CONFIDENTIAL UNTIL PUBLISHED

Evidence Review Group Report commissioned by the NIHR HTA Programme on behalf of NICE

Necitumumab for untreated advanced, metastatic, squamous non-small-cell lung cancer

Produced by Southampton Health Technology Assessments Centre (SHTAC)

Authors
Karen Pickett, Research Fellow, SHTAC
Micah Rose, Research Fellow, SHTAC
Jill Colquitt, Senior Reviewer / Partner, Effective Evidence LLP
Emma Loveman, Senior Reviewer / Partner, Effective Evidence LLP
Geoff Frampton, Senior Research Fellow, SHTAC
Andy Clegg, Senior Reviewer / Partner, Effective Evidence LLP
Joanne Lord, Professorial Fellow in Health Economics, SHTAC

Correspondence to
Karen Pickett
Southampton Health Technology Assessments Centre
University of Southampton
First Floor, Epsilon House
Enterprise Road, Southampton Science Park
Southampton SO16 7NS
www.southampton.ac.uk/shtac

Date completed 24 March 2016

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 15/69/01.

Acknowledgements

We would like to thank a consultant oncologist, who has requested not to be named, for providing clinical advice and commenting on a draft of this report. We would also like to thank: Karen Welch, Information Scientist, SHTAC, for appraising the literature search strategies in the company's submission, running updates of the company's clinical and cost-effectiveness searches and searching for ongoing studies; and Jonathan Shepherd, Principal Research fellow, for acting as internal editor for the ERG report.

Declared competing interests of the authors

None from the authors. The consultant oncologist who advised the ERG declared all potential competing interests.

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:

Pickett K, Rose M, Colquitt J, Loveman E, Frampton G, Clegg A, Lord J. Necitumumab for untreated advanced, metastatic, squamous non-small-cell lung cancer: A Single Technology Appraisal. Southampton Health Technology Assessments Centre, 2015.

Contributions of authors

K Pickett (Research Fellow) critically appraised the clinical effectiveness systematic review, drafted the report and project managed the review. M Rose (Research Fellow) critically appraised the health economic systematic review, critically appraised the economic evaluation and drafted the report. J Colquitt (Senior Reviewer / Partner, Effective Evidence LLP) and E Loveman (Senior Reviewer / Partner, Effective Evidence LLP) critically appraised the clinical effectiveness systematic review and drafted the report. G Frampton (Senior Research Fellow) critically appraised the clinical effectiveness systematic review and the network meta-analysis (NMA), drafted the report and is the project guarantor. A Clegg (Senior Reviewer / Partner,

Effective Evidence LLP) critically appraised the NMA and drafted the report. J Lord (Professorial Fellow in Health Economics) critically appraised the health economic systematic review, critically appraised the economic evaluation and drafted the report.

Word count: 48,669

Key to colour highlighting used in report

Commercial in confidence (CIC) information in blue

Academic in confidence (AIC) information in yellow.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A&E	Accident and Emergency
AE	Adverse event
ASBI	Average Symptom Burden Index
BIC	Bayesian information criterion
BSA	Body surface area
CG	Clinical guideline
CHMP	Committee for Medicinal Products for Human Use
CR	Complete response
CS	Company's submission
CSR	Clinical study report
DCarbo	Docetaxel plus carboplatin
DCis	Docetaxel plus cisplatin
DIC	Deviance Information Criterion
DSU	Decision Support Unit
ECOG	Eastern Cooperative Oncology Group
EGFR	Epidermal growth factor receptor
EOL	End-of-life
ERG	Evidence Review Group
FDA	Food and Drug Administration
GCarbo	Gemcitabine plus carboplatin
GCis	Gemcitabine plus cisplatin
GCis + N	Necitumumab in combination with gemcitabine plus cisplatin
GP	General Practitioner
G + P	Gemcitabine in combination with paclitaxel
HR	Hazard ratio
HRQoL	Health-related quality of life
H-score	Immunohistochemistry score
ICER	Incremental cost-effectiveness ratio
ICTRP	International Clinical Trials Registry Platform
ITT	Intention-to-treat
KM	Kaplan-Meier
LCSS	Lung Cancer Symptom Scale
Nab-PCarbo	Nab-paclitaxel plus carboplatin
NHS	National Health Service
NICE	National Institute for Health and Care Excellence

NMA	Network meta-analysis
NSCLC	Non-small-cell lung cancer
ORR	Objective response rate
OS	Overall survival
PCarbo	Paclitaxel plus carboplatin
PCis	Paclitaxel plus cisplatin
PFS	Progression free survival
PR	Partial response
PS	Partitioned survival
PSA	Probabilistic sensitivity analysis
QALY	Quality-adjusted life year
RCT	Randomised controlled trial
RECIST	Response Evaluation Criteria in Solid Tumours
SAE	Serious adverse event
S-1Carbo	S-1 plus carboplatin
SmPC	Summary of Product Characteristics
TA	Technology Assessment
TK	Tyrosine kinase
TTD	Time to discontinuation
TTF	Time to treatment failure
UKCRN	UK Clinical Research Network
VCarbo	Vinorelbine plus carboplatin
VCis	Vinorelbine plus cisplatin
VCis + C	Vinorelbine plus cisplatin and cetuximab
VEGF	Vascular endothelial growth factor
VGD	Vinorelbine in combination with gemcitabine and docetaxel
WHO	World Health Organisation

SUMMARY

Scope of the company submission

The company's submission (CS) generally reflects the scope of this appraisal issued by the National Institute for Health and Care Excellence (NICE). This was to appraise the clinical and cost-effectiveness of necitumumab within its marketing authorisation for the treatment of untreated advanced, metastatic, squamous non-small-cell lung cancer (NSCLC). The necitumumab marketing authorisation states that necitumumab in combination with gemcitabine and cisplatin (GCis + N) is indicated for patients with locally advanced or metastatic epidermal growth factor receptor (EGFR) expressing squamous NSCLC who have not received prior chemotherapy. The company's original evidence submission for this appraisal did not include analyses of the efficacy, safety or cost-effectiveness of GCis + N among patients with EGFR expressing squamous NSCLC. The company, however, supplied additional clinical and cost-effectiveness analyses for this patient population during the appraisal, in their response to clarification questions from NICE and the Evidence Review Group (ERG). The submission assesses the clinical and cost-effectiveness of GCis + N compared with five of the eight comparator combination drug regimens specified in NICE's scope:

- Cisplatin in combination with gemcitabine (GCis)
- Cisplatin in combination with paclitaxel (PCis)
- Carboplatin in combination with gemcitabine (GCarbo)
- Carboplatin in combination with paclitaxel (PCarbo)
- Cisplatin in combination with docetaxel (DCis)

Insufficient evidence was available to enable a comparison with the remaining three:

- Carboplatin in combination with docetaxel (DCarbo)
- Cisplatin in combination with vinorelbine (VCis)
- Carboplatin in combination with vinorelbine (VCarbo)

Summary of submitted clinical effectiveness evidence

The company's submission to NICE included:

- A systematic literature review of direct evidence, which included one Phase III randomised controlled trial (RCT) (the SQUIRE trial[1]).
- A systematic review to inform a network meta-analysis (NMA), which included a total of 10 RCTs in four networks to provide direct and indirect evidence of the efficacy of GCis + N compared to GCis alone and the other squamous NSCLC treatments specified in

the scope (where evidence was available). The outcomes in the four networks were median OS, hazard ratio of OS, median PFS and hazard ratio of PFS.

The SQUIRE trial was a large (N = 1093), open-label, multicentre RCT carried out in 26 countries, including the UK [redacted]. It compared the efficacy of GCis + N to GCis alone in people with stage IV squamous NSCLC who had not received prior chemotherapy for advanced NSCLC. The trial included patients with Eastern Cooperative Oncology Group (ECOG) performance statuses of 0-2. The drug doses of necitumumab, gemcitabine and cisplatin used in the trial matched those specified in the necitumumab and gemcitabine marketing authorisations. In line with the marketing authorisation, participants treated with necitumumab received treatment with GCis + N, followed by, where appropriate, maintenance treatment with necitumumab alone. The scope-relevant outcomes the trial assessed were overall survival (OS), progression free survival (PFS), objective response rates (ORR), health-related quality of life (HRQoL) and adverse events (AEs). The company supplied the results of these outcomes in its submission for the intention-to-treat population (ITT; defines as all randomised patients) and some of these outcomes for three trial subgroups:

- a post-hoc Western European subgroup (including patients from Germany, France, Spain, Greece, Italy, UK, Portugal, Austria and Belgium);
- a post-hoc subgroup of patients with EGFR expressing squamous NSCLC (the subgroup most relevant to the licensed indication) from the ITT population;
- a post-hoc subgroup of patients with EGFR expressing squamous NSCLC from the Western European subgroup.

The company stated that the Western European subgroup was a more generalisable population to patients in England than the ITT population, but did not provide a clear rationale for this or demonstrate a statistically significant treatment interaction for this subgroup.

The SQUIRE trial showed that GCis + N resulted in a median OS benefit compared with GCis of an additional 1.7, 1.6, [redacted] and [redacted] months, respectively, in the EGFR expressing subgroup, ITT population, Western Europe subgroup and the EGFR expressing Western Europe subgroup. The associated hazard ratios (HRs) were statistically significant for all populations. Median PFS was statistically significantly slightly longer with GCis + N compared to GCis in the EGFR expressing and ITT populations, but not the two Western Europe subgroups. Objective response rates were statistically significantly higher with GCis + N than with GCis in the EGFR Western Europe subgroup only (statistical significance not reported for the ITT population).

[redacted]. The proportion of patients experiencing at least one serious adverse event (AE) was marginally higher during the treatment phase with GCis + N than during treatment with GCis. Venous thromboembolic events were experienced more frequently in those treated with GCis + N than GCis alone for any grade. In the ITT population, the GCis + N group also experienced rashes, hypomagnesaemia, and conjunctivitis more frequently than the GCis group alone. In the EGFR expressing subgroup from the ITT population, patients treated with GCis experienced higher rates of hypomagnesaemia than patients treated with GCis in the ITT population. Rates of any grade of rash were lower in the both treatment arms in the EGFR expressing subgroup than in the ITT population.

The company's systematic review conducted for the NMA identified enough evidence to enable comparisons of GCis + N against PCarbo, GCis, PCis, DCis and GCarbo on the OS and PFS outcomes only (no evidence was available for HRQoL or toxicity, which are the other outcomes specified in the inclusion criteria for the review). A comparison with VCis could only be made for median OS data analyses. The NMA mainly included subgroup analyses of patients with squamous NSCLC from trials including patients with other histological subtypes of NSCLC. Only one trial (the SQUIRE trial) included in the NMA focused exclusively on patients with squamous NSCLC. The NMA is broader than the licensed population in that it did not focus solely on patients with EGFR expressing squamous NSCLC.

[redacted]

Summary of submitted cost effectiveness evidence

The company's submission to NICE includes:

1. A review of published economic evaluations.
2. A report of a model developed by the company to estimate the cost-effectiveness of GCis + N compared with GCis, GCarbo, PCarbo and DCis for previously untreated patients with locally advanced or metastatic squamous NSCLC eligible for first-line treatment.

Ten papers were identified from the review of economic evaluations, but none were considered suitable for the NICE decision problem. After completion of the systematic review, a US

economic evaluation based on the SQUIRE trial results was identified. The company expressed concern about the applicability of this study in the UK, and did not discuss or critique it further.

The CS reported an economic evaluation conducted for the appraisal, based on a de novo model. This was generally consistent with the specified decision problem and with the NICE reference case. A revised version of the model submitted during the assessment focussed on patients with EGFR expressing tumours, in line with the marketing authorisation. The analysis excluded some comparators specified in the scope: vinorelbine combinations and DCarbo because HR estimates were not available from the company's NMA; and PCis as the company argued that it is infrequently used in practice. Some utility values used in the model did not conform to NICE's preferred methods for the measurement and valuation of health-related quality of life.

