, with the company claiming that this was done to ensure consistency with the curves selected to represent OS and PFS. The Gompertz and log-logistic distributions showed better statistical fit than the generalised gamma distribution but the company argued that these two distributions produced long tails with patients still being on treatment after 5 and 10 years, which lacked clinical validity (Table 5.5). The impact of using alternative distributions to estimate TTD was explored in sensitivity analyses.

Table 5.4: TTD estimation based on different parametric time-to-event models

Time	TTD estimation		
	Generalised gamma	Gompertz	Log-logistic
1 year	17.6%	21.4%	22.1%
2 year	8.3%	16.7%	12.7%
3 year	5.1%	15.9%	8.9%
4 year	3.2%	15.8%	6.9%
5 year	2.1%	15.8%	6.0%
10 year	0.2%	15.8%	2.8%

Source: company's cost effectiveness model
[a] Parametric time-to-event model used in the company's base-case analysis

TTD of the comparators was based on their respective PFS curves because it was assumed that comparator treatment would continue until disease progression or unacceptable toxicity. Treatment with paclitaxel was assumed to stop after 6 (model) cycles (if treatment was not discontinued yet), i.e. 24 weeks. This represented the clinical use of paclitaxel in the UK[15] and was confirmed by clinical experts.[5] The company assumed that BSC was administered until death.

Table 5.5: Statistical fit measures of the distributions representing time to treatment discontinuation

Endpoint	Distribution	AIC	BIC
Time to treatment discontinuation	Exponential	2381.86	2385.71
	Weibull	2329.96	2337.67
	Gompertz	**2318.29**	**2325.99**
	Lognormal	2341.69	2349.40
	Log-logistic	2322.93	2330.63
	Generalised gamma	2328.48	2340.04

Source: Table 30 of the CS[2]
Abbreviations: AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion
Bold printed values represent the distributions with the lowest AIC or BIC (i.e. the 'best fitting' time-to-event models)

ERG comment: The ERG's concerns include (1) inconsistency in estimating TTD compared with estimating OS and PFS (the use of a conventional, not response-based, approach to estimate TTD), and (2) the choice of parametric distributions for TTD.

(1) Unlike OS and PFS, the parametric time-to-event models estimating TTD were not estimated based on a landmark and response-based analysis but on the pooled CheckMate 032 and CheckMate 275 trials dataset. This was inconsistent with the analysis of OS and PFS and no justification was provided. The ERG requested from the company to implement a response-based, landmark, analysis for TTD, assuming that treatment duration may be influenced by response status, especially given that treatment with nivolumab should continue *'as long as clinical benefit is observed or treatment is no longer tolerated by the patient.'*[2] The company provided an updated cost effectiveness model in which TTD can be estimated in the same way as OS and PFS, i.e. using a response-based analysis. However, the ERG noticed that the company calculated the proportion of responders and non-responders based on the sum of patients in the OS and PFS health states, thereby double-counting patients. The ERG considered it more appropriate to use all responders alive for the calculation of proportion of responders.

(2) The company justified the use of the generalised gamma distribution by the lack of clinical plausibility of the alternative parametric time-to-event models (e.g. Gompertz and log-logistic distributions). This argument was not supported by clinical expert opinion, and the ERG considers there to be uncertainty about the likely treatment duration. Within the response-based analysis provided in response to the clarification questions,[7] the company explored the influence of using Gompertz or log-logistic distributions for both responders and non-responders. Both scenario analyses increased the ICERs (Section 5.2.10). However, the company considered that the proportion of patients who were still receiving treatment after five years or more was not representative of clinical practice in both scenarios (Table 5.6).

In conclusion, the ERG adopted a conventional, non-response based approach in the base-case, using the generalised gamma distribution for estimating TTD, in line with the CS. The ERG furthermore explored the influence of using a response-based and landmark analysis for OS, PFS and TTD in a scenario analysis. In this scenario analysis, the generalised gamma was used to estimate TTD of the responders and non-responders, and in a second analysis, the Gompertz and log-logistic distributions were used for responders and non-responders, respectively.

Table 5.6: TTD estimation based on different parametric time-to-event models (landmark and response-based analysis)

Time	TTD estimation				
	Generalised gamma[a]	Generalised gamma[b]	Gompertz	Log-logistic	Best fitting parametric time-to-event models[c]
1 year	17.6%	19.6%	20.1%	20.7%	20.0%
2 year	8.3%	11.4%	13.2%	11.8%	13.2%
3 year	5.1%	8.4%	10.1%	8.0%	10.6%
4 year	3.2%	7.0%	8.4%	5.9%	9.1%
5 year	2.1%	6.1%	7.3%	4.6%	8.2%
10 year	0.2%	4.1%	2.2%	2.1%	3.1%

Source: updated cost effectiveness model submitted with the response to the clarification letter
[a] Used in the company base-case
[b] Estimation based on the landmark and response-based analysis

[c] Based on the landmark and response-based analysis, the log-normal and the Gompertz distributions were the best fitting parametric time-to-event models for the responders and non-responders, respectively.

5.2.7 Adverse events

Table 5.7 presents the adverse events that were included in the cost effectiveness model. Grade 3-4 adverse events were incorporated in the model if their incidence was ≥5%. The impact of adverse events on quality of life and costs were incorporated in the first cycle of the model (see sections 5.2.8 and 5.2.9 for more details).

Table 5.7: Adverse event rates incorporated in the cost effectiveness model

Adverse event	Nivolumab	Docetaxel	Paclitaxel	BSC	Cisplatin plus gemcitabine[a]
Neutropenia	1.00%	14.00%	6.00%	0.90%	66.67%
Anaemia	1.48%	1.00%	0.00%	8.10%	42.42%
Thrombocytopenia	NR	NR	0.00%	0.90%	33.33%
Asthenia	1.48%	6.00%	5.00%	17.90%	0.00%
Nausea/vomiting	0.37%	NR	0.00%	0.90%	0.00%
Diarrhoea	1.85%	0.00%	2.00%	NR	NR
ALT increase	0.74%	0.00%	2.00%	NR	NR
Leukopenia	0.00%	0.00%	0.00%	NR	45.45%
Source	Checkmate 275[10]	Choueiri *et al.* (2012)[27]	Jones *et al.* (2017)[15]	Bellmunt *et al.* (2009)[26];Bellmunt *et al.* (2013)[41]	Gondo *et al.* (2011)[13]

Source: adapted Table 31 of the CS[2]

[a] The trials informing the comparison versus cisplatin plus gemcitabine were conducted in gemcitabine-naïve populations, and therefore cannot be considered to provide relevant data for the retreatment with first-line platinum-based chemotherapy. This comparison has been briefly included in Appendix O as a scenario analysis only and results should be interpreted with caution.

Abbreviations: ALT, alanine transaminase; BSC, best supportive care; NR, not reported

ERG comment: The ERG's concerns relate to (1) the selection of sources for AEs associated with nivolumab, (2) selection of sources for AEs associated with the comparators, (3) the inclusion of both neutropenia and leukopenia, and (4) an inconsistency between the inclusion criteria for AEs and the actually included AEs.

(1) For the nivolumab arm, the CheckMate 275 trial was the only source informing the adverse event rates in the cost effectiveness model while the clinical effectiveness of nivolumab was estimated based on both CheckMate 032 and CheckMate 275 studies. The company justified this choice in its response to the clarification letter by stating that it simplified the analysis and that adverse events did not have a meaningful impact on the results[7]. Hence, the use of both trials instead of CheckMate 275 only for the estimation of adverse event rates would not affect the conclusions of the analysis. The company did not provide evidence to support this argument.

(2) Another issue with the adverse events are that the company did not justify the selection of the source used to estimate AE rates of the comparator. In the response to the clarification letter, the company

explained that these sources were selected to ensure consistency by using the same sources as for the relative effectiveness estimation of nivolumab versus the comparators. The company did not argue why these sources were the most appropriate.

(3) Both neutropenia and leukopenia were incorporated in the cost effectiveness model. The ERG was unsure whether this was appropriate, given that neutropenia is a subtype of leukopenia. However, this is not likely to have a significant impact on model outcomes.

(4) Finally, AEs were included in the cost effectiveness model when their incidence was ≥5%. However, nausea/vomiting, diarrhoea, and ALT increase have an incidence <5% for all treatments included in the cost effectiveness model. Hence it is inconsistent to include these AEs in the cost effectiveness model. The ERG removed these adverse events from its analyses.

5.2.8 Health-related quality of life

Within the economic SLR, six records of four unique studies were identified that included HRQoL in locally advanced or metastatic UC.[34, 36, 42-45] None of these studies were consistent with the NICE reference case and therefore data to inform utilities of the economic evaluation were taken from the CheckMate 275 trial where the EQ-5D-3L was used and valued with UK preference weights.

5.2.8.1 EQ-5D-3L data from CheckMate 275 trial

In absence of alternative data that was consistent with the NICE reference case, the utilities derived from the CheckMate 275 study were deemed most appropriate for this appraisal. Utility estimates derived from the CheckMate 275 study were stratified according to progression-free and post progression health states. Data were available at baseline for 261/270 (96%) patients. During follow-up, the completion-rate declined but remained above 70% at 49 weeks (Table 5.8).

Table 5.8: EQ-5D-3L questionnaire completion rates over time (total enrolled population)

	EQ-5D-3La	
Assessment	**n/N**	**%**
Week 1 (baseline)	261/270	96.7
Week 9	144/167	86.2
Week 17	97/116	83.6
Week 25	75/91	82.4
Week 33	54/70	77.1
Week 41	24/32	75.0
Week 49	6/7	85.7

[a] Completion rates = patients who completed the PRO with ≥1 score at the assessment time point/expected population (total population minus patients who have died or dropped out)
Abbreviations: EQ-5D-3L: EuroQoL 5-Dimensions 3-Levels; PRO: patient reported outcomes.
Source: Table 32 of the CS[2]

In total 794/1,465 (54%) observations were missing. After interpolation of observations made for measurement times deviating from the measurement schedule, 788/1,465 (54%) of observations were available. The remaining missing observations were partly (204/1,465 = 14%) due to the immaturity of the dataset, i.e. patients had not reached all follow-up measurements yet. The company acknowledged that discontinued treatment, progressive status and female gender seemed to be predictors of missing observations, and thus data might not have been missing completely at random. All missing observations were imputed using multiple imputation by chained equations and predictive mean matching, where the number of imputations was set to 40.

The company used a mixed-effects model to reflect within subject variance. This resulted in health state utilities of 0.718 and 0.604 pre-progression and post-progression respectively (Table 5.9). It is noteworthy that imputed pre-progression utilities were similar to observed utilities, but imputed post-progression utilities were lower than the observed utilities. The company furthermore explored the effect of time on progression effect. The pattern seen in post-progression utilities however was deemed different from what was seen in clinical practice, and the company therefore used one set of time-independent utilities.

Table 5.9: Summary of utility values for cost effectiveness analysis

State	Utility/disutility value: mean (standard error)	95% CI	Source
Pre-progression	Imputed value: 0.718 (0.016) Observed value: 0.713 (0.017)	Imputed value: 0.686 to 0.75 Observed value: 0.679 to 0.747	Imputed from Checkmate 275
Change in utility – pre-progression to post-progression	Imputed value: -0.115 (0.0291) Observed value: -0.061 (0.0167)	Imputed value: -0.143 to -0.087 Observed value: -0.123 to -0.055	Imputed from Checkmate 275
Post-progression	Imputed value 0.603 (N/A) Observed value: 0.623 (N/A)	N/A	Checkmate 275

Abbreviations: ALT: alanine transaminase; CI: confidence interval; N/A: not applicable; NR: not reported.
Source: Table 35 of the CS [2]

Adverse event disutilities

The company applied disutilities to several AEs (see Table 5.10); these were based on studies reporting utilities in patients with non-small cell lung cancer, pancreatic cancer and leukaemia. Disutilities were not treatment-specific and were applied as one-off events at the beginning of treatment, based on the proportion of patients experiencing the adverse event and the duration of the adverse event.

Table 5.10: Disutilities used in comparison to previous nivolumab appraisal ID971

Adverse event	Disutility ID995	Source	Disutility ID971	Source
Neutropenia	-0.18	Attard et al. (2014)[46]	-0.09	Nafees (2008)[47]
Anaemia	-0.09	Beusterien et al. (2010)[48]	-0.07	Nafees (2008)[47]
Thrombocytopenia	-0.18	Attard et al. (2014)[46]		
Asthenia/Fatigue	-0.12	Attard et al. (2014)[46]	-0.07	Nafees (2008)[47]
Nausea/vomiting	-0.05	Nafees et al. (2008)[47]	-0.05	Nafees (2008)[47]
Diarrhoea	-0.29	Attard et al. (2014)[46]		
ALT increase	-0.05	NICE TA347 (2015)[49]		
Leukopenia	-0.09	Frederix et al. (2013)[50]		

Sources: Table 35 of the CS [2], previous nivolumab appraisal ID971 [51]

ERG comment: The ERG identified several inconsistencies and choices lacking justification in the handling of utility values. The main issues include (1) inconsistencies in reported observations, (2) the use of utilities derived only from CheckMate 275, (3) the imputation of immature data, (4) the use of multiple imputation instead of the mixed model to adjust for missing data, (5) lack of justification for not using time-dependent utilities, and (6) disutilities for adverse events were inconsistent with those used for a previous nivolumab appraisal.