The model structure reflected the process of treatment and disease progression for a cohort of patients starting first-line induction treatment with GCis + N or conventional chemotherapy. It was a Markov-type model, with five health states: three prior to progression, for patients on induction treatment, maintenance treatment and off treatment; a post-progression state and death. The model used a one week time step. After completion of induction treatment with GCis + N, patients were assumed to proceed to maintenance treatment with necitumumab alone. Induction and maintenance treatment could terminate at any time due to adverse events or patient choice, disease progression or death. AEs were not modelled explicitly, but costs and effects associated with common AEs (>2.5% of patients and febrile neutropenia) were estimated. Similarly, second line treatments and palliative care were not modelled explicitly, but costs were included for proportions of patients after disease progression.

Rates of treatment discontinuation, disease progression and mortality for GCis + N and GCis were based on data from SQUIRE. PFS and OS for the other comparators were modelled using HRs from the NMA, relative to the survival curves for GCis + N. SQUIRE provided Kaplan-Meier PFS and OS estimates for up to three years following randomisation, and parametric survival functions were then used to extrapolate to the end of the model horizon (lifetime). Alternative functional forms were considered for extrapolation of PFS and OS. The company argued that the best approximations the Kaplan-Meier curves were provided by log-logistic survival functions, fitted separately for the two treatment groups. But to make use of the NMA results for the indirect comparisons, the proportional hazards assumption is necessary.

The company concluded that, of the proportional hazards survival functions tested, Weibull provided the best fit. There was no need to extrapolate estimates for Time to Treatment Discontinuation (TTD), since nearly all patients in SQUIRE had stopped treatment by the end of follow up. TTD estimates were not available from the NMA, so for the indirect comparisons it was assumed that the HRs for treatment discontinuation would match the HRs for PFS. Data from SQUIRE was used to estimate AE risks for GCis + N and GCis, and for the indirect comparisons it was assumed that the relative risks of AEs would equal those for GCis versus GCis + N from SQUIRE.

In order to calculate QALYs, health-related quality of life values ('utilities') were attached to the pre-progression and post-progression health states, and 'disutilities' to the included AEs. EQ-5D data were collected before progression for SQUIRE trial participants. As the company found no between-group differences in EQ-5D, they pooled data for GCis + N and GCis. A systematic review was used to identify sources for utility in the post - progression health state and for AE disutilities.

The model included costs for drugs used in first-line and second-line treatment, drug administration, disease monitoring and management, treatment of AEs and palliative care. Health care resource use was estimated based on a retrospective medical chart review, and consultation with clinical experts. Unit costs of healthcare items were based on national tariffs and data sources.

The company's preferred analysis was based only on direct evidence from the SQUIRE trial for the Western European subgroup of patients with EGFR expressing tumours: which yielded an Incremental Cost Effectiveness Ratio (ICER) of £57,725 per QALY gained for GCis + N compared with GCis. Including other comparators from the NMA for this same patient group, they cited an ICER of £116,344 for GCis + N compared with PCarbo, which was the the next-best, non-dominated alternative.

These results were based on deterministic versions of the model. The company did conduct a Probabilistic Sensitivity Analysis (PSA), but did not report ICERs from this. However, they did use graphical methods to illustrate the wide uncertainty around the estimated incremental costs and effects, and the low probability that GCis + N would be cost-effective below a willingness-to-pay threshold of around £200,000 per QALY. Deterministic analysis was used to show that the

ICER was most sensitive to estimates of OS, and to a lesser extent to PFS and TTD for GCis+N. The estimated ICER was shown to be much higher for the ITT population with EGFR expressing tumours (a figure of £151,152 per QALY gained was cited by the company, but we believe the correct estimate from the company model to be £110,248). Results were also sensitive to the methods used to extrapolate beyond the Kaplan-Meier OS curves; ICERs were in the region of £80,000 per QALY gained using Weibull or exponential functions, or a five-year time horizon, which effectively cuts off the tail of the survival function.

It should be noted that none of these estimates include a cost for the test of EGFR expression that would be required to comply with the marketing authorisation. Thus in practice, the cost of the GCis + N arm (and hence the ICER) would be rather higher than estimated.

Commentary on the robustness of submitted evidence

Strengths

- The company's searches for the systematic reviews of direct and indirect evidence and the reviews of cost-effectiveness studies and data used appropriate search techniques, although they were out-of-date (having been conducted in August 2015, January 2015 and April/May 2014, respectively). The company appears to have included all relevant phase III RCTs in its systematic review; the ERG's update searches for the systematic review of direct evidence (conducted for the period from January 2015 to February 2016) did not identify any other relevant RCTs, although the searches did identify four conference abstracts relating to the SQUIRE trial published before August 2015 that were not identified in the company's searches.
- The inclusion criteria for the systematic reviews of direct evidence and for the NMA generally reflect NICE's scope and the company's decision problem.
- The company's systematic review of direct evidence included a large phase III trial that provided direct evidence of the efficacy of GCis + N compared with GCis, which is one of the most commonly used platinum doublets in clinical practice. The patients included in the trial are representative of those seen in practice.
- The company has, on the whole, appropriately synthesised the evidence in its systematic review of direct evidence.
- The company's economic model is well designed and appropriate for the decision problem.

- Most model parameters are estimated from best-available evidence. The Kaplan-Meier survival estimates from the SQUIRE trial were based on a large patient population, with long follow up, which was not subject to cross over or other serious sources of bias. Systematic searches were used to identify post-progression and AE utility values. And the costing was very thorough, including a retrospective case note review. Although we are critical of some of the methods of used to analyse AE and EQ-5D data, the resulting parameter estimates appear to be reasonable.
- The model is also well implemented. We identified few errors or inconsistencies, and none that made any sizeable difference to the results.
- The model also provides a good platform for exploring parametric and structural uncertainties, including the patient population and methods for extrapolating survival curves.

Weaknesses and areas of uncertainty

- The risk of systematic error in the company's clinical effectiveness systematic reviews is uncertain. The searches for the NMA review were one year out-of-date and the company made post-hoc exclusions of studies from the NMA, not all of which the ERG agrees with.
- The ERG's quality assessment of the included SQUIRE trial differed to the company's assessment. The ERG identified that HRQoL data reported from SQUIRE in the CS are at risk of selective outcome reporting bias, as a number of analyses of HRQoL detailed in the clinical study report (CSR) provided by the company as part of its submission were not reported in the CS. The ERG also noted that subgroup analyses results by age, ECOG performance status [redacted] were not presented in the CS in line with the pre-specified comparison categories.
- The OS results supplied for the EGFR expressing subgroup do not match those reported for this subgroup in a publicly available Food and Drug Administration (FDA) briefing document about necitumumab. [redacted].
- The ERG considers that that company's argument that the Western European subgroup is a more generalisable population to patients in England than the ITT population is not inadequately justified. The ERG considers the EGFR expressing subgroup to be the most relevant population for this appraisal, as this is in line with the marketing authorisation treatment indication. The ERG did not identify a clinical justification for why

data from particular geographical regions rather than the total trial population would be more relevant to England. Furthermore the company did not find a statistical interaction for efficacy effects by region. Clinical expert advice to the ERG was that the baseline characteristics of patients in the ITT population in the SQUIRE trial are representative of those seen in clinical practice.

- The company does not report what a clinically meaningful change in OS would be in the CS; therefore it is unclear if the OS benefits seen with GCis + N in comparison to GCis are clinically meaningful. Clinical expert advice to the ERG is that the improvement in OS in the EGFR expressing subgroup (the population most relevant to the licensed indication) is clinically meaningful.
- The treatment effect estimates from the NMA networks are highly uncertain, as: it is unclear how similar the studies included in the NMA networks were in terms of length of follow-up; the proportion of patients with an ECOG performance status of 2 differed across the studies, and this might have modified treatment outcomes as the analyses were unadjusted; the NMA included mainly subgroup analyses that were likely to be underpowered; the company appears to have made some inappropriate post-hoc exclusions of studies from the NMA; and, most of the comparisons were based on indirect evidence, so consistency with direct evidence could not be assessed.
- The company only presents cost-effectiveness results for the Western European subgroup. As argued above, we do not believe that this is justified.
- The extrapolations of OS curves beyond the three-year follow-up available from SQUIRE are influential on modelled estimates of QALY gain, and are subject to considerable uncertainty. The company's base case estimates rely on log-logistic curves, which have a long tail compared with Weibull curves. Evidence of goodness of fit is similar for these two functional forms.
- We also question the method of extrapolating from the last observations from the SQUIRE data. This places undue emphasis on the tails of the Kaplan-Meier curves, which are based on very small numbers of patients and so are subject to very wide confidence ranges. In the company preferred base case, this has the effect of separating the tails of the extrapolated curves, increasing the estimates of QALY gains.
- We also question whether the long-term survival predictions from the log-logistic extrapolated curves are realistic for the SQUIRE population: 7% and 1% at five years with GCis + N and GCis, respectively, in the company preferred base case.

- The company presents ICERs estimated from the deterministic version of the model, rather than using the correct approach based on mean incremental costs and mean incremental QALYs estimated from the PSA. The deterministic ICERs are lower for the PSA-based ICERs (due to the skew in QALY estimates as illustrated on the cost-effectiveness scatterplots).
- The company does not present correct incremental analyses, but instead presents pairwise comparisons for GCis + N with other included comparators.

Summary of additional work undertaken by the ERG

The ERG conducted verification checks on the model. We started by reviewing the model structure and formulae to look for errors or inconsistencies; cross-checked the model assumptions and inputs against those reported in the CS, and with the cited data sources (where available); compared that the results and sensitivity analyses reported in the CS and clarification report with model outputs. The model included a rather complex system of interacting input sheets and intermediate calculations, and some complicated macros. We therefore chose to replicate the model in a separate Excel file, to check that the calculations and macros yielded the expected intermediate and final results. We found a small number of minor errors and inconsistencies, none of which led to big changes in the model results.

We then conducted a range of additional analyses to test the robustness of the company model to changes in structural assumptions. This included an alternative 'base case' reflecting our best judgement about the most plausible set of assumptions. We then used this base case to explore other possible scenarios and uncertainties over key parameters. The key changes that we made to the company model in our base case were:

- ITT population with EGFR expressing tumours.
- Indirect comparators included, based on the NMA. Despite uncertainty over the completeness and robustness of the NMA, we believe this to provide the best-available evidence relevant to the specified decision problem.
- We added PCis, which was included in the company NMA but not in the model.
- Kaplan-Meier curves were extrapolated from the point at which the number of patients remaining in each arm had declined to 20 or fewer.
- Weibull curves were used for the extrapolations of PFS and OS.

- Results of the PSA were used to calculate ICERs for our base case and all scenario and sensitivity analyses.

This resulted in an estimated ICER of £169,612 per QALY gained for GCis + N compared with GCis (which was the next-best, non-dominated comparator in the incremental analysis). We note that the probabilistic version of the company model including indirect comparisons yielded an ICER of similar magnitude for the ITT (EGFR-expressing) population: £154,024 compared with GCis and £189,779 compared with PCarbo (the best-best, non-dominated option in this case). In our version of the model, the estimated probability that GCis + N would be the most cost-effective treatment option was near to zero below cost-effectiveness thresholds of £100,000 per QALY.

We conducted 16 scenario or sensitivity analyses, selected as those that had proved to be influential in the company analyses and to explore other uncertainties that we had. These analyses further highlighted the sensitivity of results to the way in which OS was extrapolated beyond the Kaplan-Meier data, and the absolute levels of OS for GCis + N and GCis. Results were also somewhat sensitive to PFS and time to discontinuation of GCis + N. However, in all cases the estimated ICER remained high: above £100,000 per QALY except for the most optimistic scenario that we tested, using log-logistic curves for GCis + N and Weibull for GCis, which maximises the separate between the tails of the two curves, and gave an ICER of £84,188 per QALY gained.

As with the company reported ICER estimates, our estimates do not include the cost of a test for EGFR expression that would be required to meet the marketing authorisation. This would further increase the estimated ICERs.

1 Introduction to ERG Report

This report is a critique of the company's submission (CS) to NICE from Eli Lilly and Company on the clinical effectiveness and cost effectiveness of necitumumab for untreated advanced, metastatic, squamous non-small-cell lung cancer (NSCLC). It identifies the strengths and weaknesses of the CS. A clinical expert was consulted to advise the ERG and to help inform this review.

Clarification on some aspects of the CS was requested from the company by NICE and the ERG on 17th February 2016. A response from the company via NICE was received by the ERG on 4th March 2016 and this can be seen in the NICE committee papers for this appraisal.

2 BACKGROUND

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

The ERG considers that the CS provides a clear and accurate overview of the prevalence, cause and prognosis of both lung cancer and, more specifically, squamous NSCLC, as well as the impact of lung cancer on patients and society (CS p. 31 to p. 34). A clinical expert consulted by the ERG advised that most patients with squamous NSCLC are referred to secondary care by their general practitioner (GP) through the two week wait referrals pathway, and in line with the company's statement in the CS, median survival in this patient population is poor (typically less than one year).

The ERG notes that the company has provided limited background information about epidermal growth factor receptor (EGFR) expression in squamous NSCLC. This is an important consideration in this appraisal, as the summary of product characteristics (SmPC)[2] states that necitumumab is indicated for patients with locally advanced or metastatic EGFR expressing squamous NSCLC. The CS details that around 82% to 95% of patients with squamous NSCLC have tumours with EGFR protein expression, with the 82% estimate referring to the proportion with intermediate to high EGFR expression (CS p. 26). The ERG notes that the company has not discussed in the CS how tumour overexpression of the EGFR protein is related to patient prognosis either generally or when treated with an anti-EGFR monoclonal antibody, such as

necitumumab. The clinical expert consulted by the ERG stated that there is currently no reliable evidence linking EGFR expression to drug efficacy generally.

2.2 Critique of company's overview of current service provision

The CS provides a generally clear and accurate overview of how squamous NSCLC is currently managed in clinical practice. As is noted on CS p. 34, NICE clinical guideline (CG) 121[3] provides recommendations for good practice in the management of lung cancer in England. There is currently no guidance specific to the management of squamous NSCLC. The CS correctly notes that NICE CG 121[3] recommends chemotherapy for patients with stage III or IV NSCLC who have a good performance status using a platinum doublet. As stated in the CS, CG 121[3] recommends that either cisplatin or carboplatin is combined with one of the following third-generation drugs: docetaxel, gemcitabine, vinorelbine or paclitaxel. The CS also correctly notes that a single third-generation drug may be used in patients who are unable to tolerate a platinum doublet. In line with the clinical pathway presented in CS Figure 1 (p. 36), the clinical expert consulted by the ERG stated that patients with stages IIIB or IV disease receive first-line treatment with chemotherapy. The ERG's expert stated that patients with stage IIIA disease can receive radiotherapy or chemoradiotherapy, with some receiving chemotherapy. As noted in the CS (p. 35), NICE CG 121[3] emphasises that the aim of chemotherapy treatment is to control the patient's symptoms, to improve their quality of life and to extend their life. The clinical expert consulted by the ERG concurred, stating that chemotherapy can provide palliation and symptom control (if the patient is sufficiently fit to tolerate toxicity) and is given for quality of life reasons, so it is important to know if it is working at an early stage. The expert emphasised that quality of life is a key consideration when treating this patient population.

The ERG notes that the SmPCs for docetaxel,[4] gemcitabine[5] and paclitaxel[6] state that these drugs are to be administered in combination with cisplatin for treating NSCLC. The clinical expert consulted by the ERG indicated, however, that in clinical practice, each of these third generation drugs is used in combination with either cisplatin or carboplatin. The SmPC for vinorelbine[7] states that it can be used with either cisplatin or carboplatin for combination treatment of NSCLC. The expert indicated that carboplatin is quicker to administer and has fewer side effects than cisplatin.

The clinical expert consulted by the ERG stated that in clinical practice, cisplatin in combination with gemcitabine (GCis) or carboplatin in combination with gemcitabine (GCarbo) are the most commonly used platinum doublets. This concurs with the company's statement on CS p. 37 that gemcitabine is the most commonly used first-line treatment for squamous NSCLC in the UK and the company's statement in the decision problem (CS Table 1, p. 15) that GCis and GCarbo are the current standard of care in the National Health Service (NHS). The clinical expert consulted by the ERG stated that all the platinum doublet combinations are equally efficacious, therefore all the combinations are used in practice and all are the current standard of care. The choice of which to use is usually governed by expectations of what patients will be able to tolerate and their quality of life.

The CS (p. 29 and p. 35) states that patients receive chemotherapy for four to six cycles, but does not state the cycle length. Clinical expert advice to the ERG is that patients receive chemotherapy in three-week cycles. Patients undergo two cycles and then have a scan to check that the treatment is working. If it is, they then receive another two cycles of treatment. A full course of treatment takes 12 to 18 weeks (i.e. patients receive four to six cycles in 12 to 18 weeks). In line with the CS, the clinical expert advised that patients may receive between four to six cycles. Therefore, as acknowledged on CS p. 30, the introduction of necitumumab, which will require up to six cycles of treatment in the induction phase (mean 4.6 cycles in the SQUIRE trial, CS Table 6 p. 29) and then maintenance treatment (mean 6 cycles in the SQUIRE trial,[8] CS Table 6 p. 29) (please see section 2.3 below for a description of the induction and maintenance treatment phases), will be associated with additional costs to the NHS, including up to an extra two cycles of treatment in the induction phase.

The necitumumab SmPC states that it is indicated for patients who have epidermal growth factor (EGFR) expressing squamous NSCLC. The company has not, however, discussed in the CS current clinical practice regarding testing patients for EGFR expression nor how the introduction of necitumumab might impact on service provision regarding this. The cost of testing for EGFR expression was not included in the company's cost-effectiveness analyses. The ERG's clinical expert advised that patients are not currently routinely tested for EGFR expression. They are only currently tested for mutations in the EGFR gene. Patients would need to be tested for EGFR expression prior to administration of necitumumab and this would be a new test. The ERG's clinical expert commented that it is unclear how the costs of this would be funded.

As part of the submitted economic model for this appraisal of necitumumab, the company has included the costs of second-line treatment (with either docetaxel or erlotinib). In the overview of current service provision in the CS, the company has stated that patients with NSCLC receive second-line treatment with either docetaxel or erlotinib, as recommended in NICE CG 121[3] and Technology Assessment (TA) 162,[9] respectively. The ERG concurs with the company that docetaxel is recommended for second-line treatment in CG 121.[3] The CS states that TA 162[9] recommends erlotinib for all patients with NSCLC. The ERG notes, however that TA 162[9] has been updated and replaced by TA 374.[10] TA 374 recommends erlotinib as a second-line treatment only in patients who test positive for the EGFR-tyrosine kinase (TK) mutation and who have had non-targeted chemotherapy due to a delay in confirmation of mutation status, or where a patient's EGFR-TK status is unknown, under particular circumstances. The clinical expert consulted by the ERG stated that in clinical practice patients tend to receive second-line treatment with docetaxel given in six doses over three weeks or nivolumab in the context of a clinical trial given indefinitely [nivolumab is licensed but not funded; it is currently undergoing two separate NICE technology appraisals for the treatment of patients with metastatic squamous NSCLC (ID811) and locally advanced or metastatic non-squamous NSCLC (ID900) NSCLC]. The ERG additionally notes that ramucirumab is currently being appraised by NICE as a second-line treatment for patients with metastatic NSCLC (ID838). The ERG's clinical expert estimated that around 40% of patients who receive first-line treatment receive second-line treatment.

2.3 Critique of the company's definition of the decision problem

Population

The population specified in the company's decision problem is people with "locally advanced/metastatic (stage IV)" (CS Table 1, p. 15) squamous NSCLC who have not received prior chemotherapy for this condition. The patient population matches the final scope issued by NICE and is in line with the SmPC indication for necitumumab, in that it is indicated for patients with locally advanced or metastatic disease who have not received prior chemotherapy. Where the population specified by the company does not fully match the SmPC, is that the SmPC more specifically states that necitumumab is indicated for patients who have EGFR expressing squamous NSCLC (as discussed above). The company acknowledges on CS p. 15 that the

population specified in the decision problem is not fully consistent with the SmPC indication, but does not explain why. The ERG therefore believes that the population specified in the decision problem is not appropriate for the potential use of necitumumab in the NHS and that the most appropriate population would be people with locally advanced or metastatic EGFR expressing squamous NSCLC. The company provided clinical effectiveness and cost-effectiveness results for subgroups of patients with EGFR expressing tumours in response to clarification questions from NICE and the ERG (please see discussion under Subgroups below) to reflect the SmPC indication (clarification response A1).

The ERG notes that the Food and Drug Administration (FDA) has approved necitumumab in combination with GCis (GCis + N) for the first-line treatment of metastatic squamous NSCLC, but the FDA has not limited the indication to patients with EGFR expressing squamous NSCLC nor specified locally advanced NSCLC.[11]

As mentioned above, the patient population specified by the company matches the SmPC indication for necitumumab in terms of patients' prior treatment (patients who have not received prior chemotherapy). The final scope specifies that the population should be those "untreated" for advanced, metastatic disease. While the company has more specifically stated that the population is those who have "not received prior chemotherapy", the ERG's clinical expert advised that clinically this is the same as "untreated advanced" disease. The ERG's expert advised that some people may have had resected or irradiated cancer before chemotherapy, but this is essentially the same as presenting with untreated metastatic disease.

Intervention

In accordance with the final scope, the intervention described in the decision problem is GCis + N (necitumumab's brand name is Portrazza). Necitumumab is a monoclonal antibody that works by targeting EGFR-1. In December 2015, the Committee for Medicinal Products for Human Use (CHMP) recommended the granting of a marketing authorisation for necitumumab, and this has now been granted. As outlined in the CS, the SmPC recommends that necitumumab is given to patients at a flat dose of 800 mg via intravenous infusion over 60 minutes on days one and eight of each 3-week chemotherapy cycle, for up to six cycles. The company states that a gemcitabine dose of 1250 mg/m^2 is to be administered through intravenous infusion on days one and eight of each cycle, with a cisplatin dose of 75mg/m^2 administered on day one of each cycle. The ERG notes that these stated doses of gemcitabine and cisplatin match those

specified in the gemcitabine SmPC[5] for combination therapy for NSCLC. The ERG further notes the gemcitabine SmPC[5] states that gemcitabine should be given as a 30-minute intravenous infusion. Cisplatin is also given by intravenous infusion.[12] The CS states that following induction combination therapy, patients who have not experienced disease progression receive necitumumab monotherapy at a flat dose of 800 mg on days one and eight of each three-week cycle until the patients experience disease progression or unacceptable toxicity. The ERG notes that this matches the SmPC. Overall, the intervention described in the decision problem is appropriate for the NHS.

Comparators

The CS decision problem includes all eight platinum doublets that are currently used in the NHS and which were specified in the final scope (i.e. carboplatin or cisplatin in combination with gemcitabine, docetaxel, paclitaxel or vinorelbine). In the economic analysis, however, the company has included (as outlined on CS p. 162):

- GCis,
- GCarbo,
- Carbopatin in combination with paclitaxel (PCarbo), and
- Cisplatin in combination with docetaxel (DCis).

The company did not identify any relevant clinical evidence to be able to include carboplatin in combination with docetaxel (DCarbo) or carboplatin in combination with vinorelbine (VCarbo). The evidence identified for assessing the comparative efficacy of cisplatin in combination with vinorelbine (VCis) was unsuitable for use in the model and so VCis is also not included in the model. The ERG considers the company's justification for not including these comparators is reasonable. Although data were available for cisplatin in combination with paclitaxel (PCis), the company excluded this from their economic model. In response to a clarification question, the company stated that this was due to its infrequency of use in UK practice (assumed that less than [redacted] of patients receive PCis, based on market share data). The clinical expert consulted by the ERG confirmed that PCis is not widely used in the UK and that, in practice, its use is limited to clinical trials. Nevertheless, PCis was included in the scope, so the ERG believes it should have been included in the company model. Table 1 summarises the comparators specified in the scope and those included in the economic model.