(1) The ERG noted a small inconsistency in the reported number of observations. As they were reported in the response to the clarification letter, the number of interpolated observations (117), imputed observations (677) and valid observations (661) do not add up to the total of observations (1465), but deviated by 10 observations.[7]

(2) The exclusion of utilities of the CheckMate 032 trial, which was in accordance with the reference case, is inconsistent with the pooling of other outcomes from CheckMate 275 and CheckMate 032 trials. In response to clarification question B16.C, the company reported utilities pooled from both CheckMate 032 and CheckMate 275 trials.[7] In this analysis pre- and post-progression utilities were higher compared to the utilities used by the company, and this resulted in a decrease in the ICERs for all nivolumab comparisons.[7]

(3) The ERG considers the company's decision to impute immature data as unjustified, and is concerned that it works with the unlikely assumption that none of the immature observations will be censored due to death of patients. The ERG wants to stress that the appropriateness of imputation as a substitution for follow-up is highly questionable. The impact of this on utility values is unclear, especially given that the company did not explore the assumptions made and the uncertainty surrounding the immaturely imputed utilities.

(4) The ERG considers the approach to adjust for missing data not sufficiently justified. The company could have used the mixed model, employed to calculate health state utilities, to adjust for missing observations, but instead used multiple imputation. In response to clarification question B16.B, the company presented utilities using only a mixed model.[7] These closely resembled the utilities produced using multiple imputation and led to only a small difference in ICERs. The ERG was satisfied that the use of multiple imputation to adjust for missing data did not have a large impact on model outcomes.

(5) Unfortunately, the company did not respond to the ERG request for an explanation how it was determined that time-dependent utilities were '… *seen to increase and decrease in a manner that would not be expected in clinical practice*' and were not used in the economic evaluation[2] (clarification question B16.G[7]). However, the company additionally added a variable of on- and off-treatment into the mixed model. The utilities presented were thus for four health states: pre- and post-progression, before and after treatment discontinuation, respectively. In this scenario, the disutility of treatment discontinuation was larger than the disutility of progression (Table 5.11), which was in line with the expectation of the ERG. This analysis raises the question whether on- and off-treatment are better predictors of utility values than pre- or post-progression. However, for consistency with other TAs and because progression is commonly accepted to be a predictor for health state values, the ERG maintained the company's pre- and post-progression utility values.

Table 5.11: Final utility values with linear mixed model including treatment discontinuation as a variable

	Pre-progression	Post-progression
On treatment	0.723	0.666
Off treatment	0.650	0.573
Source: Table 29 of the company's response to request for clarification from the ERG [7]		

(6) AE disutilities used were inconsistent with those used in ID971.[51] The disutilities used in the CS stemmed from multinational trials on various cancers, were not evaluated in UK UC patients and are larger than in ID971.[51] Given the prevalence of AEs, it can be expected that the disutilities used favour the cost effectiveness of nivolumab. This is explored in the ERG's sensitivity analysis. It is of note that leukopenia was not associated with a utility decrement or cost. The company did not apply a cost to leukopenia because of the overlap of leukopenia with neutropenia and because no cost was applied in ID971[2]. For consistency, a utility decrement for leukopenia should therefore also not be applied. However, this inconsistency is not influential.

In conclusion, the ERG adopted the pooled utility estimates in its base-case and explored alternative AE disutilities in an exploratory analysis.

5.2.9 Resources and costs

Resource use and unit costs data to inform the economic model were based on a number of sources, including:

- CheckMate 275;
- national databases;
- published sources (both sources identified and not identified in the SLR described in Section 5.1 of this report) and;
- clinical advice.

Additionally, assumptions were necessary in the absence of evidence. These assumptions were validated through discussions with clinicians.

Drug, administration and monitoring costs

The British National Formulary (BNF) was used to obtain unit prices for nivolumab (40mg and 100mg). A PAS, [redacted], was incorporated in the model. The unit prices for docetaxel, paclitaxel and gemcitabine plus cisplatin were taken from the electronic market information tool (EMIT).

The dose/number of vials required per administration were estimated based on the dosage scheme and the dose intensity (reflecting missed doses). For this calculation an average weight of 77.3 kg (SD 16.34) and Body surface area (BSA) of 1.90 m^2 (SD 0.205) were assumed (both based on the CheckMate 275 trial). Using a normal distribution the proportions of patients in different weight and BSA categories were calculated (see CS Tables 36 and 37[2]). Additionally, the calculation of the dose intensity (93.4%) was based on data from the CheckMate 275 and CheckMate 032 trials and based on the assumption that all delayed doses represent missed doses. In absence of evidence, the company assumed that the dose intensity for docetaxel, paclitaxel, gemcitabine plus cisplatin was equal to that of nivolumab.

The average drug costs per patient per four weeks were calculated by combining the drug unit prices, the vials required per administration, the dose intensity and the number of administrations per four weeks.

In addition to the drug costs, administration costs of £198.94 per dose were incorporated (derived from NHS reference costs 2015-16). These costs were incorporated independent of the dose intensity as it was assumed that for missed doses, the chair time would still have been reserved for the patient. The total drug and administration costs per 4 weeks ranged between £304 for docetaxel and [redacted] for nivolumab (see Table 5.12).

Monitoring costs (while on treatment) included in the model (Table 5.13) consisted of regular follow-up visits with an oncologist, CT scans and various blood tests (full blood count, hepatic function test, renal function test, thyroid function test, pituitary function test). The resource use was based on expert opinion (i.e. advisory board feedback) while the unit prices were based on NHS reference costs 2015-16. The total monitoring costs per four weeks ranged between £272 for docetaxel and £556 for gemcitabine plus cisplatin (see Table 5.13).

Table 5.12: Drug and administration costs for nivolumab (with PAS) and comparators

	Per vial		Per dose			Per 4 weeks			
	Vial size (mg)	Costs per vial	Dosage scheme	Dose intensity	Average dose[a]	Number of administrations	Drug costs	Administration costs	Total costs
Nivolumab	40	[redacted]	3 mg/kg	93.4%	260.27	2.00	[redacted]	£397.88	[redacted]
	100	[redacted]							
Docetaxel	80	£12.47	75 mg/m2	93.4%	185.02	1.33	£38.45	£265.25	£303.71
Paclitaxel	100	£8.50	80 mg/m2	93.4%	200.17	3.00	£51.04	£596.82	£647.86
Gemcitabine	1000	£178.56	1000 mg/m2	93.4%	2312.79	3.00	£1,238.92	£596.82	£2,057.66[b]
Cisplatin	50	£6.99	70 mg/m2	93.4%	164.36	1.00	£22.98	£198.94	

[a]This includes wastage (as no vial sharing is assumed) and dose intensity (reflecting missed doses)

[b]Total costs of cisplatin + gemcitabine

Table 5.13: Monitoring costs

	Oncologist follow-up visit per 4 weeks		CT scans per 4 weeks		Various blood tests[a] per 4 weeks		Total per 4 weeks
	Frequency	Costs	Frequency	Costs	Frequency	Costs	Costs
Nivolumab	2.00	£326.00	0.50	£57.50	10.00	£10.00	£393.50
Docetaxel	1.33	£217.33	0.44	£51.11	4.00	£4.00	£272.44
Paclitaxel	3.00	£489.00	0.44	£51.11	9.00	£9.00	£549.11
Gemcitabine plus cisplatin	3.00	£489.00	0.50	£57.50	9.00	£9.00	£555.50

[a]Full blood count, hepatic function test, renal function test, thyroid function test, pituitary function test (all costing £1)

Best supportive care costs

For the BSC comparator, BSC costs were administered until death. For the remaining comparators, BSC costs were incorporated after treatment discontinuation (i.e. discontinuation of nivolumab, docetaxel, paclitaxel or cisplatin + gemcitabine) until death.

BSC costs included GP home visits, community nurse specialist visits and blood transfusions as well as drug costs for prednisolone, morphine, gabapentin and alendronic acid. The total BSC costs per 4 weeks amounted to £170.21 (see CS Table 39[2]).

Adverse event costs

Although not described in the CS, treatment dependent AE costs were incorporated as one-off event costs for patients on treatment during the first cycle of the model based on the occurrence (See Table 5.7) and costs (CS Table 41[2]) of AE. The sum of these costs is provided per treatment in Table 5.14.

Table 5.14: Total AE event costs

	Total AE event costs	Total AE event costs (alternative costs per AE event)[a]	Difference
Nivolumab	£147.24	£115.33	-£31.91
Docetaxel	£773.55	£284.67	-£488.88
Paclitaxel	£408.62	£205.91	-£202.71
Cisplatin + gemcitabine	£5,389.57	£2,477.92	-£2,911.65
BSC	£819.63	£847.52	£27.89

Source: economic model submitted by the company

[a]Alternative costs per AE were retrieved from ID971[51] (nivolumab for treating recurrent or metastatic squamous-cell carcinoma of the head and neck after platinum-based chemotherapy). Moreover, the costs for leukopenia were set to £0 given 1) the overlap with neutropenia; 2) AE occurrence was missing for all comparators except one and; 3) given that this is consistent with ID971 as in this assessment no costs for leukopenia were considered. Finally, the fatigue AE costs from ID971 were assumed to be applicable for asthenia.

Subsequent treatment costs

Following discontinuation of nivolumab, docetaxel, paclitaxel or cisplatin plus gemcitabine, a proportion of patients received subsequent radiotherapy and/or surgery (9.3% and 3.3% respectively based on CheckMate 275). The unit prices were based on NHS reference costs 2015-16 and amounted to £128.22 and £3,201.68 for radiotherapy and surgery respectively. The costs were incorporated as one-off event costs after treatment discontinuation.

Terminal care costs

Terminal care costs were incorporated in the model as event costs of £6,152.64 related to the transition to death. These costs were an average of the acute care and community costs for cancer patients in their last eight weeks of life.[52]

ERG comment: The ERG identified several technical errors, inconsistencies and assumptions that lacked justification. These included a technical error (1) in calculating the dose intensity; inconsistencies, namely (2) using the average weight and BSA from CheckMate 275 (not using CheckMate 032), (3) using the subsequent treatment proportions from CheckMate 275 (not using CheckMate 032), (4) not using cost and resource use data from TA272 (identified in the SLR), and (5) using different AE unit costs compared with ID971[51] (nivolumab for treating recurrent or metastatic squamous-cell carcinoma of the head and neck after platinum-based chemotherapy); as well as three

assumptions, namely (6) assuming an administration scheme that is inconsistent with UK clinical practice for cisplatin + gemcitabine, (7) Assuming that all delayed doses are missed doses for calculating nivolumab dose intensity, and (8) assuming that the dose intensity for the comparators is equal to that of nivolumab.

(1) The identified technical errors entailed the incorporation of dose intensity in the economic model. Drug dose intensity was incorporated in the calculation of the total dose required per weight category, which was subsequently used to calculate the number of vials per weight category. This is incorrect as the dose intensity is related to the number of missed doses and not to the number of vials per weight category. Hence the dose intensity should be applied after calculating the number of vials per weight category. This is corrected in the ERG base-case.

(2) The company assumed a weight and BSA of 77.3 kg and 1.90 m^2 respectively to calculate the dose/number of vials per administration. This was based on CheckMate 275 only. This is inconsistent given that the company combined data from the CheckMate 275 and CheckMate 032 in the majority of their analyses, presumably assuming that the combined population is most relevant for the decision problem being considered. Although the ERG requested clarification on this inconsistency (clarification question B17.A[7]), no further details were provided. Moreover, given that the mean weight was 83.51 kg in CheckMate 032 (mean BSA was not provided in the CSR[11]), this inconsistency resulted in an underestimation of the nivolumab drug costs. Hence, an average weight of 80.405 kg based on both CheckMate 275 and CheckMate 032 was used in the ERG analyses. Given that mean BSA from CheckMate 032 was not provided, the mean BSA of 1.90 m^2 from CheckMate 275 was retained. Moreover, this seems appropriate given that in TA272[34] a similar BSA (of 1.85 m^2) was used (as stated by the company in response to clarification question B17.D[7]).

(3) Similar to the previous inconsistency, the proportions of patients receiving subsequent radiotherapy and/or surgery (9.3% and 3.3% respectively), following discontinuation of nivolumab, docetaxel, paclitaxel or cisplatin + gemcitabine, was retrieved from CheckMate 275 only. These proportions were 11.5% and 6.4% in CheckMate 032. For consistency, average proportions based on both CheckMate 275 and CheckMate 032 were used in the ERG analyses (10.40% and 4.85% for patients receiving subsequent radiotherapy and/or surgery respectively).

(4) The company identified TA272 (the only other NICE submission in this indication) in its SLR. This source was nevertheless not used to inform costs and resource use.[34] The company stated (response to clarification question B18.A[7]) that NHS reference costs for 2007/2008 (from TA272) would be inappropriate to use in 2017. This argument is inconsistent with other costs used by the company (e.g. the leukopenia cost estimate was derived from a paper published in 2004[35]). However, considering the response to clarification question B18[7], it seems reasonable not to use the monitoring and BSC costs from TA272. In response to this clarification question the company states that treatment-related monitoring costs in TA272 did not include oncologist visits and CT scans and were dependent on progression status (instead of treatment status as preferred by the company). Regarding BSC costs, the company stated in TA272 these costs included hospice costs while the company prefers to incorporate these costs as part of the terminal care costs.[34]

(5) The AE unit costs are reported in CS Table 41[2]. These AE unit costs however differ from previous nivolumab assessments (e.g. ID971[51]) and no justification is provided for the sources used to obtain the AE unit costs. This is of particular concern for the AE unit costs for neutropenia and nausea and vomiting as these were based on NHS reference costs for paediatrics. To illustrate the impact of the inconsistency with ID971[51] (nivolumab for recurrent or metastatic head and neck cancer), the ERG calculated alternative AE costs based on ID971[51] (Table 5.14; see footnote for calculation details).