Overall, the comparators specified in the decision problem are appropriate for the NHS.

Table 1 The eight comparators specified in NICE's final scope and those included in the economic model

Scope specified comparator	Comparator included in the company's economic model (✓ indicates 'yes')
GCis	✓
GCarbo	✓
DCis	✓
DCarbo	
PCis	
PCarbo	✓
VCis	
VCarbo	

GCis, cisplatin in combination with gemcitabine; GCarbo, carboplatin in combination with gemcitabine; DCis, cisplatin in combination with docetaxel; DCarbo, carboplatin in combination with docetaxel; PCis, cisplatin in combination with paclitaxel; PCarbo, carboplatin in combination with paclitaxel; VCis, cisplatin in combination with vinorelbine; VCarbo, carboplatin in combination with vinorelbine.

Outcomes

The company has listed all the outcomes specified in the final scope in their decision problem:

- Overall survival (OS)
- Progression free survival (PFS)
- Response rates
- Adverse events (AEs)
- Health-related quality of life (HRQoL)

These outcomes are appropriate and clinically meaningful to patients. The ERG considers that the company has included all important outcomes in the decision problem. Given that the ERG's clinical expert advised that treatment of this patient population is palliative with a focus on patients' quality of life, the ERG suggests that HRQoL is a particularly clinically important outcome. Clinical expert advice to the ERG is that the aim of chemotherapy is to improve or maintain quality of life.

Economic analysis

The economic analysis specified in the decision problem largely matches the final scope and is appropriate for the NHS. The company have conducted a cost-utility analysis with a lifetime horizon. This is an appropriate time horizon when considering differences in costs and outcomes between treatments for patients with squamous NSCLC. Utility estimates for the main health states in the model are based on EQ-5D data from patients, valued by a representative sample of the UK population (UK tariff). However, disutility estimates for adverse events are derived from patients by direct valuation (standard gamble). Costs are considered from the NHS

and Personal Social Services perspective. Discount rates of 3.5% per year are applied to health outcomes (QALYs) and costs.

[redacted].

Other relevant factors

Subgroups

The final scope does not specify any patient subgroups for examination in this appraisal and the company has not specified any in their decision problem in the CS. The company has, however, argued that a post-hoc subgroup of patients from Western Europe in the SQUIRE trial is a more generalisable population to patients in England than the ITT population (all randomised patients). The company provided trial results for both the ITT population and the Western Europe subgroup in the CS but based efficacy and cost-effectiveness conclusions on the results of the Western European subgroup analyses. The Western European subgroup included patients from Germany, France, Spain, Greece, Italy, UK, Portugal, Austria and Belgium.

In the CS, the company state in their decision problem (CS Table 1 p. 15) that additional analysis would be provided to NICE at a later stage to reflect the SmPC population. In response to NICE and the ERG's clarification request about this (clarification response A1), the company provided clinical effectiveness results for two further post-hoc subgroups:

1. Patients with EGFR expressing tumours from the total SQUIRE trial population
2. Patients with EGFR expressing tumours from the Western Europe subgroup of the SQUIRE trial

The company supplied additional cost-effectiveness analyses and a revised economic model, which used data from the EGFR expressing tumours Western European subgroup as the base case. The four groups of patients included in the company's submission and clarification response are summarised in Table 2.

Table 2 Summary of populations and subgroups

Population	Source	Used in company's analyses	Used in ERG's analyses
ITT population All randomised patients in SQUIRE trial	Original CS	Clinical effectiveness only	Clinical effectiveness only
Western European subgroup	Original CS	Clinical effectiveness Base case in original CS	Clinical effectiveness subgroup analysis only
EGFR expressing subgroup of ITT population (the licensed indication)	Clarification response	Clinical effectiveness only	Clinical effectiveness **ERG's preferred base case**
EGFR expressing Western European subgroup	Clarification response	Clinical effectiveness **Base case in updated analysis**	ERG's scenario analysis

CS, company submission.; EGFR, epidermal growth factor receptor; ERG, evidence review group; ITT, Intention-to-treat.

The company does not provide a clear rationale for why patients from countries in Western Europe are considered to be the most generalisable to patients in England. The ERG notes that the SQUIRE trial included patients from other countries, such as Australia, the US and Canada, who might also be considered similar to the patient population in England, and who could have been included in the relevant subgroup, but the company has not discussed why these countries have not been included. The company has also not included Eastern European countries (such as Hungary and Poland) in the subgroup considered generalisable to England. The company states in the CS that patients in Hungary and Poland performed better in the GCis than the GCis + N arm in the trial, and that there were no differences between arms in patient demographics, characteristics, prognostic factors or treatment received that explained this difference (CS p. 21 to p. 22). Instead the company suggests this finding may be due to "unobserved treatment effect modifiers" (CS p. 22), including the disease burden of squamous NSCLC and environmental causes of the disease, including heavy smoking. The ERG considers this explanation unconvincing, since these factors would likely equally affect both arms in the trial due to patient randomisation and no rationale is given as to why these factors would result in worse outcomes with necitumumab.

NICE and the ERG sought further clarification from the company on the rationale for the choice of countries included in the Western Europe subgroup, and asked why other countries such as Australia and Canada had not been included. In response, the company stated that all countries in Europe that were not included in the pre-specified Eastern Europe subgroup were included in

the post-hoc Western Europe subgroup, and that Australia and Canada were not included as they are not part of Europe. The company also stated that it is believed that the Western Europe subgroup is more generalisable to clinical practice in England than the populations across Australia, Canada and Europe combined (clarification response A6), however no additional information was provided.

Clinical expert advice to the ERG is that data from patients from all geographical regions would be representative of patients in England, with perhaps the exception of Asia (8% of the ITT population). Patients in Asia have a higher frequency of EGFR mutations, which would make an EGFR receptor drug more effective. The ERG also notes that the company stated that there was not a statistically significant treatment interaction between the post-hoc Western Europe subgroup and other patients in the SQUIRE trial (CS p. 229). Overall, the ERG considers that the company's use of the Western Europe subgroup in the base case is not sufficiently justified. The ERG considers the subgroup of patients with EGFR expressing tumours from the ITT population is the most relevant patient group to the marketing authorisation and to patients in England.

On CS pp. 68 to 69, the company additionally lists a number of planned subgroup analyses by geographical region and countries with an enrolment >40 patients, but has not provided the results of these in the CS. These were requested by NICE and the ERG, and while subgroup analyses by region were provided in clarification response A6c Appendix 6, the regions analysed differed to those pre-specified.

The company also provides details of other planned subgroup analyses on CS p. 69, including:

- age (<70 versus ≥70 years; and <65 versus ≥65 years);
- gender (women versus men);
- race (White versus non-White);
- ECOG PS (0 versus 1 versus 2 and 0-1 versus 2); and,
- smoking history [never smoker (non-smoker and light ex-smoker combined) versus smoker].

CS Table 11 p. 51 also states that patients who displayed a rash within the first cycle was a pre-specified subgroup, however results are not presented in the CS.

Results of the subgroup analyses by age, gender, race, ECOG performance status and smoking history are provided in CS Figure 9 (p. 70) for the ITT population only. However, the presentation of the results for the age and ECOG performance status analyses is not entirely in line with the pre-specified comparison categories. Clinical expert advice to the ERG is that EGFR receptor drugs are more efficacious in women, people of an Asian ethnicity and smokers. The ERG's expert also advised that a patient's performance status can impact treatment efficacy. The ERG therefore considers that while these subgroup analyses are appropriate, the deviation in how the results are presented from those pre-specified means the results may be at risk of selective reporting bias. The results of the subgroup analyses by age and performance status are not, however, used in the company's economic model.

The CS also presents pre-specified subgroup analyses of OS and PFS by EGFR expression status, classified by immunohistochemistry score (H-score) on a scale of 0-300[1] (H-score of <200 and H-score ≥200) for the ITT population (CS pp. 71 to 73). The CS does not provide a rationale for using an H-score of 200 as the cut-off, although the trial publication[1] refers to a previously reported study, the FLEX trial of cetuximab in NSCLC.[13] An FDA Briefing Document[14] about necitumumab identified by the ERG notes that the cutpoint value of 200 was chosen based on a post-hoc subgroup analysis of the FLEX study, in which patients with NSCLC who had an EGFR H-score >200 experienced greater improvement in OS with cetuximab compared to patients with an H-score <200.

The FDA Briefing Document states that additional analyses were undertaken to evaluate all patients with EGFR expressing squamous NSCLC together (H-score > 0) and those with no detectable EGFR expression (H-score=0, where H-score=0 is defined as 100% of cells with undetectable EGFR staining).[14] Results of these analyses are not reported in the CS but were provided in the company's clarification response. The ERG has reproduced the results from the FDA briefing in section 3.3, along with the results provided by the company in the clarification response. This is because the results the company provided for the EGFR expressing subgroup differ slightly to those reported in the FDA document. The company only provided a brief comment on the results for the subgroup of patients with no detectable EGFR expression without provding supporting data, although these data are available in the FDA document.

The ERG and the clinical expert consulted by the ERG did not identify any other key subgroups that should have been considered.

Equality issues

The final scope does not identify any equity or equality issues related to the implementation of GCis + N in the NHS and the company has not specified any in its decision problem. The ERG and the ERG's clinical expert have also not identified any equity or equality issues.

3 CLINICAL EFFECTIVENESS

3.1 Critique of company's approach to systematic review

3.1.1 Description of company's search strategy

The ERG considers that the searches for the main systematic review of direct evidence, the systematic review informing the network meta-analysis (NMA), and the reviews of cost-effectiveness studies and data were appropriate. There was one minor typographical error and slight inconsistencies in approach across the searches, but the ERG considers that these would not impact the results. As the main systematic review and the cost-effectiveness searches were out-of-date (conducted in August 2015 and April/May 2014, respectively), the ERG ran update searches for these relating to necitumumab on the following databases: Embase, Medline and Medline in Process and other Indexed Citations via the Ovid Platform. The search for the review to inform the NMA was also out-of-date, having been conducted in January 2015, but the ERG did not elect to update these searches. No new trials or cost-effectiveness publications were identified. The ERG's searches (conducted from January 2015 to February 2016) for the main systematic review, however, found four conference abstracts reporting analyses appertaining to the SQUIRE trial, which were not identified in the company's searches. These were screened by two ERG reviewers who considered that two of the abstracts[15 16] were relevant to the appraisal and met the company's main systematic review inclusion criteria (please see section 3.1.3 of this report for details). Clinicaltrials.gov is documented in the CS as searched for recently completed and not yet published studies. The ERG elected to widen the ongoing study searches to incorporate UK Clinical Research Network (UKCRN), World Health Organisation (WHO) International Clinical Trials Registry Platform (ICTRP), ISRCTN, and clinicaltrials.gov. The ERG search results were screened by one reviewer. The results yielded two relevant trials of GCis + N (please see section 3.1.3 for details).

3.1.2 Statement of the inclusion/exclusion criteria used in the study selection

The company clearly states the inclusion and exclusion criteria for both the main systematic review of studies evaluating GCis + N (in CS Table 9, p. 39) and the systematic review underpinning the NMA (in CS Table 21, p. 85). The ERG's critique of the eligibility criteria used in the review for the NMA and details about the studies identified for inclusion are provide section 3.1.7.

The main systematic review of direct evidence included trials of first-line treatment with GCis + N for patients with locally advanced or metastatic squamous NSCLC who were naïve to treatment, compared with a platinum doublet (i.e. GCarbo, GCis, DCarbo, DCis, PCarbo, PCis, VCarbo or VCis). Trials had to assess OS, PFS, response rates, HRQoL or safety to be included. Inclusion was limited to Phase III and IV randomised controlled trials (RCTs) and English language references. The company did not specify treatment setting as an inclusion criterion nor place any limits on inclusion relating to the quality of the RCTs, which is appropriate. The inclusion criteria reflect the decision problem, the licensed indication for necitumumab (although drug doses are not specified in the criteria), current service provision and the potential use of GCis + N in the NHS. Overall, the ERG considers the inclusion criteria reasonable, but suggests that the company could have considered including phase II RCTs for efficacy and safety data. The company, however, did not restrict study eligibility for inclusion in the systematic review that informed the NMA by RCT phase and so any relevant phase II RCTs are likely to have been identified by the review for NMA (the review identified none; please see section 3.1.7).

The CS includes a flow diagram showing the number of studies included and excluded at stage of the main systematic review (CS Figure 2, p. 41). The flowchart for the main syste review does not provide reasons for the exclusion of six publications at the full text screer stage of the main review; however these were provided in clarification response A2 and the exclusions appear justified.

Overall, the ERG considers that the eligibility criteria used in the main systematic review were appropriate and matched the company's decision problem.

3.1.3 Identified studies

The main systematic review identified one relevant Phase III RCT of GCis + N – the SQUIRE trial[1] (shown in Table 3) – reported in one publication. In the CS, the company has also referred to supplementary information, in addition to the primary publication. The company did not identify any non-RCTs, as they restricted inclusion to RCTs only. Details of the studies identified in the systematic review underpinning the NMA are provided in section 3.1.7.

Table 3 Details of the included SQUIRE RCT[1]

Design, patient population and legth of follow-up	Intervention	Comparator
Design: Phase III, open-label, multicentre RCT carried out in 26 countries, including the UK [redacted] *Patient population:* Adults with stage IV squamous NSCLC, who had not received previous chemotherapy for advanced NSCLC. ECOG PS 0-2. N=1093 (545 GCis + N; 548 GCis). *Median length of follow-up:* GCis + N arm: 25.2 months; GCis arm: 24.8 months.	A maximum of six 3-week cycles of gemcitabine 1250 mg/m^2 (administered intravenously over 30 min on days 1 and 8 of each cycle) and cisplatin 75 mg/m^2 (administered intravenously over 120 min on day 1), plus necitumumab 800 mg (administered intravenously on days 1 and 8 over a minimum of 50 min). At the end of chemotherapy, patients who had not experienced disease progression received necitumumab alone as a maintenance therapy until disease progression, AEs leading to discontinuation, or consent withdrawal.	A maximum of six 3-week cycles of gemcitabine 1250 mg/m^2 (administered intravenously over 30 min on days 1 and 8 of each cycle) and cisplatin 75 mg/m^2 (administered intravenously over 120 min on day 1).

ECOG, Eastern Cooperative Oncology Group; GCis, gemcitabine plus cisplatin; GCis + N, Necitumumab with gemcitabine plus cisplatin; NSCLC, non-small-cell lung cancer; RCT, randomised controlled trial; PS, performance status.

The company supplied the ERG with electronic copies of the SQUIRE trial primary publication[1] and the clinical study report (CSR). The trial was sponsored by Eli Lilly and Company.

The ERG agrees that the SQUIRE trial meets the systematic review inclusion criteria and is relevant to the final scope and the company's decision problem. The trial only included patients with metastatic (stage IV) squamous NSCLC, so no data on the efficacy or safety of GCis + N were available in the company's systematic review for people with locally advanced (stage III) disease. The population is therefore narrower than that outlined in the scope and included in the

SmPC indication (GCis + N is indicated for patients with stage III and IV disease). The ERG notes that the trial used the drug doses and regimens outlined in the necitumumab draft SmPC and the gemcitabine SmPC,[5] except that necitumumab was delivered for a minimum of 50 minutes, while the draft SmPC states it should be delivered for a minimum of one hour.

The trial patient population, however, is wider than the licensed indication, as the SQUIRE trial was not limited to patients with EGFR expressing NSCLC, which is the licensed indication. The CS provides subgroup analyses of OS and PFS according to whether patients had high or low expressing EGFR tumours (defined as H-scores of ≥200 and <200, respectively), but does not provide a combined subgroup analysis of all patients with EGFR expressing NSCLC compared with patients without EGFR expressing NSCLC. The ERG identified results from this analysis in a FDA Briefing Document[14] and has presented these findings in section 3.3. As noted above, the company provided subgroup analysis results for patients with EGFR expressing tumours in the SQUIRE trial, in its response to NICE and the ERG's clarification questions (clarification response A1 Appendix 1). The company also provided a comment in clarification response A7 that a subgroup analysis of patients without EGFR expression (H-score = 0) was carried out but did not present data.

The CS provides an overview of the SQUIRE trial design and interventions used (CS p. 42 and p. 45 to 46). The patient inclusion and exclusion criteria are provided on CS p. 42 to 43. In clarification response A8 the company stated that the SQUIRE exclusion critieria incorporated prior anticancer therapy with monoclonal antibodies, signal transduction inhibitors, or any therapies targeting the EGFR, vascular endothelial growth factor (VEGF), or VEGF receptor; or previous chemotherapy for advanced NSCLC (patients who had received adjuvant chemotherapy were eligible if the last administration of the prior adjuvant regimen occurred at least one year prior to randomisation). Clinical expert advice to the ERG is that it was reasonable for the SQUIRE trial to include patients who had received adjuvant chemotherapy. Baseline characteristics for both the ITT and Western European populations are reported in the CS (Table 13 on CS p. 58 and p. 59). Race characteristics are missing for the Western European population and were requested by NICE and the ERG. These are provided in clarification response A6b Appendix 5, but are for the EGFR expressing Western European subgroup only and not the whole Western European subgroup from the ITT population. The data provided showed that race characteristics were balanced across arms.

The CS provides a CONSORT flowchart (Figure 3, p. 55) showing the number of patients randomised, treated and the number who completed each stage of the trial or discontinued. The number of patients eligible for the trial is not reported. [redacted] There appears to have been no participant cross-over (treatment switching) in the trial. The CS details the primary and secondary outcomes assessed (CS p. 52 and p. 53, including the definitions of each outcome and the HRQoL measures used. The sample size and power calculation (for the primary outcome of OS) are provided on CS p. 53, subgroup analyses are detailed on CS p. 49 to p.50 and the ITT population is defined on CS p. 18. The CS details how the statistical analyses of the primary and secondary outcomes were performed on pp. 52 to 53, except how HRQoL data were analysed.

Baseline characteristics

The CS states that baseline patient characteristics were similar across treatment arms within the SQUIRE trial in the ITT population (CS p. 57), and the ERG agrees with this conclusion. The ERG also agrees with the company that there are differences in age and ECOG performance status between the trial arms in the Western Europe population. As the CS notes, the proportion of patients aged ≥ 70 years was higher in the GCis + N arm than in the GCis arm (23% versus 14%). The GCis arm had a higher proportion of patients with an ECOG performance status of 2 than the GCis + N arm (7% versus 2%). Clinical expert advice to the ERG is that performance status can affect treatment outcomes and that the fittest patients are best suited to treatment with GCis. The ERG suggests that the performance status differences between trial arms may have marginally favoured the GCis + N arm in the post-hoc Western European subgroup analyses. Clinical expert advice to the ERG is that prognosis is determined by disease stage and ECOG performance status more than by age.

The CS does not report on the proportion of patients who had EGFR expressing tumours at baseline, but provided these data in clarification response A7. The company clarified that of the patients in the SQUIRE trial with tumour samples available (90%), 95% had tumours expressing EGFR protein. The company also provided patients' baseline characteristics for the main EGFR expressing subgroup and the EGFR expressing Western Europe subgroup in their clarification response A1 Appendix 1. The ERG notes that baseline characteristics were similar between treatment arms in the main EGFR expressing subgroup. In the EGFR expressing Western Europe subgroup (similarly to the overall Western Europe subgroup) proportionally fewer patients in the GCis + N arm were aged ≥18 - <65 years [redacted] than in the GCis arm

[redacted]and proportionally fewer patients in the GCis + N arm had an ECOG performance status of two[redacted] than in the GCis arm[redacted].

The CS states that the patient baseline characteristics in the SQUIRE trial were representative of patients with advanced, squamous NSCLC, and the ERG agrees. The clinical expert consulted by the ERG stated that the baseline characteristics for both the ITT and Western Europe populations are broadly representative of patients seen in practice in England. The SQUIRE trial included fewer patients with an ECOG performance status of 2 than 0 or 1, and the ERG's clinical expert advised that this reflects the patient population treated in practice.

ERG's appraisal of whether all relevant studies were included in the review

The CS appears to have included all relevant Phase III RCTs. The ERG's searches did not identify any other relevant studies, but did identify four conference abstracts[15-18] reporting results from the SQUIRE trial that were not included in the company's systematic review. All these abstracts were published between November 2014 and May 2015. The company did not appear to find these publications during their searches (which were undertaken in August 2015), as they are not listed among the 34 excluded references listed by the company in their clarifications response. Of the four abstracts identified by the ERG, the ERG considered that two met the company's inclusion criteria for the systematic review:

- One[15] reported on the planned ECOG performance status subgroup analyses, in line with all the pre-planned categories. The ERG has summarised the results from these analyses in section 3.3 of our report.
- One[16] reported on the safety and efficacy of treatment with necitumumab alone during the maintenance phase following treatment with GCis + N; this reported the proportions of patients receiving maintenance treatment, the median OS, and PFS, and two-year survival (these outcomes were not reported in the CS), and adverse events of special interest [reported in the CS, apart from the proportion of patients experiencing venous thrombolic events (2.5%)].

Ongoing studies

The CS lists six ongoing Phase I and II trials of necitumumab for treating squamous and non-squamous NSCLC (CS pp. 136 to 137). The CS provides the trial identifiers and details about the patient populations. All but one of the trials are single arm studies of the safety and efficacy of necitumumab used in combination with other drugs, including standard chemotherapy with

PCarbo and with experimental agents. The trials include patients with either stage IV NSCLC or squamous NSCLC, or, in one trial, patients with EGFR mutation-positive stage IV or recurrent NSCLC who have progressed after previous treatment with an EGFR-TK inhibitor.

The ERG searched for ongoing trials and identified one additional relevant RCT (NCT01763788; not listed among the ongoing studies identified by the company on CS p. 136-137). This is an open-label RCT of GCis + N versus GCis in people with Stage IV squamous NSCLC. The study has two phases: Phase 1b is a dose escalation study (gemcitabine 1000 or 1250 mg/m^2) to determine the recommended dose for the subsequent Phase 2 portion of the study. Phase 2 evaluates efficacy. Estimated enrolment is 189 and study completion date is June 2017. In addition, one single arm, open label, phase II study of GCis + N in people with Stage IV squamous NSCLC was identified (NCT01788566) (also not listed on CS p. 136-137), with a study completion date of December 2015.

Summary

The CS appears to include all relevant RCTs of GCis + N for treating squamous NSCLC; there appears to be only one relevant RCT available (the SQUIRE trial). As the ERG identified two conference abstracts from the ERG's searches that met the company's inclusion criteria for the main systematic review that were not identified in the CS or listed among the studies excluded in the company's clarification response, it is uncertain if the company's searches identified all relevant publications relating to the SQUIRE trial.

3.1.4 Description and critique of the approach to validity assessment

The CS includes a quality assessment of the SQUIRE trial (CS Table 12 p. 56), but not of the trials included in the NMA. In response to a request by NICE and the ERG, the company provided quality assessment for trials included in the NMA (clarification response A21 Appendix 10), and the ERG discusses this further in section 3.1.7.

The company's quality assessment of the SQUIRE trial is presented in tabular format containing detailed factual information, although judgement or discussion on the criteria by the company is limited. The company used the criteria suggested by NICE for quality assessment of the SQUIRE trial, and the ERG agrees with most of the company's assessment (please see Table 4). However, the ERG notes that whilst the ITT population was similar in both trial arms at the

outset of the study in terms of prognostic factors, there were differences in the post hoc Western Europe subgroup (used in the company's base-case analysis) with respect to age group (≥18 to <65 years GCis+N [redacted], GCis [redacted]; ≥70 years, GCis+N [redacted], GCis [redacted]) and ECOG performance status (ECOG performance status 1 GCis+N [redacted], GCis [redacted]; ECOG performance status 2 GCis+N [redacted], GCis [redacted]).