Given the lack of clarity and justification for the AE unit costs reported in CS Table 41, the alternatively calculated AE unit costs, based on ID971, were used in the ERG exploratory analyses.

(6) In the CS it is stated that '*In UK clinical practice, cisplatin plus gemcitabine is given in the first-line setting as gemcitabine (1250mg/m2) plus cisplatin (70mg/m2) on days 1 and 8 of a 21 day cycle (cisplatin on day 1 only)*'.[2] However, in response to clarification question B17.E[7] the company responded that, in the economic model, it assumed the administration regimen with gemcitabine on days 1, 8 and 15 and cisplatin on days 1 and 2. This was based on the administration regimen from the Gondo (2011) study[13] and justified by stating that this study was the key source for efficacy data. The ERG performed scenario analyses incorporating the cisplatin + gemcitabine administration scheme that is likely applicable to UK clinical practice.

(7) In response to clarification question B17.B[7] the company stated that dose delays that exceed the duration of a nivolumab treatment cycle (i.e. 14 days) can reasonably be assumed to be missed. Hence, the company assumed that all delayed doses were missed doses. This seems reasonable to the ERG if all dose delays exceed the duration of a nivolumab treatment cycle. However, it is highly questionable whether this is applicable to all dose delays. Particularly given that the length of dose delays was less than one week in 34.6% and 38.5% of all delayed doses for CheckMate 275 and CheckMate 032 and the large majority of dose delays (71.7% and 80.8% respectively) does not exceed the duration of a nivolumab treatment cycle[10, 11]. Therefore, in the ERG base-case a missed dose was only assumed in case the length exceeded seven days; resulting in a proportion of unadministered drug doses of 6.6% (CS dose intensity) × 36.6% (the proportion of dose delays that exceeded 14 days; averaged for CheckMate 275 and CheckMate 032) = 2.4% (i.e. dose intensity of 97.6%).

(8) The calculated dose intensity of 93.4% for nivolumab was assumed to be applicable for the comparators; assuming that 6.6% of the doses would be missed. In response to clarification question B17.C[7], the company stated that this was assumed in absence of evidence. In addition, the company stated that assuming no dose intensity for the comparators would induce bias in favour of nivolumab.[7] However, the ERG questions whether the current approach (assuming a dose intensity of 93.4% for all comparators) does not induce bias in favour of nivolumab as well. Particularly considering the AE occurrence that was used for the comparators (Table 5.7), it is not unlikely that that the number of missed doses is higher for (some of) the comparators than for nivolumab. Hence the drug costs for the comparators might be overestimated.

5.2.10 Cost effectiveness results

In the deterministic base-case analysis, nivolumab was associated with larger QALY and LY gains than docetaxel, paclitaxel and BSC (Table 5.15). The main benefit of nivolumab versus these comparators stemmed from QALY gains post-progression ([redacted], [redacted] and [redacted] of incremental QALYs in post-progression health state for the comparisons with docetaxel, paclitaxel and BSC respectively). Compared with cisplatin plus gemcitabine, nivolumab's incremental QALYs were increased in pre-progression and decreased in post-progression.

Nivolumab also induced larger life time costs than docetaxel, paclitaxel and BSC. Incremental costs mainly stemmed from higher treatment costs ([redacted]), which reflect the technology costs of nivolumab, and to a minor degree stemmed from higher costs in the post-progression health state ([redacted]) (Table 5.16). With the PAS, nivolumab treatment resulted in incremental cost effectiveness ratios (ICERs) of £37,646, £44,960 and £38,164 per QALY gained versus docetaxel, paclitaxel and BSC respectively (Table 5.17).

Table 5.15: Summary of quality-adjusted life year gains by health state

	Nivolumab		Docetaxel			Paclitaxel			Cis+ gem			BSC		
	QALYs	LYG	QALYs	LYG	Incremental QALYs vs. Nivolumab	QALYs	LYG	Incremental QALYs vs. Nivolumab	QALYs	LYG	Incremental QALYs vs. Nivolumab	QALYs	LYG	Incremental QALYs vs. Nivolumab
Health state														
Pre-progression	[redacted]	1.06	[redacted]	0.75	[redacted]	[redacted]	0.47	[redacted]	[redacted]	0.47	[redacted]	[redacted]	0.32	[redacted]
Post-progression	[redacted]	1.72	[redacted]	0.65	[redacted]	[redacted]	0.71	[redacted]	[redacted]	1.99	[redacted]	[redacted]	0.70	[redacted]
Adverse events	[redacted]		[redacted]		[redacted]	[redacted]		[redacted]	[redacted]		[redacted]	[redacted]		[redacted]
Total	[redacted]	2.78	[redacted]	1.40	[redacted]	[redacted]	1.19	[redacted]	[redacted]	2.47	[redacted]	[redacted]	1.01	[redacted]

Abbreviations: QALY: quality-adjusted life year; LYG: life years gained; Cis+gem: cisplatin plus gemcitabine; BSC: best supportive care.
Source: Table 67 of the CS Appendix J [20]

Table 5.16: Summary of costs by health state

	Nivolumab	Docetaxel		Paclitaxel		Cis+ gem		BSC	
	Costs	Costs	Incremental costs vs. Nivolumab	Costs	Incremental costs vs. Nivolumab	Costs	Incremental costs vs. Nivolumab	Costs	Incremental costs vs. Nivolumab
Treatment	[redacted]	£3,113	[redacted]	£3,515	[redacted]	£12,381	[redacted]	£2,310	[redacted]
Monitoring	[redacted]	£2,716	[redacted]	£2,734	[redacted]	£3,455	[redacted]	£0	[redacted]
Post-progression	[redacted]	£1,521	[redacted]	£1,864	[redacted]	£4,492	[redacted]	£0	[redacted]
Adverse events	[redacted]	£739	[redacted]	£411	[redacted]	£5,378	[redacted]	£806	[redacted]
Terminal care	[redacted]	£5,857	[redacted]	£5,902	[redacted]	£5,630	[redacted]	£5,940	[redacted]
Total	[redacted]	£13,945	[redacted]	£14,426	[redacted]	£31,337	[redacted]	£9,056	[redacted]

Abbreviations: Cis+gem: cisplatin plus gemcitabine; BSC: best supportive care
Source: Table 68 of the CS Appendix J [20]

Table 5.17: Base-case results – with PAS

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER of nivolumab vs each comparator (£/QALY)
Nivolumab	[redacted]	2.78	[redacted]				
Paclitaxel	£14,426	1.19	0.76	[redacted]	1.60	[redacted]	£37,647
Docetaxel	£13,945	1.40	0.92	[redacted]	1.38	[redacted]	£44,960
BSC	£9,056	1.01	0.64	[redacted]	1.77	[redacted]	£38,164
Cis+gem	£31,337	2.47	1.49	[redacted]	0.31	[redacted]	£71,608

Abbreviations: BSC: best supportive care; Cis+gem: cisplatin plus gemcitabine; ICER: incremental cost-effectiveness ratio; LYG: life years gained; QALYs: quality-adjusted life years.
Source: Table 44 of the CS [2]

ERG comment: The ERG comments relate to (1) the exclusion of cisplatin plus gemcitabine from the base-case, and (2) the driving factor of incremental QALYs being the extended post-progression survival.

(1) Cost effectiveness results were not presented for nivolumab compared with cisplatin plus gemcitabine within the company's base-case. This is not in line with the scope. The ERG requested this analysis in the clarification letter but the company continued to exclude this analysis from the base-case, arguing in their response to question B13.A[7], that '*… it is not considered a relevant comparator in the context of second-line UK clinical practice*'[7]. The ERG disagrees with this statement, especially given that this comparator was named in the scope. More detail on this is presented in Section 4.

(2) In a previous nivolumab appraisal ID971,[51] it has been discussed that incremental QALYs were mainly driven by extended survival post-progression and after treatment discontinuation. Such a pronounced effect of nivolumab after progression or treatment discontinuation had not been seen in clinical practice, [51] thus the extrapolation in the model has been criticised in previous committee appraisals. The ERG wishes to flag up that in the company's base-case the issue of the QALY gain coming almost entirely from the post-progression health state was less pronounced but still accounted for over 50% of incremental gains for all comparators in the company's base-case.

5.2.11 Sensitivity and scenario analyses

Probabilistic sensitivity analysis (PSA) and deterministic sensitivity analysis (DSA) were undertaken and presented by the company. Patient age, weight and BSA, costs, resource use, utilities, TTD, PFS and OS were varied (further information in Table 46 of the CS[2]).

Results of the PSA using 1,000 iterations are shown in Table 5.18. Incremental costs increased and incremental QALYs decreased compared to the deterministic results, resulting in ICERs of £46,209 and £44,698 per QALY gained for nivolumab versus paclitaxel and BSC, and an ICER of £54,220 per QALY gained for nivolumab versus docetaxel. The company reasoned that the PSA ICER increases were mainly driven by a reduction in PFS and OS in the PSA (compared with the deterministic analysis). As PFS and OS are greater in nivolumab than in the comparators, the effect on nivolumab was more pronounced than on the comparators. Probability of cost effectiveness at a threshold of £50,000 per QALY gained was 72.1% versus paclitaxel, 49.0% versus docetaxel, 76.3% versus BSC and 6.9% versus gemcitabine plus cisplatin.

Table 5.18: Probabilistic CS results

Technologies	Incremental costs (£)	Incremental QALYs	ICER (£/QALY)	Probability of cost effectiveness[a]
Paclitaxel	[redacted]	[redacted]	£46,209	72.10%
Docetaxel	[redacted]	[redacted]	£54,220	49.00%
BSC	[redacted]	[redacted]	£44,698	76.30%
Cis+gem	[redacted]	[redacted]	£103,568	6.9%

[a]The probability of nivolumab being cost-effective versus the stated comparator at a cost-effectiveness threshold of £50,000/QALY.
Abbreviations: Cis+gem: cisplatin plus gemcitabine; BSC: best supportive care, ICER: incremental cost-effectiveness ratio; QALYs: quality-adjusted life years.
Sources: Table 47 of the CS [2], Table 79 of the CS Appendix O [20]

The company stated that individual one-way DSAs were conducted including all parameters other than survival curves. The parameters were varied within their respective 95% CI or, if not applicable, within a ± 50% range of the deterministic base-case value. The DSA results including the PAS were presented using tornado diagrams with the 10 key model drivers (CS Figures 46-48 [2]). Ranked by importance, the following parameters were identified as most influential on the cost effectiveness of nivolumab versus paclitaxel, docetaxel and BSC:

1. Mean age (65; 47-84)
2. Cost per 100mg Nivolumab (£1,097; £548.50-£1,645.50)
3. Mean weight (77,3; 45-100)
4. Nivolumab dose intensity (93%; 47%-100%)

The company performed six deterministic scenario analyses, which are presented in Table 5.19. In summary, the scenario analyses indicated that the choice of nivolumab parametric OS, PFS and TTD curves, the position of the landmark, as well as the choice of the fractional polynomial model were major drivers of model results, mostly resulting in less favourable cost effectiveness estimates for nivolumab versus its comparators (see Table 5.19).

Table 5.19: Deterministic scenario analyses

Scenario			**ICER vs. Paclitaxel**	**ICER vs. Docetaxel**	**ICER vs. BSC**
Base case		**Gen. gamma**	**£37,647**	**£44,960**	**£38,164**
1 Survival curves	Landmark week 8	Weibull	£101,994	£114,823	£91,372
		Gompertz	£49,010	£59,858	£50,201
		Lognormal	£52,900	£72,044	£53,634
		Log-logistic	£58,279	£78,063	£59,695
		Exponential	£57,998	£70,582	£59,564
	Landmark week 26	Gen. Gamma	£34,541	£40,246	£34,774
		Weibull	£50,060	£62,866	£51,378
		Gompertz	£35,655	£41,933	£35,269
		Lognormal	£38,834	£48,610	£38,192
		Log-logistic	£42,475	£54,235	£43,097
		Exponential	£60,279	£76,786	£61,389
2 Fractional polynomial model[a]		p1=1, p2=1	£56,073	£59,504	£43,554

Scenario		ICER vs. Paclitaxel	ICER vs. Docetaxel	ICER vs. BSC
3 Exponential piecewise model	Piecewise exponential at 8 weeks	£53,616	£65,450	£55,597
	Piecewise exponential at 26 weeks	£55,681	£71,147	£57,293
4 Vial sharing		£35,651	£42,630	£36,333
5 Stopping rule[b]		£31,561	£37,781	£32,743
6 Alternative TTD parametric curves	Weibull	£33,562	£40,141	£34,525
	Gompertz	£183,467	£216,984	£168,053
	Lognormal	£61,810	£73,465	£59,688
	Log-logistic	£61,994	£73,683	£59,851
	Exponential	£28,331	£33,971	£29,866

[a] Second-best fitted fractional polynomial model
[b] Stopping rule applied where are the end of 2 years treatment, 75% of patients still receiving treatment will discontinue treatment
Sources: Tables 48 – 54 [2]

ERG comment: The ERG identified several inconsistencies and limitations regarding the DSA and PSA presented by the company. These relate to (1) the exclusion of parameters from the DSA, (2) the exclusion of parameters from the PSA, (3) the number of iterations used in the PSA, along with (4) the unexplained differences between deterministic and probabilistic results, and (5) the absence of cisplatin plus gemcitabine from the fully incremental PSA.

(1) In the DSA, the contribution of survival curves were not explored and even though stated by the company, HRs were not varied either. The ERG concludes the DSA does not accurately reflect uncertainty of the cost effectiveness of nivolumab versus the comparators.

(2) The PSA excluded HRs and Kaplan-Meier estimates used to estimate nivolumab survival before the landmark, and erroneously included patient characteristics. In response to the clarification questions, the company included Kaplan-Meier curves in the PSA, but stated that it did not include hazard ratios because '*inclusion of hazard ratios would generate illogical results due to the time-varying nature of the hazard ratios […]*' resulting in '*changes in PFS and OS that are not clinical plausible*'[7]. This was not further elaborated on and methods to correct for this were not explored. The ERG agrees that varying the HR in each time period could result in counterintuitive results but the ERG also thinks that this could have been corrected for, for example, by using a fixed set of random numbers. The company furthermore stated that the comparators' OS was accounted for via the OS estimates of nivolumab. However, it is the relative effectiveness that has the greatest effect on the model and on uncertainty and the ERG therefore does not consider this to be a valid argument and concludes that the PSA does not fulfil the NICE reference case and does not reflect a significant part of the uncertainty. The ERG therefore chose not to present the CEACs.

(3) The PSA presented by the company used 1,000 iterations, a number criticised as too small by the ERG. In response to the clarification letter, the company increased the number of iterations to 10,000, which is considered to be more appropriate. However, the ERG tested the use of 20,000 in its base-case and still noted discrepancies in incremental costs and QALYs between two runs (not in excess of £100 in costs and third decimal place utility values), thus indicating that a large number of PSA iterations is required to achieve stable results.

(4) Unfortunately, the company did not provide further information in response to the ERG clarification question on why nivolumab OS and PFS in the PSA might be lower compared to the deterministic analysis. The discrepancy between probabilistic and deterministic results persisted in the ERG's response-based analysis, with probabilistic results being the more conservative. However, the ERG noticed that using conventional, not response-based, survival analysis resulted in probabilistic model outcomes that reflected much more closely the deterministic results. The large discrepancy between probabilistic and deterministic results is likely the result of a combination of the increased uncertainty associated with the response-based approach (which in turn is caused by fitting parametric models to smaller sample sizes based on responder and non-responder groups and only using data after the landmark), the skew of the used distributions and the quantitative difference in survival between the response-based and conventional approaches (response-based approach yields an average of 2.45 and 2.8 probabilistic and deterministic nivolumab life years respectively and the conventional approach an average of 1.82 and 1.84 probabilistic and deterministic nivolumab life years respectively).

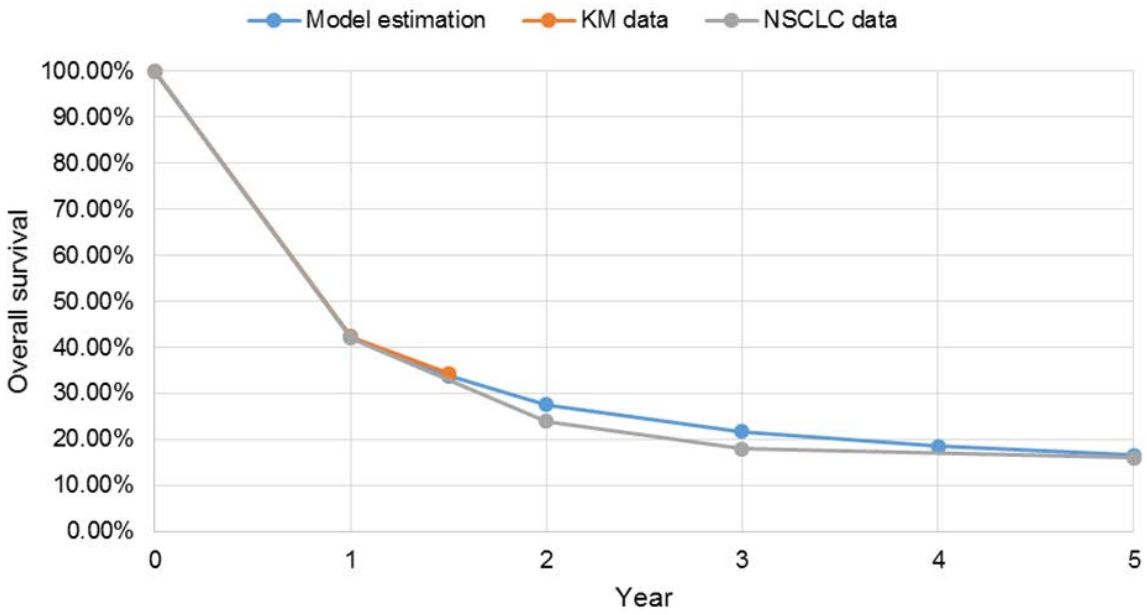
(5) In response to clarification question B13.A[7], the company provided a model that allowed for a simultaneous comparison of nivolumab to docetaxel, paclitaxel and BSC in fully incremental analysis. Despite the ERG's request to include the comparator cisplatin plus gemcitabine in the base-case, cisplatin plus gemcitabine remained excluded from the incremental PSA.

In conclusion, the ERG extended the incremental PSA to contain 10,000 iterations and to include cisplatin plus gemcitabine as a comparator.

5.2.12 Model validation and face validity check

The company undertook efforts to validate their cost effectiveness estimates for both nivolumab and comparators. The predictions of the model regarding OS and PFS were compared against expert feedback and other long-term nivolumab data in NSCLC and other solid tumours, using five-years follow up data from the CheckMate 003 study.[53] Clinical experts stated that lung cancer would be the most similar to bladder cancer, in relation to the strong link to smoking, the choice of treatment used in clinical practice, and the poor outcomes associated with both diseases without treatment. A comparison between the prediction of the generalised gamma and the CheckMate 003 data is shown in Figure 5.12.

Figure 5.12: Validation of model predictions of OS with nivolumab



Source: CS Figure 49

Validation of comparator estimates also involved comparison against expert opinion and the KM estimates derived from available clinical data (see Table 5.20). Two clinical experts stated that they would not expect more than 5% of patients to be alive at two years, when treated with the comparators. This feedback was deemed to be most closely aligned with outcomes for paclitaxel, informed by the UK PLUTO trial (see Table 5.20).[15] The company states that, because of this expert opinion, it might be that overall survival may be slightly over-estimated in the model.

Table 5.20: Comparison of overall survival extrapolation in model against observed data

Data source	Survival curve	Proportion alive, %					
		1 year	1.5 years	2 years	3 years	4 years	5 years
Nivolumab							
Model estimates for OS	Gen. Gamma (Base case)	42.34%	33.82%	27.54%	21.66%	18.51%	16.55%
CheckMate 275	Kaplan-Meier data	XXX	XXX	-	-	-	-
CheckMate 003 (NSCLC)	-	42%	-	24%	18%	-	16%
Docetaxel							
Model estimates for OS	Gen. Gamma (Base case)	25.01%	15.67%	11.05%	7.67%	6.36%	5.69%
Choueiri *et al.* (2012)[30]	Kaplan-Meier data	24.33%	13.03%	-	-	-	-

Sideris *et al.* (2016)[54]	Kaplan-Meier data (Bytescout)	19%	8%	6%	-	-	-
Paclitaxel							
Model estimates for OS	Gen. Gamma (Base case)	31.41%	17.40%	10.56%	5.66%	3.94%	3.15%
Jones *et al.* (2017)[31]	Kaplan-Meier data	31.58%	15.08%				
Sideris *et al.* (2016)[54]	Kaplan-Meier data (Bytescout)	19%	8%	6%	-	-	-
BSC							
Model estimates for OS	Gen. Gamma (Base case)	14.00%	8.96%	6.64%	5.03%	4.42%	4.09%
Bellmunt *et al.* (2013)[55]	Kaplan-Meier data	21.30%	10.65%	7.41%	1.39%	-	-
Source: CS table 55 Abbreviations: BSC: best supportive care; NSCLC: non-small cell lung cancer; OS: overall survival.							

ERG comment: The ERG's concerns include (1) the lack of internal and cross validity efforts as well as sparse use of expert opinion, (2) external validation efforts that are based on a lung cancer study, (3) the use of only CheckMate 275 for validating model predictions, as well as (4) transparency issues with the model.

(1) The company focused on external validation only. There is no description of face validity checks or cross validity checks (for instance, model outcomes could have been compared with those from TA 272[34]). It is also noteworthy that clinical experts were only consulted prior to model development at an advisory board. Clinical experts therefore did not provide feedback on the distributions used for estimating OS and PFS in the company's base-case response-based approach.

(2) The CS cites clinical experts as stating that bladder cancer is most similar to lung cancer. However, the ERG considers it questionable whether lung cancer really is similar enough to bladder cancer to enable data from the CheckMate 003 trial to be used for external validation of model predictions in bladder cancer. The cited study also was not identified through a SLR. This is of even more concern given that there are significant molecular differences in the two diseases.[5] The comparison does show a slight over-estimation of longer-term OS using the company's base-case model predictions when compared with longer-term OS data from the NSCLC study.[53]

(3) In the comparison of model predictions for OS in nivolumab patients, the company only provides data of CheckMate 275, and not the pooled estimates from CheckMate 275 and 032. This discrepancy impairs the credibility of this validation effort.

(4) The ERG wishes to highlight a few transparency issues with the submitted model file. Hidden columns on several sheets, the practice of not naming cells, the practice of disabling headings for columns and rows and the missing macro for generating the CEAC caused the ERG unnecessary difficulties in validating and amending the model.

5.3 Exploratory and sensitivity analyses undertaken by the ERG

Table 5.20 summarises all main issues highlighted by the ERG in Section 5.2, indicates the expected direction of bias introduced by these issues and whether these are examined in any analyses/incorporated in the ERG base-case.

Table 5.21: Main ERG critique of company's submitted economic evaluation

Issue	Bias introduced[a]	ERG analyses	Addressed in company analysis?
Model structure (section 5.2.2)			
• Combination of responder and non-responder groups instead of creating separate health states	+/-	NA	Requested, not addressed
Interventions and comparators (section 5.2.4)		ERG base-case (FV)	Requested, not addressed
• Exclusion of cisplatin plus gemcitabine	+		
Treatment effectiveness and extrapolation (section 5.2.6)			
• Pooling of CheckMate studies	+/-	NA	Requested, not addressed
• Response-based analysis and the use of landmark analysis, with the main issues including:	+	ERG base-case (FV), and exploratory analysis: Scenario analysis	Partly addressed, listed for each :
o the choice of landmark of 8 weeks	+/-		Requested, partly addressed
o the use of KM estimates up to the landmark	+/-		Not addressed
o rejection of proportional hazards between responders and non-responders	+/-	NA	Not addressed
o simultaneous choice of parametric time-to-event models for OS, PFS and TTD	+/-	NA	Company enabled differential selection
o a posteriori combination of responder and non-responder groups	+/-	Scenario analysis	Requested, not addressed
o application of HRs (which are derived from all patients) on the a posteriori group	+/-	NA	Not addressed
• Background mortality: error in use of UK life tables and converting rate to probability	+/-	NA	NA
• Effectiveness data derived from single-arm studies using a simulated treatment comparison results in large uncertainty and potential bias that was not quantified	+/-	ERG base-case (FE)	Not addressed
• Use of time-varying HRs	+	Exploratory analysis	Company provided time-fixed HRs, but ERG's own estimates differed
• Estimation of HRs for PFS of BSC and cisplatin plus gemcitabine based on assumptions around similarity in comparative effectiveness to vinflunine and paclitaxel	+/-	Exploratory analysis	Not addressed
• Inconsistency in that TTD analysis was not response-based, when OS and PFS were	+/-	Exploratory analysis	Company provided response-based TTD
• Proportion of responders for TTD analysis based on a sum of PFS and OS patients	-	ERG base-case (FV)	NA
Adverse events (sections 5.2.7-5.2.9)			
• Use of only CheckMate 275	+/-	NA	Requested, not addressed

Issue	Bias introduced[a]	ERG analyses	Addressed in company analysis?
• Choice of source for AE rates used for comparators not justified	+/-	NA	Not addressed
• Inclusion of AEs with incidence of <5% not in line with inclusion criteria	+/-	ERG base-case (FV)	Not addressed
Health-related quality of life (section 5.2.8)			
• Utilities only derived from CheckMate 275	-	ERG base-case (FV)	Company provided pooled utilities
• AE disutilities inconsistent with those used in ID971	+	Exploratory analysis	Not addressed
Resources and costs (section 5.2.9)			
• Technical error incorporating dose intensity	-	ERG base-case (FE)	NA
• Inconsistency in estimating weight and subsequent treatment proportions, based on CheckMate 275 only	+	ERG base-case (FV)	Not addressed
• AE unit costs inconsistent with ID971	+	Exploratory analysis	Not addressed
• Cisplatin plus gemcitabine administration scheme not reflective of UK practice	+	Exploratory analysis	Not addressed
• Assumption that all delayed doses were missed doses	+	ERG base-case (MJ)	Not addressed
• Assumption that dose intensity for the comparators is equal to that of nivolumab	+	NA	NA
Cost-effectiveness analyses (sections 5.2.10 and 5.2.11)			
• Relative effectiveness not considered in the PSA	+/-	NA	Requested, not addressed
• Patient characteristics included in PSA	+/-	ERG base-case (FV)	Not addressed
• OS and PFS under-estimated in PSA compared to deterministic analysis	+/-	NA	Not addressed
Validation (section 5.2.12)			
• Insufficient validation of the model	+/-	NA	Not addressed

Abbreviations: NA, not applicable; FE, fixing error; FV, fixing violations; MJ, matters of judgement

[a]Likely conservative assumptions (of the intervention versus all comparators) are indicated by '-'; while '+/-' indicates that the bias introduced by the issue is unclear to the ERG and '+' in indicates that the ERG believes this issue likely induces bias in favour of the intervention versus at least one comparator.