SQUIRE was an open label study, with outcome assessors at Eli Lilly blinded to treatment assignment, except for serious adverse event (SAE) data. In response to a question from NICE and the ERG, the company clarified that the assessment of progressive disease or toxicity to define whether maintenance therapy was given was completed by investigators who were not blinded to treatment allocation (clarification response A9). The trial is therefore at risk of performance bias and detection bias on these measures. The ERG notes, however, that the trial paper states that safety data were assessed by an independent data monitoring committee. [8] The company confirmed that an independent review of the assessment of PFS, ORR and time to treatment failure (TTF) was not conducted (clarification response A10), meaning that the results for these outcomes are also at risk of detection bias. Limited HRQoL data are presented in the CS. The ERG also notes that a number of the analyses and corresponding results of the Lung Cancer Symptom Scale (LCSS) detailed in the CSR are not reported in the CS. Furthermore, subgroup analyses by age, ECOG performance status and region were not presented in the CS and the company's clarification response in accordance with the pre-planned analyses. The ERG therefore considers there to be a risk of bias due to selective outcome reporting in the CS.

Table 4 Company and ERG assessment of trial quality

NICE QA Criteria for RCT	CS response (selected information from CS Table 12 p. 56)	ERG response
1. Was the method used to generate random allocations adequate?	Description given but judgement not provided	Yes
2. Was the allocation adequately concealed?	Description given but judgement not provided	Yes
3. Were the groups similar at the outset of the study in terms of prognostic factors, e.g. severity of disease?	Yes	Yes (ITT population)[a]
4. Were the care providers, participants and outcome assessors blind to treatment allocation? If any of these people were not blinded, what might be the likely impact on the risk of bias (for each outcome)?	Care providers and participants not blinded. Outcome assessors at Eli Lilly blinded to treatment assignment with the	Care providers and participants: no Outcome assessors: no[c]

	exception of SAE data.	
5. Were there any unexpected imbalances in drop-outs between groups? If so, were they explained or adjusted for?	No	No
6. Is there any evidence to suggest that the authors measured more outcomes than they reported?	No	Yes[b]
7. Did the analysis include an intention to treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Primary analyses include all randomised patients following the ITT principle, regardless of compliance with the treatment regimen and protocol.	Yes (OS, PFS, ORR and TTF analyses, but not the HRQoL analyses),[d] yes

[a] For the Western Europe subgroup, there were imbalances in age group (≥18 to <65 years GCis+N [redacted], GCis [redacted]; ≥70 years, GCis+N [redacted], GCis [redacted]) and ECOG performance status (ECOG PS 1 GCis+N [redacted], GCis [redacted]; ECOG performance status 2 GCis+N [redacted], GCis [redacted]).
[b] Limited HRQoL data were presented in the CS. The ERG notes the CSR also reports [redacted]. The ERG also notes that subgroup analyses by patients' age and ECOG performance status presented in the CS and subgroup analyses by region presented in Appendix 6 of the company's response to clarifications questions from NICE and the ERG are not presented in line with the pre-specified analyses.
[c] CS p. 42 states that the company had blinded access to the clinical data provided to it during the trial (except for SAEs), but it is unclear who assessed patient outcomes. Based on information provided in the SQUIRE trial paper, [8] and the company's clarifications question response to NICE and the ERG (clarification responses A9 and A10), outcome assessors (the investigators) were not blinded to treatment allocation for assessments of any outcome (clarification response A9), although safety data were assessed by an independent data monitoring committee.[8]
[d] AE analyses were conducted in the safety population.

3.1.5 Description and critique of company's outcome selection

The NICE scoped outcomes were OS, PFS, response rates, AEs and HRQoL. The outcomes in the decision problem addressed by the company (CS p. 15) are the same as in the NICE scope. The primary outcome in the SQUIRE trial was OS, with secondary outcomes including PFS, objective response rate (ORR), time to treatment failure (TTF), safety and HRQoL. The company used the results from the analyses of the OS, PFS, AE and HRQoL (EQ-5D data) outcomes from the SQUIRE trial in the economic models submitted with the CS and the model submitted in the company's clarifications response.

OS was defined as the time from the date of randomisation to the date of death from any cause, and PFS was defined as the time from randomisation until the first radiographic documentation of objective progression as defined by Response Evaluation Criteria in Solid Tumours (RECIST) Version 1.0, or death from any cause (CS Table 11, p. 50). An independent review of the assessment of PFS, ORR and TTF was not conducted (clarification response A10). For OS, patients who did not die or who were lost to follow-up were censored at the last date they were

known to be alive. For PFS, the CS states that patients were censored from the PFS analysis at the date of their last radiographic tumour assessment if they did not experience disease progression or if they were lost to follow-up. Patients were also censored at the date of their last radiographic assessment if they died or experienced disease progression after two missing assessment visits or if they began using a different cancer treatment before disease progression.

ORR was defined as the proportion of patients achieving a best overall response of confirmed partial or complete response according to RECIST Version 1.0 from the start of treatment until disease progression or recurrence (taking as reference for progressive disease the smallest measurements recorded since the treatment started).

HRQoL was measured using the LCSS and EQ-5D-3L (visual analogue scale and health index score) prior to treatment (within 14 days of randomisation), prior to the first infusion of Cycles 1-6, and every 6 weeks (± 3 days) thereafter (i.e. concurrent with radiological evaluation after discontinuation of chemotherapy) until progressive disease. The LCSS is a self-reported disease and lung cancer specific instrument consisting of nine items including six major lung cancer symptoms and three global measures of symptom distress, activity and quality of life. Each item is assessed with a 100-mm visual analogue scale with higher ratings equating to poorer quality of life. The CS does not define a clinically meaningful difference, however in the CSR p. 63 this is defined as a ≥15 mm change from baseline.

The CS states that the instruments were completed where there was a validated language/cultural translation in a language/culture in which the patient was fluent. Other lung cancer-specific instruments are available, such as EORTC Quality of Life Questionnaire 30 item core instrument (QLQ C30) or FACT-L (Functional Assessment of Cancer Therapy – lung cancer module); however, the CS does not provide a rationale for selecting the LCSS.

The CS presents time to deterioration of LCSS and time to deterioration of ECOG performance status (CS Figure 14 p. 77), but does not provide a definition for these. In the CSR p.63 deterioration in LCSS was defined as a ≥15 mm increase from baseline in LCSS score, but assessment of deterioration of ECOG PS was not defined. Deterioration of ECOG performance status is not a NICE scoped outcome.

The ERG notes that the CSR also reports that a number of other analyses of the LCSS were undertaken, but the results of these analyses are not reported in the CS. This means there is a risk of selective outcome reporting in the data presented in the CS.

The CS also reports TTF (not a NICE scoped outcome), defined as the time from randomisation to the first observation of progressive disease, death due to any cause, early discontinuation of treatment or initiation of new anticancer therapies. As TTF is not a NICE scoped outcome, we do not consider it further in our report.

Treatment-emergent adverse events (TEAEs) were defined as: those with an onset date that occurred any time during or after the administration of the first dose of study treatment or up to 30 days after the last dose of study treatment (or up to any time if serious and related to study treatment); or those that occurred prior to the date of first dose and worsened while on therapy or up to 30 days after the last dose of study treatment (or up to any time if serious and related to study treatment). A serious adverse event (SAE) was defined as any untoward medical occurrence that at any dose: resulted in death; was life-threatening; required inpatient hospitalisation or caused prolongation of existing hospitalisation; resulted in persistent or significant disability/incapacity; was a congenital anomaly/birth defect; required intervention to prevent permanent impairment/damage; and/or was an important medical event (defined as a medical event that may not be immediately life-threatening or result in death or hospitalisation but, based upon appropriate medical and scientific judgment, may jeopardize the patient, or may require intervention to prevent one of the aforementioned serious outcomes).

Overall the ERG considers that the outcomes listed in the NICE scope are appropriately addressed by the CS.

3.1.6 Description and critique of the company's approach to trial statistics

The company confirmed in the clarification response to NICE and the ERG that data for the EGFR expressing subgroup and ITT population in SQUIRE were analysed using the same methods (clarification response A1). Below, we summarise and critique the company's approach to summarising the trial statistics in the CS for each outcome.

Overall survival

The CS reports trial results for OS (CS p. 59-61) presenting the numbers, numbers censored, the stratified log-rank p-value, the stratified HR and 95% CI. The HR for OS was estimated from a stratified Cox proportional hazards model; the stratification factors were ECOG PS (0-1 vs 2) and geographic region (North America, Europe and Australia vs South America, South Africa and India vs Eastern Asia). A Kaplan-Meier (KM) curve for the ITT population is presented in CS figure 4 and outcomes of median OS and survival rates (6-month, 1-year, 18-months, 2-year) are presented together with 95% CIs in CS Table 14.

The trial publication for SQUIRE[1] concurs with the CS that censoring for OS was based on the last date that the patient was known to be alive.

The trial publication for SQUIRE[1] it states that a Cox proportional hazards model was fitted for OS to formally test the proportional hazards assumption, with the following predictors: treatment, and an interaction term of treatment and log of event time.

We note that CS p. 53 reports the power calculation for the trial which was adequately powered to detect a statistically significant difference on OS.

Progression free survival

The CS reports trial results for PFS (CS pp. 63 to 65), presenting the number of patients in the analysis, the number censored, the stratified log-rank p-value, the stratified hazard ration (HR) and 95% confidence intervals (CIs). The HR for PFS was estimated from a stratified Cox proportional hazards model; the stratification factors were the same as for OS. A KM curve for the ITT population is presented in CS Figure 6 (p. 64) and outcomes of median PFS and 3-month and 6-month PFS are presented with 95% CIs in CS Table 15 (p. 79).

Objective response rate

For ORR (CS pp. 65 to 66) the CS reports trial results for the ITT population as numbers and proportions, 95% CIs (from the Wilson test) as appropriate and p-values based on the Cochran-Mantel-Haenszel test adjusted for the stratification factors. The treatment effect for the ITT population is not provided in the CS despite the CS stating on page 53 that the stratified OR (Group B over Group A) and the estimated difference (A minus B) in ORR are presented along with the corresponding 95% CIs.

Analysis population for OS, PFS and ORR

The ITT population was used for OS, PFS, and ORR and included all randomly assigned patients. The CS states on page 54 that all analyses are based on the observed data. On page 52 of the CS it states that additional analyses were completed for PFS on a per protocol population, however, the outcomes from these analyses are not reported. For PFS analyses, page 52 of the CS states that additional analyses using unstratified log-rank tests were performed; however, the results of these are also not presented in the CS. In addition, the CS states (pp. 52 to 53) that sensitivity analyses for PFS, using alternative censoring rules, w performed according to rules specified in the statistical analysis plan. No further details or results from these sensitivity analyses are presented.

Health-related quality of life

HRQoL outcomes are presented as HRs and 95% CIs for time to deterioration (undefined – see Section 3.1.5) for the LCSS scale items and composite scores, and as descriptive statistic for the LCSS scales and for the EQ-5D-3L. These reported analyses reflect those in the CSR. It is unclear in the CS what the pre-specified analysis plan for HRQoL was and therefore wh what is reported is in line with the analysis plan. In response to a question from NICE and the ERG, the company clarified that no formal statistical test of difference between arms was planned for EQ-5D (clarification response A13). Summary statistics, including change fro baseline, for the index score (using UK weights) and the VAS for each assessment visit in the chemotherapy phase (up to cycle 6) by treatment arm were planned. These are not presented in the CS, but the company supplied the results in Appendix 7 of the clarifications response. The CS described the analysis as being undertaken on all patients with a baseline value and at least one post-baseline value. Patients without baseline and/or post-baseline assessments were censored at the randomisation date (clarification response A12). The CSR also states, that HRs for time to deterioration in LCSS (the outcome presented in the CS, see Section 3.1.5) were estimated using Cox proportional hazards. For EQ-5D, the CSR reports only that summary statistics were used.