Based on all considerations from Section 5.2 (summarised in Table 5.21), the ERG defined a new base-case. This base-case included multiple adjustments to the original base-case presented in the previous sections. These adjustments made by the ERG form the ERG base-case and were subdivided into three categories (derived from Kaltenthaler 2016):[56]

- Fixing errors (correcting the model where the company's submitted model was unequivocally wrong)
- Fixing violations (correcting the model where the ERG considered that the NICE reference case, scope or best practice had not been adhered to)
- Matters of judgement (amending the model were the ERG considers that reasonable alternative assumptions are preferred)

Additionally, exploratory sensitivity analyses were performed by the ERG to examine the potential impact of alternative assumptions on the cost effectiveness estimates.

The ERG's base-case:

Fixing errors

1. Error in the use of UK life tables and conversion of background mortality rate to probability

   The ERG corrected the error.
2. Error in calculating dose intensity

   The ERG corrected the error by applying dose intensity after calculating the number of vials per weight category, instead of before.

Fixing violations

3. Exclusion of cisplatin plus gemcitabine from base-case and fully incremental analysis in PSA.

   The ERG added cisplatin plus gemcitabine to the base-case and fully incremental analysis in the PSA.
4. Calculation of responder and non-responder proportions for response-based TTD analysis based on OS and PFS, thereby double-counting patients.

   The ERG used only OS to calculate the responder and non-responder proportions used for response-based TTD analysis.
5. Adverse events with an incidence <5% were included in the model, despite the company stating that these should be excluded.

   The ERG removed adverse events with an incidence <5% from the analysis.
6. Use of utilities from CheckMate 275 only.

   The ERG employed the pooled utility estimates from both CheckMate 275 and 032 studies.
7. Use of BSA and weight from CheckMate 275 only.

   The ERG employed the pooled weight from CheckMate 275 and 032, but, due to BSA data not being available from CheckMate 032, kept the BSA estimate from CheckMate 275 only. It should be noted that the re-calculation of weight categories was based on the pooled mean only, the standard deviation was unchanged.
8. Inappropriate parameters in PSA: Patient characteristics were included in the PSA, although they are considered first order uncertainty and typically not reflected in cohort model PSAs. Comparator treatment costs were included in the PSA, but are not typically included.

   The ERG removed patient characteristics and comparator treatment costs from the PSA.

Matters of judgment

9. Use of response-based analysis, without sufficient justification and despite it introducing additional uncertainty.

   The ERG used a not response-based, conventional, survival analysis in its base-case, making redundant the choice of a landmark and retaining the same parametric time-to-event models as chosen by the company (goodness-of fit suggests it is second for OS and first or second for PFS).
10. The assumption that all delayed doses are missed doses.

    The ERG assumed only doses delayed by 7 days or more to be missed doses.

5.3.1 Probabilistic ERG base-case

The ERG performed a PSA to obtain the ERG base-case incorporating all abovementioned adjustments. This resulted in ICERs (probabilistic) of £87,709, £68,519 and £69,515 for nivolumab (with PAS) versus docetaxel, paclitaxel and BSC, respectively (Table 5.22). Cisplatin plus gemcitabine dominated

nivolumab. The individual effects of each change on costs, QALYs and ICERs are presented in Section 6, Table 6.1. For comparison, the deterministic ERG base-case ICERs were £83,397, £65,411 and £67,175 per QALY gained, with cisplatin plus gemcitabine dominating nivolumab.

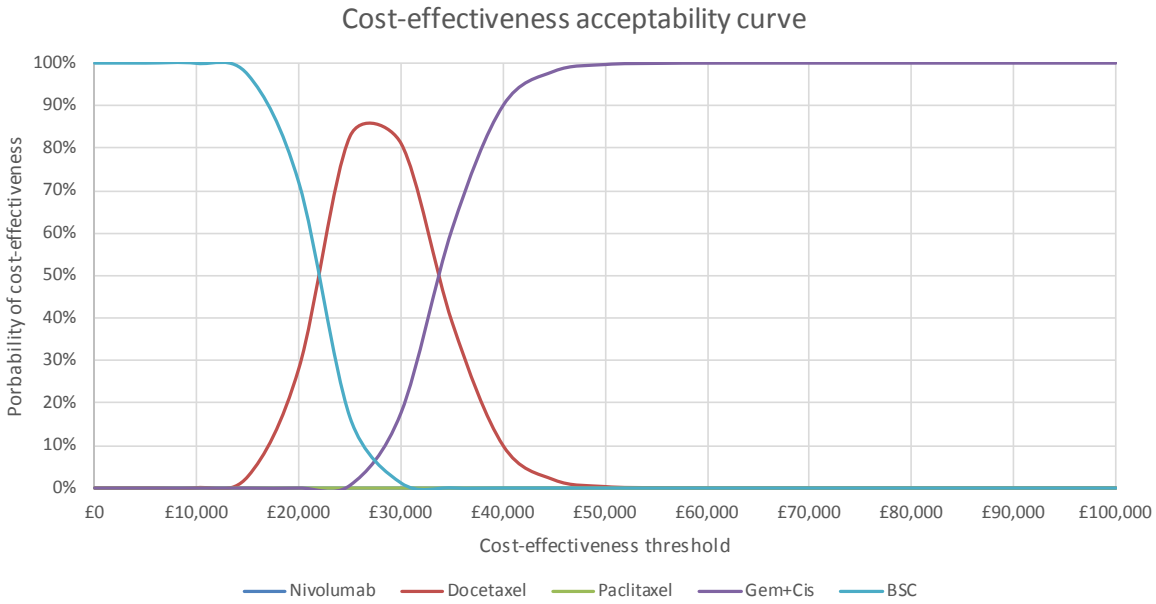
Table 5.22: ERG base-case (probabilistic)

	Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	Nivolumab ICER (£/QALY)
ERG base-case	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£12,493	0.74	[redacted]	[redacted]	£87,709
	Paclitaxel	£13,866	0.63	[redacted]	[redacted]	£68,519
	Cis + gem	£29,384	1.24	[redacted]	[redacted]	Nivolumab is dominated
	BSC	£8,696	0.56	[redacted]	[redacted]	£69,515

ERG = Evidence Review Group; ICER = incremental cost effectiveness ratio; QALY = quality-adjusted life year

The CEACs based on the ERG base-case (Figure 5.13) show that nivolumab has a probability of being cost effective of 0% and 0% at thresholds of £30,000 and £50,000 per QALY gained, respectively.

Figure 5.13: Cost effectiveness acceptability curve for ERG base-case



The ERG wishes to reiterate that the probabilistic model results are different from the deterministic results. This difference was more pronounced using the company's base-case (with fixed errors) than when using the ERG base-case. The difference is explained by using the response-based approach. However, it is not clear what in the response-based approach causes the probabilistic results to deviate as much from the deterministic results. The ERG considers it to be related to a) the increased uncertainty introduced by the response-based approach, b) the skew of the parametric models used and c) potentially the significant quantitative difference in OS and PFS caused by the response-based compared to the conventional approach.

5.3.2 Additional exploratory analyses performed based on the ERG base-case

Additional sensitivity analyses were performed to examine the potential impact of the following alternative assumptions on the cost effectiveness estimates. These included two scenario analyses: a) exploratory analyses performed using the ERG base-case, and b) exploratory analyses performed using the ERG base-case, except that a response-based approach was used (ERG base-case apart from 9.). Results are presented in Tables 6.2 in Section 6.

a) Exploratory analyses using the ERG base-case:

1. Alternative parametric time-to-event models: use of the lognormal distribution for OS (best-fitting according to AIC/BIC) and log-logistic for PFS (best fitting according to BIC, second-best according to AIC).
2. Use of alternative specifications for the fractional polynomial model, by employing a 'mini-PSA' across the different p1 and p2 values provided by the company in response to clarification questions. Results are presented as credible intervals about incremental costs and QALYs and the resulting range of ICERs in Table 6.3 in Section 6.
3. Use of naïve comparison performed by the ERG, instead of the STC, to derive HRs for OS and PFS. The ERG noticed that the code supplied to estimate the time-dependent HRs only estimated them up to a time horizon of 256 weeks, ending much before the end of the model time horizon. It is not clear where the time-dependent HRs implemented after 260 weeks were sourced from. The ERG used the company's time-dependent HRs after 260 weeks, which should not be influential and work in favour of nivolumab.
4. Use of time-independent HRs for OS and PFS derived by the ERG instead of time-dependent HRs.
5. Use of HRs for OS as proxy for HR for PFS for the comparisons with BSC and cisplatin plus gemcitabine.
6. Use of adverse event disutilities and resource use from technology appraisal ID971.
7. Use of the UK dosage schedule for cisplatin plus gemcitabine.
8. An extreme scenario of assuming no treatment effect of nivolumab vs comparators.

b) Exploratory analyses on the ERG base-case using response-based analysis for OS, PFS and TTD:

1. Maintaining the company's base-case choice of parametric time-to-event models, i.e. the generalised gamma for responders' and non-responders' OS, PFS and TTD.
2. Use of parametric time-to-event models with the best fit for OS and PFS (based on AIC/BIC) for responder OS and PFS (generalised gamma), non-responder OS and PFS (Weibull), but maintaining responder and non-responder TTD as the generalised gamma.
3. Use of parametric time-to-event models with the best fit (based on AIC/BIC) for responder OS and PFS (generalised gamma), non-responder OS and PFS (Weibull), responder TTD (lognormal) and non-responder TTD (Gompertz).
4. Use of 26-week landmark instead of 8-week landmark

5.3.3 Subgroup analyses performed based on the ERG base-case

No subgroup analyses were performed.

5.4 Conclusions of the cost effectiveness section

The majority of the cost effectiveness searches in the CS were well documented and easily reproducible, and were carried out in line with the NICE guide to the methods of technology appraisal Sections 5.2.2 and 5.2.4.[33]

Reviewing the overall evidence, the ERG confirmed that there was no existing cost effectiveness model for nivolumab for the current indication, and thus that development of a de novo model was necessary. The economic model described in the CS is considered by the ERG to meet the NICE reference case, with the notable exceptions of a) the exclusion of a comparator that was identified in the scope, and b) a PSA that excludes crucial parameters, includes parameters usually not included in the PSA (such as patient characteristics), and yields results significantly different from the deterministic results. The company model follows a logical structure with respect to the nature of the disease. The economic model was primarily informed by the CheckMate 275 and CheckMate 032 studies, both single-arm studies. Relative treatment effectiveness were informed based on a simulated treatment comparison using studies that were identified through the systematic literature review on the comparators docetaxel, paclitaxel, cisplatin plus gemcitabine and BSC.

The company base-case ICERs (probabilistic) of nivolumab (with PAS) compared with docetaxel, paclitaxel, cisplatin plus gemcitabine and BSC were £54,220, £46,209, £103,568 and £44,698 per QALY gained respectively. The cost effectiveness results were not robust to scenario and one-way sensitivity analyses conducted by the company. Scenario analyses indicated that the choice of nivolumab parametric OS, PFS and TTD curves, the position of the landmark, as well as the choice of the fractional polynomial model used for the NMA were major drivers of model results, mostly resulting in less favourable cost effectiveness estimates for nivolumab versus its comparators.

The ERG incorporated various adjustments to the company's base-case. The ERG base-case resulted in ICERs (probabilistic) of £87,709, £68,519 and £69,515 per QALY gained for nivolumab (with PAS) versus docetaxel, paclitaxel and BSC respectively. In the ERG base-case, cisplatin plus gemcitabine dominated nivolumab, with a larger QALY gain and lower costs. For comparison, the deterministic ERG base-case ICERs were £83,397, £65,411 and £67,175 per QALY gained, with cisplatin plus gemcitabine dominating nivolumab. The single most influential adjustment made by the ERG in its base-case was the use of conventional survival analysis instead of adopting the company's preferred response-based approach.

The ERG identified substantial issues and uncertainties that affected the cost effectiveness analysis. The main issues with the analysis include the use of a response-based survival analysis approach, which was not appropriately and sufficiently justified, necessitated a number of additional assumptions and therefore caused additional uncertainty. These additional assumptions included the choice of a landmark; the use of KM estimates up to the chosen landmark; assumptions surrounding the proportionality of hazards between responders and non-responders; increased uncertainty surrounding the choice of parametric time-to-event models for OS, PFS and TTD; the a posteriori combination of responder and non-responder groups; and the application of HRs in this artificial a posteriori population, which is not the same as the one that HRs were derived from. The ERG deemed the introduction of these additional uncertainties, some of which were shown to have a substantial effect on the ICERs in the ERG's exploratory analysis, as unjustified, given that the need for response-based analysis and its improvement over conventional analysis was not demonstrated. Further issues related to the exclusion of cisplatin plus gemcitabine as a comparator, inconsistencies in the source for nivolumab-related effectiveness, resource use, utilities and adverse event data (use of CheckMate 275 and CheckMate 032 for effectiveness, use of CheckMate 275 only for the others), the inclusion of adverse events with incidence smaller than 5%, the calculation of dose intensity, and the exclusion of important parameters from, and inclusion of inappropriate parameters in, the PSA.