Subgroups

The CS states that the SQUIRE trial had pre-planned exploratory subgroup analyses for OS and PFS (which is consistent the with trial publication). These included analyses by geographic region (five subgroup comparisons): Korea and Taiwan combined vs. all others; Eastern Asia

vs. all others; Eastern Europe vs. Eastern Asia vs. all others; Eastern Europe vs. all others; Each non-Eastern country with >40 patients randomized vs. Eastern Asia vs. all others; Each country with >40 patients randomized vs. all others. Subgroup analyses were also conducted by age, gender, race, ECOG performance status, smoking history and EGFR expression (by H-score values <200 or ≥200, see Section 3.1.5 for description of H-score). Each analysis was completed using the same methodology as for the primary analyses except that tests were unstratified. Results for these subgroups (other than geographical region) are presented in a forest plot (CS figure 9, p. 70) which concurs with the trial publication, and for the EGFR expression groups in various Kaplan-Meier plots (CS pp 71 - 73) which concurs with the CSR (see Section 2.3 for further discussion of these subgroups). As discussed in section 2.3 above, the subgroup analyses by age and ECOG performance status are not fully presented in line with the pre-specified categories. That is, ECOG performance status results are presented for 0, 1 and 2 only and not for a combined 0-1 category (which was also planned). The ERG identified a conference abstract reporting the pre-planned subgroup analyses during its searches[15] – please see section 3.1.3 of this report for details and section 3.3 for a summary of the results. The age results are not presented for patients aged <70 versus ≥70 years and <65 versus ≥65 years (as planned), but are instead provided for subgroups of patients aged <65 years, ≥65 to <70 years, and ≥70 years. Results from the age and ECOG performance status subgroups are not, however, used in the economic model.

Results of the pre-planned subgroup analyses by geographic region are not presented in the CS and were requested by NICE and the ERG. Clarification response A6c Appendix 6 presents subgroup analyses by geographic region but not in the groupings stated on CS pp. 68 to 69.

Post hoc subgroup analyses were completed for patients in countries classified by the CS as being in Western Europe (includes participants from Austria, Belgium, Germany, France, Greece, Italy, Portugal, Spain and UK, see Section 2.3 'Subgroups' for further details). Results are presented from analyses using the same approaches as the ITT analyses described above. Results are presented in the CS following the presentation of each of the ITT analyses for the outcomes of OS, PFS, ORR and TTF (CS pp. 61 to 62 for OS; pp. 64 to 65 for PFS; p. 66 for OR; pp. 67 to 68 for TTF). Please see Section 2.3 for ERG's critique of this post hoc subgroup.

Clinical significance

The CS does not discuss the clinical significance of the results seen for the OS and PFS outcomes, other than commenting that the OS difference in the Western European subgroup could be considered clinically significant. The company, however, did not define what a clinically meaningful change would be. The ERG is unclear what the minimally clinically important differences are on these outcomes for patients with NSCLC. The clinical expert consulted by the ERG stated that previous trials that have found an improvement in survival of around one month have been considered practice changing, although the expert was unable to comment on precisely what would be a clinically meaningful difference. The CSR reports that a clinically meaningful difference on the scales of the LCSS is a change of ≤ 15 mm, and this was used to categorise patients as having either improved, stable or worsened status in the LCSS results presented in the CS.

Summary

The ERG considers the trial statistics to be appropriate for survival outcomes, response rates and the pre-planned subgroup analyses. The ERG has reservations about the reliability of the post hoc Western European subgroup analysis (see also Section 2.3) and believe it is unclear if the OS benefits reported in the CS are clinically meaningful for all the populations included in the CS and clarification response.

3.1.7 Description and critique of the company's approach to the evidence synthesis

SQUIRE trial

Given that only one trial (SQUIRE) was included comparing necitumumab with one of the scoped comparators (GCis), a pairwise meta-analysis was not feasible.

A narrative review of the evidence from the SQUIRE RCT is presented in the CS. Where possible, the ERG has checked key data presented in the CS against those in the publication [8] and CSR. CS Table 40 p. 125 has an error in the number of Grade 3, 4 and 5 arterial thrombotic events. The correct data from the publication is reproduced in the ERG report. Otherwise, the adverse events results in the CS are consistent with those in the trial publication. [redacted]

[redacted]

Network meta-analysis

To enable comparison of necitumumab against scoped comparators for which there is no direct evidence, the company conducted a NMA comprising four networks to capture indirect evidence:

- An analysis of OS HR data
- An analysis of OS median data
- An analysis of PFS HR data
- An analysis of PFS median data.

The systematic review conducted for the NMA appears to broadly follow conventional guidelines for systematic review (e.g. a systematic search for evidence was undertaken). However, the searches were out-of-date (conducted in January 2015). As mentioned in section 3.1.4, the CS does not include a quality assessment of the included studies. In response to a request by NICE and the ERG, the company provided a quality assessment for trials included in the NMA (clarification response A21 Appendix 10); however, this is presented in tabular format only, without a summary of the overall quality of the evidence base as requested.

Inclusion and exclusion of studies

The company clearly states the inclusion and exclusion criteria for the systematic review underpinning the NMA (CS Table 21, p. 85). To be included in the NMA, studies had to be RCTs that included patients with squamous NSCLC and that evaluated any first-line chemotherapy or concurrent radiation therapy and chemotherapy treatment in each trial arm. Study inclusion was not limited to just the scope and decision problem specified intervention (i.e. GCis + N) and comparators. To be eligible, RCTs had to report results for OS, PFS, toxicity or HRQoL. RCTs including patients with other histological subtypes of NSCLC in addition to patients with squamous NSCLC were eligible, but had to provide a separate analysis for patients with "advanced or metastatic (Stage IV)" (CS p. 85) squamous NSCLC on at least one outcome of interest. In response to a question from NICE and the ERG, the company clarified that only patients with advanced or metastatic squamous NSCLC (stages IIIB and IV) were

included in the NMA (clarification response A5). Inclusion was restricted to English language publications, published from 1995 onwards. Setting was not used as an eligibility criterion and no restrictions were placed on the quality of the RCTs for inclusion in the review.

At the final analysis stage of the NMA systematic review, after full text screening, a second set of eligibility criteria was applied and the company excluded studies of agents that are not used to treat NSCLC (n=1); those used only in non-squamous NSCLC (n=6) and those that did not contain a comparator that enabled connection to a common comparator in the NMA networks (n=5). The company additionally excluded trials of unapproved experimental agents and agents without a marketing authorisation in any country (but not necessarily limited by histology) (n=10), although those recommended by clinical treatment guidelines and/or used off-label for the first-line treatment of advanced or metastatic squamous NSCLC were included (clarification response A4a). One additional trial was excluded as it compared two dosing schedules of the same regimen.

The ERG considers that the company's wide inclusion criterion related to the intervention (evaluation of any first-line chemotherapy) is appropriate, even though this meant that studies including interventions outside the scope in at least one trial arm could be included. It is appropriate to include these studies if they contribute evidence to the network, as long as they are clinically relevant (i.e. include the same patient groups and outcomes as other studies included in the network). The ERG, though, does not agree with all the exclusions the company made on the basis of studies using experimental or unapproved agents at the second screening stage. The ERG also does not agree with some of the other post-hoc exclusions of studies. Please see 'Identified studies' sub-section below for a further discussion. Overall, however, the ERG considers that the eligibility criteria adequately reflect the decision problem, except that response rate was not specified as an outcome of interest, so evidence for this outcome was not included.

The CS includes a flow diagram showing the number of studies included and excluded at each stage of inclusion for the systematic review for the NMA (CS Figure 22, p. 86). The flow diagram does not reflect the number of publications (n = 23) subsequently excluded from the NMA (as described above), but the company has summarised reasons for these exclusions in the CS text and provided a list of the 23 excluded studies excluded in CS Appendix 5. The ERG agrees with

the exclusions of trials comparing different doses of the same regimens. The ERG, however, considers that the following exclusions were insufficiently justified:

- Lynch et al. (2012) was excluded due to not having a comparator similar enough to the other trials in the network to enable connection with the network. The ERG notes, however, that CS Table 6 in Appendix 5 indicates that PCarbo was a comparator arm, and, based on this information, it appears that the trial could have been connected to the network via this arm.
- Eight of the 10 trials excluded due to using experimental or unapproved agents potentially could have been connected to the network through the PCarbo (Heymach et al., 2008; Langer et al., 2014; Lara et al, 2011; Novello et al., 2014; Paz-Ares et al., 2013; Reck et al., 2013; and, Scagliotti et al., 2010) and GCarbo arms (Spigel et al., 2013) of these studies. As these studies were included based on the initial inclusion criteria for the NMA, they would appear to be clinically relevant [i.e. include the same patient group (patients with squamous NSCLC) and outcomes as other studies included in the network]. NICE and the ERG requested additional clarification on the reasons for excluding these studies. However, the company re-iteratated that these studies investigated agents without market authorisation for the first-line treatment of patients with advanced or metastatic squamous NSCLC, but not necessarily limited by histology, without further details (clarification response A4b).
- Four of the six trials excluded due to one of the treatment arms receiving a drug limited to the treatment of patients with non-squamous NSCLC (Sandler et al, 2010; Scagliotti et al 2008, Johnson et al, 2007, and Zhang et al, 2013). Again, as these studies were included based on the initial inclusion criteria, they would appear to include patients with squamous NSCLC (or at least a subgroup) and to have measured relevant outcomes. The ERG considers that these studies could potentially have been connected to the network through the PCarbo (Johnson et al, 2007; Sandler et al, 2010) and GCis (Scagliotti et al, 2008; Zhang et al, 2013) arms.
- One trial (Lee et al, 2009) was excluded due to one of the treatment arms using a regimen not used in patients with NSCLC. Similar to above, having been included at the initial screening stage, the ERG suggests that this trial could potentially have been connected to the network through the GCis + placebo arm.

The ERG therefore considers that the efficacy estimates derived from the NMA may be subject to greater uncertainty, as not all relevant trials appear to have been included in the network. This may particularly affect the treatment effect estimates for comparisons against PCarbo and

GCis. This may impact the cost-effectiveness results for comparisons of GCis + N with PCarbo, as the base case uses the OS and PFS results from the NMA networks.

Identified studies

The systematic review for the NMA identified 10 RCTs (reported in 12 publications) that met the eligibility criteria (stated in the CS as 11 RCTs in 13 publications, but corrected in clarification response A19). The company lists the studies and comparisons included in the NMA in CS Tables 22 (pp. 93 to 94) and 23 (p. 96). However there are a number of discrepancies between these tables. In response to a question from NICE and the ERG, the company provided clarification on the studies and data included in the NMA (clarification questions A17 to A19, clarification Appendix 8). One of the 11 studies listed as included in the NMA was actually excluded (Yoshioka et al. 2013 assessing S-1, which is a combination of three drugs: tegafur, a fourth generation pro-drug of 5-fluorouracil; gimeracil; and oteracil); the company stated it was not included in the NMA as it is not relevant to countries outside of Japan (clarification reponse A19). The ERG does not believe that exclusion of this study was appropriate, because, as discussed above, the ERG considers that it is appropriate to include studies of unapproved or experimental agents if they are clinically relevant and contain a scope-specified treatment arm that could be connected to the network. The ERG considers that Yoshioka et al 2013 could potentially have been connected to the network through its PCarbo arm. The ERG has summarised the studies included in the NMA networks of OS (n = 6) and PFS (n = 7) using HR data in Table 5 and Table 6. We have not presented thestudies included in the two networks for median OS and median PFS here, as the data were not used in the economic model.

The ERG has not checked the company's quality assessment of trials included in the NMA but notes that all trials have been judged to have a high risk of bias on at least one domain of

The ERG notes that some of the arms of the trials included in the NMAs used drug doses that are not specified in the drugs' SmPCs. Clinical expert advice to the ERG is that although some of the drug doses are low, none of the doses used would likely adversely impact efficacy.

None of the included studies, apart from the SQUIRE trial, measured AEs and HRQoL in the squamous population, so NMA networks could only be formed for the OS and PFS outcomes.