There is substantial uncertainty about the relative treatment effectiveness estimates, which were entirely derived from single-arm studies, using a simulated treatment comparison that aimed at correcting for differences in the study populations. The residual bias could not be quantified in the company's analysis,

and cost effectiveness results should therefore be interpreted with extreme caution. Model estimates for nivolumab were not externally validated, apart from the comparison with NSCLC data, which may not be appropriate. The uncertainty introduced by the derived time-varying HRs was unfortunately not assessed within the PSA. In exploratory analysis, the ERG attempted to give a measure of parts of this uncertainty by using a naive comparison as opposed to the STC, and time-fixed HRs as opposed to time-varying HRs.

In exploratory analysis, the ERG found that using the naïve comparison resulted in pronounced increases in the ICERs (£92,335, £64,914, dominated, £65,593 per QALY gained when comparing nivolumab against docetaxel, paclitaxel, cisplatin plus gemcitabine and BSC respectively). These further increased in an extreme scenario where no relative treatment effect was assumed for nivolumab. The use of time-independent HRs also had a significant effect on ICERs, with some ICERs increasing and others decreasing compared to the ERG base-case ICERs (£71,639, £95,775, £76,576, £55,577 per QALY gained when comparing nivolumab against docetaxel, paclitaxel, cisplatin plus gemcitabine and BSC respectively). The use of alternative parametric time-to-event models for OS (lognormal) and PFS (log-logistic) in the conventional approach produced ICERs more favourable to nivolumab (£45,721, £39,286, £72,732, £38,147 per QALY gained when comparing nivolumab against docetaxel, paclitaxel, cisplatin plus gemcitabine and BSC respectively). Using the response-based analysis with alternative time-to-event models for OS, PFS and TTD, however, resulted in a marked increase in ICERs compared with the response-based company's base-case (£77,597, £67,608, £143,923, £64,282 per QALY gained when comparing nivolumab against docetaxel, paclitaxel, cisplatin plus gemcitabine and BSC respectively). Lastly, the alternative landmark drove the company's base-case ICERs up (£75,094, £71,255, £87,022, £61,647 per QALY gained when comparing nivolumab against docetaxel, paclitaxel, cisplatin plus gemcitabine and BSC respectively). The ERG also found that the use of different parameter values for the fractional polynomial model alone resulted in large variation in absolute costs and QALYs (Table 6.3). These findings illustrate how uncertain the presented cost effectiveness results are.

In conclusion, given the ERG base-case ICERs are estimated to be substantially above £60,000 per QALY gained, and the large uncertainty regarding comparative treatment effectiveness in combination with the lack of appropriate validation, uncertainty around the cost effectiveness of nivolumab remains substantial.

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

In Section 5.3 the ERG's base-case was presented, which was based on various changes compared to the company's base-case. Table 6.1 shows how each individual change impacts the ICER plus the combined effect of all changes simultaneously. The analyses numbers in Table 6.1 correspond to the analyses numbers reported in Section 5.3. Also, the exploratory analysis is presented in Table 6.2 (conditional on the ERG base-case). Finally, the threshold analyses are discussed in Section 5.3.2. Appendix 1 contains technical details on the analyses performed by the ERG.

Table 6.1: ERG base-case (probabilistic), nivolumab with PAS

	Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	Nivolumab ICER (£/QALY)
Probabilistic Company base-case[a]	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£12,748	0.82	[redacted]	[redacted]	£54,131
	Paclitaxel	£14,186	0.71	[redacted]	[redacted]	£45,482
	Cis+gem	£30,443	1.34	[redacted]	[redacted]	£100,417
	BSC	£8,811	0.57	[redacted]	[redacted]	£44,873
Fixing errors (1) and (2)	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£12,744	0.82	[redacted]	[redacted]	£50,974
	Paclitaxel	£14,155	0.71	[redacted]	[redacted]	£42,715
	Cis+gem	£29,969	1.34	[redacted]	[redacted]	£91,773
	BSC	£8,813	0.58	[redacted]	[redacted]	£42,532
Proportions of responders based on OS for TTD (4)[b]	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£12,779	0.82	[redacted]	[redacted]	£50,889
	Paclitaxel	£14,162	0.71	[redacted]	[redacted]	£42,644
	Cis+gem	£29,960	1.35	[redacted]	[redacted]	£92,606
	BSC	£8,819	0.58	[redacted]	[redacted]	£42,435
Removing AEs with incidence < 5% (5)[b]	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£12,810	0.82	[redacted]	[redacted]	£51,023
	Paclitaxel	£14,205	0.71	[redacted]	[redacted]	£42,870
	Cis+gem	£29,982	1.34	[redacted]	[redacted]	£92,433
	BSC	£8,858	0.58	[redacted]	[redacted]	£42,566
	Nivolumab	[redacted]	[redacted]			

	Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	Nivolumab ICER (£/QALY)
Utilities from pooled CheckMate studies (6)[b]	Docetaxel	£12,803	0.84	[redacted]	[redacted]	£49,613
	Paclitaxel	£14,204	0.73	[redacted]	[redacted]	£41,605
	Cis+gem	£29,994	1.39	[redacted]	[redacted]	£91,388
	BSC	£8,849	0.59	[redacted]	[redacted]	£41,406
Weight from pooled CheckMate studies (7)[b]	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£12,763	0.82	[redacted]	[redacted]	£52,682
	Paclitaxel	£14,165	0.71	[redacted]	[redacted]	£44,199
	Cis+gem	£29,975	1.34	[redacted]	[redacted]	£98,529
	BSC	£8,819	0.58	[redacted]	[redacted]	£43,780
Excluding parameters from PSA (8)[b]	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£12,763	0.82	[redacted]	[redacted]	£51,149
	Paclitaxel	£14,178	0.71	[redacted]	[redacted]	£42,868
	Cis+gem	£29,960	1.34	[redacted]	[redacted]	£92,876
	BSC	£8,829	0.57	[redacted]	[redacted]	£42,632
Conventional instead of response-based analysis (9)[b]	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£12,507	0.72	[redacted]	[redacted]	£84,193
	Paclitaxel	£13,894	0.61	[redacted]	[redacted]	£65,302
	Cis+gem	£29,082	1.20	[redacted]	[redacted]	Dominated
	BSC	£8,736	0.55	[redacted]	[redacted]	£66,951
Missed doses when delayed > 7days (10)[b]	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£12,803	0.82	[redacted]	[redacted]	£52,858
	Paclitaxel	£14,198	0.71	[redacted]	[redacted]	£44,330
	Cis+gem	£30,315	1.35	[redacted]	[redacted]	£97,665
	BSC	£8,835	0.58	[redacted]	[redacted]	£43,958
ERG base-case (combining adjustments 1-10)	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£12,493	0.74	[redacted]	[redacted]	£87,709
	Paclitaxel	£13,866	0.63	[redacted]	[redacted]	£68,519

	Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	Nivolumab ICER (£/QALY)
	Cis+gem	£29,384	1.24	[redacted]	[redacted]	Nivolumab is dominated
	BSC	£8,696	0.56	[redacted]	[redacted]	£69,515

Note: [a] results have been reproduced by the ERG, based on the economic model submitted by the company in their clarification response; [b] this scenario is conditional on the fixing errors adjustment (adjustments 1 and 2)
ERG = Evidence Review Group; ICER = incremental cost effectiveness ratio; QALY = quality-adjusted life year

Table 6.2: Exploratory analyses; nivolumab with PAS

	Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	Nivolumab ICER (£/QALY)
Probabilistic Company base-case[a]	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£12,748	0.82	[redacted]	[redacted]	£54,131
	Paclitaxel	£14,186	0.71	[redacted]	[redacted]	£45,482
	Cis+gem	£30,443	1.34	[redacted]	[redacted]	£100,417
	BSC	£8,811	0.57	[redacted]	[redacted]	£44,873
ERG base-case	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£12,493	0.74	[redacted]	[redacted]	£87,709
	Paclitaxel	£13,866	0.63	[redacted]	[redacted]	£68,519
	Cis+gem	£29,384	1.24	[redacted]	[redacted]	Dominated
	BSC	£8,696	0.56	[redacted]	[redacted]	£69,515
Alternative parametric TTE models (lognormal for OS, log-logistic for PFS) (A.1)	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£13,173	1.01	[redacted]	[redacted]	£45,721
	Paclitaxel	£14,654	0.89	[redacted]	[redacted]	£39,286
	Cis+gem	£29,736	1.58	[redacted]	[redacted]	£72,732
	BSC	£9,235	0.72	[redacted]	[redacted]	£38,147
Naïve comparison data instead of STC results (A.3)	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£13,005	0.77	[redacted]	[redacted]	£92,335
	Paclitaxel	£13,914	0.60	[redacted]	[redacted]	£64,914
	Cis+gem	£30,910	1.56	[redacted]	[redacted]	Dominated

	Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	Nivolumab ICER (£/QALY)
	BSC	£8,630	0.52	[redacted]	[redacted]	£65,593
Time-independent HRs (A.4)	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£10,213	0.60	[redacted]	[redacted]	£71,639
	Paclitaxel	£13,081	0.78	[redacted]	[redacted]	£95,775
	Cis+gem	£26,584	0.86	[redacted]	[redacted]	£76,576
	BSC	£8,173	0.40	[redacted]	[redacted]	£55,577
Alternative assumptions for PFS HRs for BSC and cis+gem (A.5)	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£12,507	0.74	[redacted]	[redacted]	£87,863
	Paclitaxel	£13,858	0.63	[redacted]	[redacted]	£68,679
	Cis+gem	£34,999	1.26	[redacted]	[redacted]	Dominated
	BSC	£8,698	0.55	[redacted]	[redacted]	£68,369
AE disutilities and resource use from TA ID971 (A.6)	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£12,068	0.74	[redacted]	[redacted]	£89,222
	Paclitaxel	£13,695	0.63	[redacted]	[redacted]	£69,051
	Cis+gem	£26,508	1.26	[redacted]	[redacted]	Dominated
	BSC	£8,750	0.56	[redacted]	[redacted]	£69,622
UK dosage schedule for cis+gem (A.7)	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£12,476	0.74	[redacted]	[redacted]	£87,722
	Paclitaxel	£13,852	0.63	[redacted]	[redacted]	£68,621
	Cis+gem	£31,195	1.24	[redacted]	[redacted]	Dominated
	BSC	£8,678	0.56	[redacted]	[redacted]	£69,560
No treatment effect of nivolumab vs comparators (A.8)	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£13,726	1.19	[redacted]	[redacted]	£5,740,183
	Paclitaxel	£14,270	1.19	[redacted]	[redacted]	£11,382,482
	Cis+gem	£32,028	1.15	[redacted]	[redacted]	£415,600
	BSC	£10,635	1.16	[redacted]	[redacted]	£1,168,837
Response-based analysis	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£12,783	0.84	[redacted]	[redacted]	£53,273

	Technologies	Total costs	Total QALYs	Incremental costs	Incremental QALYs	Nivolumab ICER (£/QALY)
using ERG base-case (B.1)	Paclitaxel	£14,163	0.73	[redacted]	[redacted]	£44,877
	Cis+gem	£30,310	1.39	[redacted]	[redacted]	£103,186
	BSC	£8,811	0.59	[redacted]	[redacted]	£44,183
Response-based analysis using alternative TTE models for OS, PFS, but not TTD (B.2)	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£12,475	0.77	[redacted]	[redacted]	£78,795
	Paclitaxel	£13,983	0.68	[redacted]	[redacted]	£68,594
	Cis+gem	£29,893	1.25	[redacted]	[redacted]	£146,721
	BSC	£8,678	0.55	[redacted]	[redacted]	£65,249
Response-based analysis using alternative TTE models for OS, PFS and TTD (B.3)	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£12,452	0.77	[redacted]	[redacted]	£77,597
	Paclitaxel	£13,948	0.67	[redacted]	[redacted]	£67,608
	Cis+gem	£29,880	1.25	[redacted]	[redacted]	£143,923
	BSC	£8,662	0.55	[redacted]	[redacted]	£64,282
Response-based analysis using 26-week landmark (B.4)	Nivolumab	[redacted]	[redacted]			
	Docetaxel	£10,849	0.51	[redacted]	[redacted]	£75,094
	Paclitaxel	£13,689	0.52	[redacted]	[redacted]	£71,255
	Cis+gem	£28,678	0.79	[redacted]	[redacted]	£87,022
	BSC	£8,035	0.35	[redacted]	[redacted]	£61,647

Note: [a] results have been reproduced by the ERG, based on the economic model submitted by the company in their clarification response
ERG = Evidence Review Group; ICER = incremental cost effectiveness ratio; QALY = quality-adjusted life year

Table 6.3. Impact of using different parameter values in the fractional polynomial model for NMA

Technologies	Incremental costs (CI) of nivolumab vs comparators		Incremental QALYs (CI) of nivolumab vs comparators		ICER of nivolumab vs comparators	
	Lower	Upper	Lower	Upper	Range based on CIs for incremental costs and QALYs	
Docetaxel	[redacted]	[redacted]	[redacted]	[redacted]	£178,199	£52,441
Paclitaxel	[redacted]	[redacted]	[redacted]	[redacted]	£160,141	£47,615
Cis + gem	[redacted]	[redacted]	[redacted]	[redacted]	Dominated	£35,146
BSC	[redacted]	[redacted]	[redacted]	[redacted]	£96,636	£43,847

7. END OF LIFE

The company discusses the end-of life criteria in section B.2.13.2 of the CS, arguing that nivolumab fulfils the end-of-life criteria in this appraisal.[2]

This argument is partly based on lack of evidence to argue that it does not – '*no study provided evidence of OS estimates for this patient population that approached the 24 months that represents the threshold for NICE's end of life criteria'*, and partly on very weak evidence from the economic model based on a comparison of single arm studies – '*The economic analysis predicted mean life years per patient with nivolumab of 2.78 years (33.36 months). In comparison, predicted mean life years per patient with comparator therapies were 1.19 years (14.28 months) with paclitaxel, 1.40 years (16.80 months) with docetaxel and 1.01 years (12.12 months) with BSC'*.

We agree that there is no evidence to argue that nivolumab does not fulfil the end-of-life criteria in this appraisal. But, at the same time, there is no robust evidence to argue that it does.

8. OVERALL CONCLUSIONS

8.1 Statement of principal findings

Searches were carried out in accordance with the NICE guide to the methods of technology appraisal sections 5.2.2 and 5.2.4 using a good range of databases. The systematic review was performed to a good standard.

The identification of two single arm studies for nivolumab, CheckMate 275 and CheckMate 032, precluded any conventional mixed treatment comparison (MTC) or indirect meta-analysis. There were no studies that could provide a common comparator to support any indirect comparison or MTC. As a consequence the company decided to perform an unanchored (no common comparator) stimulated treatment comparison (STC). In terms of ORR the main analysis using the fixed effect model presented finds that nivolumab is significantly better than BSC and docetaxel. No significant differences were found for nivolumab paclitaxel and gemcitabine. In the random effects model nivolumab is only statistically significantly superior to BSC. In the naïve indirect comparison nivolumab is superior to all three comparators in the fixed effect model but only to BSC in the random effects model. The results of the analysis using fixed effect fractional polynomial model (allowing variation of HRs over time) based on the STC show that for OS and PFS nivolumab is superior to all comparators at most time points. However, the credible intervals for the HRs are quite wide, crossing 1 in many cases. The results of the naïve indirect comparison i.e. with the fractional polynomial model, but without the STC, were not reported. The results assuming a proportional hazards model i.e. fixed HRs were reported in the response to the clarification request, although were derived by a method that lacked validity and were quite different to those obtained by the ERG using a method advocated in the paper on which the company analysis was based. Very few of the many functional forms of the fractional polynomial model were explored.

The methods used by the company to conduct the STC largely follow those described in NICE DSU TSD 18, but, as stated in the same TSD, given no comparative data (unanchored analysis) the results obtained should be treated with caution.[1] The ERG found several serious limitations in the STC analysis. In particular, the major assumption for unanchored STC is that all effect modifiers or prognostic variables are accounted for. Not all of the key characteristics (possible effect modifiers or prognostic variables) for the STC were reported for all comparator trials, therefore imputations were required for these characteristics which were based on correlations to the baseline characteristics in the nivolumab trials. Also, the method used for the prediction models lacked transparency; the results at each stage of the stepwise selection process were not provided. In particular, it is not clear that the most parsimonious model is the best model. The ERG was able to produce the results based on a naïve comparison (without the STC), which verified the adoption of the PH model used in the STC i.e. all HRs of nivolumab versus each comparator were multiplied by a single factor, the HR of the adjusted (by the STC) vs. unadjusted hazard for nivolumab. However, it would have been useful to see an STC that was based on prediction models with more covariates including all eleven considered. The only external test of validity of the STC i.e. the 'out-of-sample' method seemed to either show insufficient reduction in bias or be inapplicable given the use of the fractional polynomial model that was used for survival analysis. As stated on page 56 of TSD 18: '*The size of this systematic error can certainly be reduced, and probably substantially, by appropriate use of…STC. Much of the literature on unanchored … STC acknowledges the possibility of residual bias due to unobserved prognostic variables and effect modifiers; however, it is not made clear that the accuracy of the resulting estimates is entirely unknown, because there is no analysis of the potential magnitude of residual bias, and hence no idea of the degree of error in the unanchored estimates. It is, of course, most unlikely that systematic error has been eliminated.*

Hoaglin,[72, 73] in a series of letters critiquing an unanchored comparison by Di Lorenzo et al.[78] based upon a matching approach similar to MAIC, remarked that, without providing evidence that the adjustment compensates for the missing common comparator arms and the resulting systematic error, the ensuing results "are not worthy of consideration".'[1]

No formal comparison of AEs including no evidence synthesis was performed, although it might be reasonable to conclude, based on few data from the comparators, that the rate of key AEs was generally similar to or lower than the comparators.

In conclusion, it is difficult to be sure what the effectiveness of nivolumab is in comparison to the comparators in the scope. Evidence from directly examining the single arms of the trial data indicates little difference between the outcomes measured from the nivolumab and comparator studies. Such a naive comparison carries a high risk of bias. STC analysis was used to try and reduce this bias, but there is also no clear evidence that risk of bias was reduced by the STC analysis. Multiple limitations in the STC were identified and the test of validity recommended by TSD 18, the 'out-of-sample' method either lack of success in reducing the bias if it is applicable at all given the lack of data and PF model. The ERG was able to estimate the unadjusted hazards for nivolumab, but not with estimates of uncertainty. The effect of an analysis based on different combinations of covariates in the prediction model used to make the adjustment remains unknown.

With regards to the health economic model submitted by the company, the ERG demonstrated that there was large uncertainty surrounding the ICERs and that a number of alternative assumptions could change the ICERs significantly. Most crucially, the ERG questioned the need for the company's response-based approach to survival analysis, which was deemed insufficiently justified. If a response-based approach was indeed deemed necessary, then other, more established methods, should be explored (spline-based or mixture cure models, as recommended in TSD 14).[38] However, it should also be noted, that the company's approach to implementing the response-based approach necessitated additional model assumptions and increased uncertainty. The resulting model predictions were different from those obtained using a conventional approach to an extent that might be implausible; the lack of validation by experts further made the ERG question the plausibility of the company's base-case. Furthermore, the exclusion of cisplatin plus gemcitabine from the base-case stood in contrast to the scope and was inappropriately justified.

Apart from this, numerous issues were identified by the ERG. The ERG was able to adjust/correct some of these in its base-case. This resulted in ICERs (probabilistic) of £87,709, £68,519 and £69,515 for nivolumab (with PAS) versus docetaxel, paclitaxel and BSC, respectively. Cisplatin plus gemcitabine dominated nivolumab.

Additional sensitivity analyses were performed to examine the potential impact of alternative assumptions on the cost effectiveness estimates. These included two scenarios in which changes were implemented: a) exploratory analyses performed using the ERG base-case, and b) exploratory analyses performed using the ERG base-case, except that a response-based approach was used. Scenarios exploring the uncertainty about the treatment and relative effectiveness evidence significantly increased the ICERs. Using one example set of alternative parametric time-to-event models within the ERG base-case decreased the ICERs significantly. Finally, using the response-based approach significantly decreased the ICER, but these ICERs were shown to increase significantly with the use of best-fitting parametric time-to-event models. In addition, alternative parameter values informing the fractional polynomial model for the NMA showed that this model feature alone could have a vast impact on the ICERs.

In conclusion, given the ERG base-case ICERs are estimated to be substantially above £60,000 per QALY gained, and the large uncertainty regarding comparative treatment effectiveness in combination with the lack of appropriate validation, uncertainty around the cost effectiveness of nivolumab remains substantial.

8.2 Strengths and limitations of the assessment

The searches for eligible studies in the CS were well documented and easily reproducible. Searches were carried out on a wide range of databases and other resources. Supplementary searches of conference proceedings and organisational websites, and the checking of references lists were undertaken by the company in order to identify additional studies not retrieved by the main searches. However, the search for English language studies only in the MEDLINE and Embase searches in the clinical effectiveness section was felt to be a limitation. The systematic review was well conducted, but no randomised controlled trials (RCTs) were identified for nivolumab and there were no studies that directly compared nivolumab with any specified comparator. Furthermore, there were no studies that could provide a common comparator to support indirect comparison or MTC. The STC analysis is compromised by many limitations (listed earlier) which impairs the ability to critique the presence of residual bias. Given that the TSD 18 states that without providing evidence that the adjustment compensates for the missing common comparator arms and the resulting systematic error, the ensuing results *"…are not worthy of consideration…"* the ERG does not think the STC methods are sufficiently reported nor validated to sustain the companies claims.[1]

The economic model had a structure similar to past NICE technology appraisals in metastatic cancer but deviated from conventional survival modelling in that it used a response-based approach. This was inconsistently implemented, insufficiently justified and alternative approaches were not explored. The uncertainty and bias potentially introduced by this approach could not be completely explored. The lack of validation of model predictions raised concerns about the validity of CS model results. Lastly, the exclusion of cisplatin plus gemcitabine from the company's base-case stands in contrast to the scope and lacked appropriate justification.

8.3 Suggested research priorities

The ERG recommends the conduct of an RCT of nivolumab versus at least one of the comparators or perhaps an investigator choice design, which might be lacking in terms of power, depending on time believed to be reasonable to recruit, but would provide at least some unbiased evidence of effectiveness.

9. REFERENCES

[1] Phillippo D, Ades T, Dias S, Palmer S, Abrams KR, Welton N. *NICE DSU Technical Support Document 18: methods for population-adjusted indirect comparisons in submissions to NICE*. Sheffield: NICE Decision Support Unit, 2016. 81p.

[2] Bristol-Myers Squibb Pharmaceuticals Ltd. *Nivolumab for treating metastatic or unresectable urothelial cancer after platinum-based chemotherapy [ID995]. Document B: Company evidence submission. Submission to National Institute of Health and Clinical Excellence. Single technology appraisal (STA)*: Bristol-Myers Squibb Pharmaceuticals Ltd, 2017. 143p.

[3] Cancer Research UK. Cancer statistics for the UK [Internet]. 2017 [accessed 20.7.17]. Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics

[4] National Institute for Health and Care Excellence. *Bladder cancer: diagnosis and management* London: NICE, 2015. 57p. Available from: http://nice.org.uk/guidance/ng2

[5] Bristol-Myers Squibb. *Meeting minutes, Clinical Advisory Board Meeting: 06 March 2017, 0930-1600 [Not in the public domain]*. 30 Euston Square, London, UK: Bristol-Myers Squibb, 2017

[6] National Institute for Health and Care Excellence. *Nivolumab for treating metastatic or unresectable urothelial cancer. Final scope [Internet]*. London: NICE, 2017 [accessed 3.7.17]. 4p. Available from: https://www.nice.org.uk/guidance/gid-ta10163/documents/final-scope

[7] Bristol-Myers Squibb Pharmaceutical Ltd. *Nivolumab for treating metastatic or unresectable urothelial cancer after platinum-based chemotherapy [ID995] – Response to request for clarification from the ERG*: Bristol-Myers Squibb Pharmaceutical Ltd, 2017. 97p.

[8] Sharma P, Retz M, Siefker-Radtke A, Baron A, Necchi A, Bedke J, et al. Nivolumab in metastatic urothelial carcinoma after platinum therapy (CheckMate 275): a multicentre, single-arm, phase 2 trial. *Lancet Oncol* 2017;18(3):312-322.

[9] Sharma P, Callahan MK, Bono P, Kim J, Spiliopoulou P, Calvo E, et al. Nivolumab monotherapy in recurrent metastatic urothelial carcinoma (CheckMate 032): a multicentre, open-label, two-stage, multi-arm, phase 1/2 trial. *Lancet Oncol* 2016;17(11):1590-1598.

[10] Bristol-Myers Squibb Pharmaceuticals Ltd. *CheckMate 275: Clinical Study Report for Study CA209275 (25th July 2016)*, 2016

[11] Bristol-Myers Squibb Pharmaceuticals Ltd. *CheckMate 032: Clinical Study Report for Study CA209032 (29th June 2016).* 2016

[12] European Medicines Agency. *Summary of opinion (initial authorisation). Maviret: glecaprevir/pibrentasvir [Internet]*. London: European Medicines Agency, 2017 [accessed 15.8.17]. 1p. Available from:

http://www.ema.europa.eu/docs/en_GB/document_library/Summary_of_opinion_-_Initial_authorisation/human/004430/WC500229790.pdf

[13] Gondo T, Ohori M, Hamada R, Tanaka A, Satake N, Takeuchi H, et al. The efficacy and safety of gemcitabine plus cisplatin regimen for patients with advanced urothelial carcinoma after failure of M-VAC regimen. *Int J Clin Oncol* 2011;16(4):345-51.

[14] Ozawa A, Tanji N, Ochi T, Yanagihara Y, Kikugawa T, Yamaguchi A, et al. Gemcitabine and cisplatin for advanced urothelial carcinomas: the Ehime University Hospital experience. *Int J Clin Oncol* 2007;12(4):279-83.

[15] Jones RJ, Hussain SA, Protheroe AS, Birtle A, Chakraborti P, Huddart RA, et al. Randomized phase II study investigating pazopanib versus weekly paclitaxel in relapsed or progressive urothelial cancer. *J Clin Oncol* 2017;35(16):1770-1777.

[16] Petrylak DP, Tagawa ST, Kohli M, Eisen A, Canil C, Sridhar SS, et al. Docetaxel as monotherapy or combined with ramucirumab or icrucumab in second-line treatment for locally advanced or metastatic urothelial carcinoma: an open-label, three-arm, randomized controlled phase II trial. *J Clin Oncol* 2016;34(13):1500-9.

[17] Canadian Agency for Drugs and Technologies in Health. *CADTH peer review checklist for search strategies [Internet]*. Ottawa: CADTH, 2013 [accessed 17.7.13]. 3p. Available from: http://www.cadth.ca/en/resources/finding-evidence-is

[18] National Institute for Health and Care Excellence. *Single technology appraisal: user guide for company evidence submission template [Internet]*. London: NICE, 2015 [accessed 3.7.17]. 52p. Available from: https://www.nice.org.uk/guidance/pmg24/resources/single-technology-appraisal-user-guide-for-company-evidence-submission-template-pdf-72286715419333

[19] Lefebvre C, Manheimer E, Glanville J. 6.4.9 Language, date and document format restrictions (Chapter 6: Searching for studies) [Internet]. In: Higgins JPT, Green S, editors. Version 5.1.0 [updated March 2011]: The Cochrane Collaboration, 2011 [accessed 7.8.17]. Available from: http://handbook.cochrane.org/

[20] Bristol-Myers Squibb Pharmaceuticals Ltd. *Nivolumab for treating metastatic or unresectable urothelial cancer after platinum-based chemotherapy [ID995]. Appendices. Submission to National Institute of Health and Clinical Excellence. Single technology appraisal (STA)*: Bristol-Myers Squibb Pharmaceuticals Ltd, 2017. 257p.

[21] National Institute for Health and Care Excellence. *Nivolumab for treating metastatic or unresectable urothelial cancer after platinum-based chemotherapy [ID995]. Clarification letter*. London: NICE, 2017

[22] Centre for Reviews and Dissemination. *Quality Assessment: Cohort Studies (Box 5.9). CRD Report 4: Undertaking systematic reviews of research on effectiveness*. York: University of York, 2001

[23] Kim YS, Lee SI, Park SH, Park S, Hwang IG, Lee SC, et al. A phase II study of weekly docetaxel as second-line chemotherapy in patients with metastatic urothelial carcinoma. *Clin Genitourin Cancer* 2016;14(1):76-81.

[24] McCaffrey JA, Hilton S, Mazumdar M, Sadan S, Kelly WK, Scher HI, et al. Phase II trial of docetaxel in patients with advanced or metastatic transitional-cell carcinoma. *J Clin Oncol* 1997;15(5):1853-7.

[25] Vaughn DJ, Broome CM, Hussain M, Gutheil JC, Markowitz AB. Phase II trial of weekly paclitaxel in patients with previously treated advanced urothelial cancer. *J Clin Oncol* 2002;20(4):937-40.

[26] Bellmunt J, Theodore C, Demkov T, Komyakov B, Sengelov L, Daugaard G, et al. Phase III trial of vinflunine plus best supportive care compared with best supportive care alone after a platinum-containing regimen in patients with advanced transitional cell carcinoma of the urothelial tract. *J Clin Oncol* 2009;27(27):4454-61.

[27] Choueiri TK, Ross RW, Jacobus S, Vaishampayan U, Yu EY, Quinn DI, et al. Double-blind, randomized trial of docetaxel plus vandetanib versus docetaxel plus placebo in platinum-pretreated metastatic urothelial cancer. *J Clin Oncol* 2012;30(5):507-12.

[28] Joly F, Houede N, Noal S, Chevreau C, Priou F, Chinet-Charrot P, et al. Do patients with advanced urothelial carcinoma benefit from weekly paclitaxel chemotherapy? A GETUG phase II study. *Clin Genitourin Cancer* 2009;7(2):E28-33.

[29] Jansen JP. Network meta-analysis of survival data with fractional polynomials. *BMC Med Res Methodol* 2011;11:61.

[30] Choueiri TK, Ross RW, Jacobus S, Vaishampayan U, Yu EY, Quinn DI, et al. Double-blind, randomized trial of docetaxel plus vandetanib versus docetaxel plus placebo in platinum-pretreated metastatic urothelial cancer. *Journal of Clinical Oncology* 2012;30(5):507-12.

[31] Jones R, Hussain S, Protheroe A, Birtle A, Chakraborti P, Huddart R, et al. Randomized Phase II Study Investigating Pazopanib Versus Weekly Paclitaxel in Relapsed or Progressive Urothelial Cancer. *J Clin Oncol* 2017;0(0):JCO.2016.70.7828.

[32] Petrylak DP, Tagawa ST, Kohli M, Eisen A, Canil C, Sridhar SS, et al. Docetaxel as monotherapy or combined with ramucirumab or icrucumab in second-line treatment for locally advanced or metastatic urothelial carcinoma: an open-label, three-arm, randomized controlled phase II trial. *J Clin Oncol* 2016;34(13):1500-9.

[33] National Institute for Health and Care Excellence. *Guide to the methods of technology appraisal 2013 [Internet]*. London: NICE, 2013 [accessed 23.8.17]. 93p. Available from: http://publications.nice.org.uk/pmg9

[34] National Institute for Health and Care Excellence. *Vinflunine for the treatment of advanced or metastatic transitional cell carcinoma of the urothelial tract [TA272]*. London: National Institute for Health and Care Excellence, 2013 [accessed 3 Feb 2017] Available from: https://www.nice.org.uk/guidance/ta272

[35] Robinson P, Maase H, Bhalla S, Kielhorn A, Aristides M, Brown A, et al. Cost-utility analysis of the GC versus MVAC regimens for the treatment of locally advanced or metastatic bladder cancer. *Expert Rev Pharmacoecon Outcomes Res* 2004;4(1):27-38.

[36] Scottish Medicines Consortium. *Vinflunine (as ditartrate), 25mg/mL, concentrate for solution for infusion (Javlor(R)) SMC No. (686/11)*, 2015 [accessed 27.02.17] Available from: https://www.scottishmedicines.org.uk/files/advice/vinflunine_Javlor_Resubmission_FINAL_June_2015_for_website.pdf

[37] Woods B, Sideris E, Palmer S, Latimer N, Soares M. *NICE DSU Technical Support Document 19: Partitioned survival analysis for decision modelling in health care: a critical review*. Sheffield: Decision Support Unit, ScHARR, 2017. 72p. Available from: http://scharr.dept.shef.ac.uk/nicedsu/wp-content/uploads/sites/7/2017/06/Partitioned-Survival-Analysis-final-report.pdf

[38] Latimer N. *NICE DSU Technical Support Document 14: Undertaking survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data*. Sheffield: Decision Support Unit, ScHARR, 2017. 52p. Available from: http://scharr.dept.shef.ac.uk/nicedsu/wp-content/uploads/sites/7/2016/03/NICE-DSU-TSD-Survival-analysis.updated-March-2013.v2.pdf

[39] Dafni U. Landmark analysis at the 25-year landmark point. *Circ Cardiovasc Qual Outcomes* 2011;4(3):363-71.

[40] National Institute for Health and Care Excellence. *Nivolumab for treating recurrent or metastatic squamous-cell carcinoma of the head and neck after platinum-based chemotherapy - committee papers [Internet]*, 2017 [accessed 9.8.17] Available from: https://www.nice.org.uk/guidance/gid-ta10080/documents/committee-papers

[41] Bellmunt J, Fougeray R, Rosenberg JE, von der Maase H, Schutz FA, Salhi Y, et al. Long-term survival results of a randomized phase III trial of vinflunine plus best supportive care versus best supportive care alone in advanced urothelial carcinoma patients after failure of platinum-based chemotherapy. *Ann Oncol* 2013;24(6):1466-72.

[42] Berdeaux G, Roskell N, Hemstock M. Influence of the method of analysis on estimates of QALY treatment difference: Phase III trial of vinflunine versus best supportive care in patients with TCCU. *Value Health* 2015;18(7):A467.

[43] von der Maase H, Hansen SW, Roberts JT, et al. Gemcitabine and cisplatin versus methotrexate, vinblastine, doxorubicin, and cisplatin in advanced or metastatic bladder cancer: results of a large, randomized, multinational, multicenter, phase III study. *J Clin Oncol* 2000;18:3068-77.

[44] Davey P, Rajan N, De Souza P, et al. Examining preferences, utility values and cost-effectiveness for gemcitabine plus cisplatin (GEM/cis) for the treatment of bladder cancer - a discrete choice conjoint analysis conducted in Australia. *Eur J Cancer* 2001;37:S163.

[45] Pickard AS, Jiang R, Lin HW, et al. Using patient-reported outcomes to compare relative burden of cancer: EQ-5D and Functional Assessment of Cancer Therapy-General in eleven types of cancer. *Clin Ther* 2016;38:769-777.

[46] Attard CL, Brown S, Alloul K, Moore MJ. Cost-effectiveness of folfirinox for first-line treatment of metastatic pancreatic cancer. *Curr Oncol* 2014;21(1):e41-51.

[47] Nafees B, Stafford M, Gavriel S, Bhalla S, Watkins J. Health state utilities for non small cell lung cancer. *Health Qual Life Outcomes* 2008;6:84.

[48] Beusterien KM, Davies J, Leach M, Meiklejohn D, Grinspan JL, O'Toole A, et al. Population preference values for treatment outcomes in chronic lymphocytic leukaemia: a cross-sectional utility study. *Health Qual Life Outcomes* 2010;8:50.

[49] National Institute for Health and Care Excellence. *Nintedanib for pre Nintedanib for previously treated locally viously treated locally advanced, metastatic, or locally recurrent advanced, metastatic, or locally recurrent non-small-cell lung cancer [TA347]*. London: National Institute for Health and Care Excellence, 2015 [accessed 23.8.17] Available from: https://www.nice.org.uk/guidance/ta347

[50] Frederix GW, Quadri N, Hovels AM, van de Wetering FT, Tamminga H, Schellens JH, et al. Utility and work productivity data for economic evaluation of breast cancer therapies in the Netherlands and Sweden. *Clin Ther* 2013;35(4):e1-7.

[51] National Institute for Health and Care Excellence. Nivolumab for treating recurrent or metastatic squamous-cell carcinoma of the head and neck after platinum-based chemotherapy [ID971] [Internet]. 2017 [accessed 17.7.17]. Available from: https://www.nice.org.uk/guidance/indevelopment/gid-ta10080/documents

[52] Addicott R, Dewar S. *Improving choice at end of life: a descriptive analysis of the impact and costs of the Marie Curie Delivering Choice programme in Lincolnshire*. London: Kings Fund, 2008. 45p. Available from: https://www.kingsfund.org.uk/publications/improving-choice-end-life

[53] Brahmer J, Horn L, Jackman D, Spigel D, Antonia S, Hellmann M, et al. Abstract CT077: Five-year follow-up from the CA209-003 study of nivolumab in previously treated advanced non-small cell lung cancer (NSCLC): Clinical characteristics of long-term survivors. *Cancer Res* 2017;77(13 Supplement):CT077-CT077.

[54] Sideris S, Aoun F, Zanaty M, Martinez NC, Latifyan S, Awada A, et al. Efficacy of weekly paclitaxel treatment as a single agent chemotherapy following first-line cisplatin treatment in urothelial bladder cancer. *Mol Clin Oncol* 2016;4(6):1063-1067.

[55] Bellmunt J, Fougeray R, Rosenberg JE, von der Maase H, Schutz FA, Salhi Y, et al. Long-term survival results of a randomized phase III trial of vinflunine plus best supportive care versus best supportive care alone in advanced urothelial carcinoma patients after failure of platinum-based chemotherapy. *Annals of Oncology* 2013;24(6):1466-1472.

[56] Kaltenthaler E, Carroll C, Hill-McManus D, Scope A, Holmes M, Rice S, et al. The use of exploratory analyses within the National Institute for Health and Care Excellence single technology appraisal process: an evaluation and qualitative analysis. *Health Technol Assess* 2016;20(26):1-48.

Appendix 1: Details of ERG analyses (for validation purposes)

Altered cells are printed in *italics*.

Fixing errors

1. Error in the use of UK life tables and conversion of background mortality rate to probability

Addition of tab "National life table"; General mortality data!CA3:CB73

2. Error in calculating dose intensity

Drug costs!E14:E20; Drug costs!E32:E36; Drug costs!E43:E47; Drug costs!E54:E58; Drug costs!E65:E69; Costs & Resource Use!E32:E35

Fixing violations

3. Exclusion of cisplatin plus gemcitabine from base-case and fully incremental analysis in PSA.

PSA Simulation!H11;PSA Simulation!R11; PSA Simulation Y14:AC10013

Of note: the ERG added the total LY for each comparator (in each PSA draw) in columns J to N of the PSA Simulation-sheet.

4. Calculation of responder and non-responder proportions for response-based TTD analysis based on OS and PFS, thereby double-counting patients.

Discontinuation!CM23:CN24

5. Adverse events with an incidence <5% were included in the model, despite the company stating that these should be excluded.

Adverse Events!I13; Adverse Events!I17; Adverse Events!J13; Adverse Events!J15; Adverse Events!K13; Adverse Events!K15

6. Use of utilities from CheckMate 275 only.

LIVE!E32:E33

7. Use of BSA and weight from CheckMate 275 only.

Set-Up!E28

8. Inappropriate parameters in PSA: Patient characteristics were included in the PSA, although they are considered first order uncertainty and typically not reflected in cohort model PSAs.

PSA Distributions!J13:J16; PSA Distributions!J19:J22

Matters of judgement

9. Use of response-based analysis, without sufficient justification and despite it introducing additional uncertainty.

PFS & OS!BS11; PFS & OS!BP18:BU30; PFS & OS!BY21:CV470; PFS & OS!DL21:DM470; Discontinuation!AH27:AH447; Discontinuation!BD27:BD447

10. The assumption that all delayed doses are missed doses.

Costs & Resource Use!I24:I28