Table 5 Trials included in the OS network using HR data

Trial	Interventions
SQUIRE	GCis + N GCis
Morabito 2013	GCis Gemcitabine
Hoang 2013[a]	GCis PCis DCis PCarbo
Socinski 2012	Sb-PCarbo Nab-PCarbo
Treat 2010	PCarbo GCarbo G + P
Kubota 2008	PCarbo VGD

[a] Not explicitly stated for this study, but assumed that the company calculated HRs by digitization of survival curves. DCis, docetaxel plus cisplatin; GCarbo, gemcitabine plus carboplatin; GCis, Gemcitabine plus cisplatin; GCis + N, necitumumab in combination with gemcitabine plus cisplatin; G + P, gemcitabine in combination with paclitaxel; Nab, nanoparticle albumin-bound; PCarbo, paclitaxel plus carboplatin; PCis, paclitaxel plus cisplatin; Sb, solvent-based; VGD, vinorelbine in combination with gemcitabine and docetaxel.

Table 6 Trials included in the PFS network using HR data

Trial	Interventions
SQUIRE	GCis + N GCis
Morabito 2013	GCis Gemcitabine
Hoang 2013[a]	GCis PCis DCis PCarbo
Socinski 2012	Sb-PCarbo Nab-PCarbo
Treat 2010	PCarbo GCarbo G + P
Kubota 2008	PCarbo VGD
Lilenbaum 2008	PCarbo Erlotinib

[a] Not explicitly stated for this study, but assumed that the company calculated HRs by digitization of survival curves. DCis, docetaxel plus cisplatin; GCarbo, gemcitabine plus carboplatin; GCis, Gemcitabine plus cisplatin; GCis + N, necitumumab in combination with gemcitabine plus cisplatin; G + P, gemcitabine in combination with paclitaxel; Nab, nanoparticle albumin-bound; PCarbo, paclitaxel plus carboplatin; PCis, paclitaxel plus cisplatin; Sb, solvent-based; VGD, vinorelbine in combination with gemcitabine and docetaxel.

Similarity of included studies

The CS does not present summary baseline characteristics from the 10 studies included in the NMA or provide other details (e.g. length of follow-up) that would enable a comparison of how similar the studies included were. Baseline data on age, sex, proportion of patients with stage IV disease and ECOG performance status 0, 1 and 2 were provided in response to a request by NICE and the ERG (clarification response A20 Appendix 9). However other requested details [race, region (proportion from Western Europe) and length of follow-up were not provided].

Squamous participants were a subgroup of about 15% to 44% of the arms in all the trials in the NMA networks, other than the SQUIRE trial. The ERG notes there were slight imbalances in baseline characteristics between the arms in some trials (clarification response Appendix 9). The trials in the NMA networks were reasonably similar with respect to patients' mean age, around 60 to 65 years, although this was about 78 years in the Chen et al. 2012 study of vinorelbine. The proportions with other characteristics in the trial arms ranged as follows: men 44% to 84%, stage IV 75% to 100%, performance status 0 4% to 42%, performance status 1 0% to 78%, and performance status 2 0% to 100%. The ERG therefore does not agree with the company's statement on CS p. 89 that covariates were similar across studies in all but two studies.

Company's approach to conducting the NMA

The CS states that networks were analysed for the outcomes of OS and PFS, respectively, through calculation of HRs as the primary analysis. The CS states that HRs were extracted from the text of the publications, calculated from available data, or, where possible, extracted from the KM plot following digitisation of the curve. Also, the CS states that a secondary analysis assessed median time to death for OS, median time to progression for PFS and as a ratio of median time to event. It is unclear how the secondary analysis was conducted. Data on HRs and 95% confidence intervals were extracted, or calculated using data or KM plots, from included studies. Where KM curves were the source, the submission reports that the

proportional hazards assumption was applied and appropriately tested through correlating the scaled Schoenfeld residuals with the default transformation of time (and by visual inspection of patterns of residuals that may indicate non-proportionality not identified by tests of non-zero slopes). Median time to event data were log transformed. Where standard errors for HRs were not available, they were estimated using the standard error for the median time to event adopting an exponential distribution of survival time and log HR or from the number of subjects with events. Fixed- and random-effects models were planned, along with adjustment for influential covariates and extensive sensitivity analyses to assess the robustness of the models. Sensitivity analyses would investigate the effects of the method of analysis (i.e. Bayesian versus frequentist), different survival outcome measures, geographical location, disease severity, patient age, study design, study quality and risk of bias.

The models adopted a Bayesian framework through the BATMAN tool, which uses the JAGS software program for the Bayesian analysis of hierarchical models. The BATMAN tool was developed in collaboration with the company to undertake NMA using a Bayesian approach through Markov Chain Monte Carlo simulation. The CS indicates that models developed in BATMAN were validated through independent replication in OpenBUGS software. All models were estimated using two chains with a 10,000 iteration burn-in and 2,000 iterations for estimating the posterior distribution. Clarifications from the company corrected the burn-in iterations to 1,000 (clarification response A24). Convergence was assessed using trace plots and autocorrelation plots (clarification response A25). Heterogeneity was planned to be explored through use of forest plots and consistency through comparison of outcomes from any closed loops in the network and through density plots of posterior distributions. Model fit was to be assessed through the variance, between-study standard deviation, residual deviance and deviance information criterion, where possible. In practice, the company did not assess model fit. The deviance information criterion (DIC) was provided for information only, but not discussed.

Table 7 shows the ERG's critical appraisal of the company's NMA. Despite the CS providing a basic outline of the approach taken to the NMAs, specific issues limit its usefulness. The r concern is the sparse nature of the evidence identified, which affects the analysis that was possible. Although the evidence networks included all the interventions in the scope exce vinorelbine in combination with a platinum drug, most comparisons only involved one study, and outcomes could be produced for only six of the eight specified comparisons. Despite the

inclusion of some comparators that were outside the scope to provide additional information to inform the NMA, the evidence base remained limited.

The submission planned to produce fixed- and random-effects models for the NMAs, however the lack of evidence meant that random-effects models would not converge (confirmed in clarification response A23). As a result, only fixed-effect models are presented, which may result in narrower credible intervals. With some of the credible intervals from the fixed-effect models being close to unity, estimation of random-effects models might have resulted in wider credible intervals that included unity, affecting the interpretation of the outcome. Only unadjusted analyses could be estimated as the inclusion of covariates could not be supported by the evidence available, with models failing to converge. Of the eight pre-planned sensitivity analyses, only two were conducted due to fragmentation of the networks. The limited evidence available will have affected the assessment of heterogeneity and consistency between direct and indirect evidence due to a lack of contributing studies that can be compared through forest plots, density plots, direct and indirect evidence for closed loops and comparisons of between-study variance. The company did not comment on how their results compared with those of other published NMAs. For example, the ERG notes that Brown et al. 2013[19] conducted mixed treatment comparison of the clinical effectiveness of first-line chemotherapy for patients with locally advanced or metastatic NSCLC. The results of the company's NMA could have been compared with those of the Brown et al. 2013 mixed treatment comparison.

In addition, the CS provides limited details of some aspects of the approach taken. Although a proportional hazards assumption is applied to the HR and tested for, only the results for a limited number of included studies are discussed. These identified some concerns around violation of the proportion hazards assumption (clarification response A26) and the company submission indicates that conclusions regarding appropriateness of the proportional hazards assumption should not be drawn across all studies in the network.

Apart from the sample OpenBUGS code provided in appendix 8 of the submission, there is no discussion regarding the prior values used for treatment effects or for the prior distributions for the variance component (i.e. alternative priors used). It is unclear if any sensitivity analyses were undertaken around the different priors. Clarification from the company outlined the OpenBUGS code for the different models (clarification response A22); however, no clarification was provided regarding sensitivity analyses on prior values or distributions. It was evident from

the OpenBUGS code and the clarifications (clarification response A24) that vague priors were used. Similarly, none of the diagnostic plots to assess convergence, autocorrelation, heterogeneity or consistency are presented or discussed. Only trace plots are used to assess convergence (Brooks-Gelman-Rubin plots would have provided further clarity as to whether convergence had occurred). These limitations make it difficult to judge whether there was a sufficient period of burn-in or additional iterations to allow model convergence. This is particularly important given the limited numbers of iterations used for burn-in and estimation of the posterior distribution. Finally, results from the frequentist analysis and validation analysis in OpenBUGS are not presented or discussed.

Summary and key caveats:

Overall, the ERG considers that the sparse evidence base and the lack of clarity around the NMA indicates that there is considerable uncertainty and the outcomes should be viewed with caution, something identified in the CS itself. Other key caveats are:

- The post-hoc exclusion of studies that the ERG considers relevant and which could have been connected to the network.
- It is unclear if the included studies were similar enough to combine as the company provided limited information about participants' baseline characteristics and the company did not provide information about length of follow-up.
- The ERG considers that the fixed effects model was not a plausible model choice and the company have not adequately explored the possibility of using a random effects model, so there is uncertainty in the model.
- The company has used a model unadjusted for covariates.

Table 7 ERG appraisal of the NMA approach

Checklist	Response (yes/no)
Does the MS present an MTC?	Yes (NMA)
Are the NMA results used to support the evidence for the clinical effectiveness of the intervention	Yes
Are the NMA results used to support the evidence for the cost-effectiveness of the intervention	Yes
Homogeneity	
1. Is homogeneity considered?	Yes, with caveats. The NMA clearly specifies the participants that will be included in the systematic review and it undertakes a limited assessment of heterogeneity (see below). However there is no discussion of how similar the studies are in terms of intervention characteristics or outcome measures. Most of the studies included patients with non-squamous histology and only data from the subset of squamous patients were used, which was <30% of the study population in most cases. Similarity of the squamous subgroup between arms within the trials was not considered.
2. Are the studies homogenous in terms of patient characteristics and study design?	Unclear. Limited baseline characteristics were provided in clarification response A20 Appendix 9. The proportion of squamous patients in each trial arm ranged from 15% to 44% and was 100% in the SQUIRE trial. Median age ranged from 59 to 78 years; the proportion male ranged from 44% to 84%; and the proportion with stage IV disease ranged from 75% to 100%.The PS varied considerably among trials, with the proportion of patients classed as PS0, PS 1 and PS2 ranging, respectively, from 0% to 42%, 0% to 78%, and 0% to 100%.
3. Is the method used to determine the presence of statistical heterogeneity adequate? (e.g. Chi-squared test, I-squared statistic)	No (visual inspection of forest plots only). Due to the limited number of studies and small patient numbers the random effects heterogeneity variance became inestimable and the random effects models did not converge in all instances. Therefore all analyses were conducted using a fixed effects model.
4. If the homogeneity assumption is not satisfied, is clinical or methodological homogeneity across trials in each set involved in the indirect comparison investigated by an adequate method? (e.g. subgroup analysis, sensitivity analysis, meta-regression)	Partially. Homogeneity: The majority of comparisons were populated with only one study. Heterogeneity was identified by wide credible intervals for one study (Morabito 2013) in the OS HR analysis and two studies (Morabito 2013, Lilenbaum 2008) in the PFS HR analysis. These studies were not central to the network connections for each of the OS and PFS analyses and could be removed without affecting the results versus the necitumumab region, therefore post hoc sensitivity analyses were undertaken. The CS does not comment on heterogeneity (identified by wide confidence intervals) for studies that were central to the network connections.
Similarity	
1. Is the assumption of similarity stated?	No

2. Have they justified their assumption?	No
Consistency	
1. Does the analysis explicitly assess consistency of direct and indirect evidence?	Not applicable (the CS states (p. 119) that the consistency assumption could not explored due to the lack of closed loops that included GCis + N).
2. Does the method described include a description of the analyses/ models/ handling of potential bias/ inconsistency/ analysis framework?	Not applicable
3. Are patient or trial characteristics compared between direct and indirect evidence trials?	Not applicable
4. If Q3 is yes, and inconsistency is reported, is this accounted for by not combining the direct and indirect evidence?	Not applicable

3.2 Summary statement of the company's approach

The ERG's quality assessment of the CS is summarised in Table 8. The quality of the company's systematic reviews is suboptimal and there is a chance of systematic error in the reviews. The company appears to have included all relevant RCTs in the main systematic review. The company, however, did not identify four conference abstracts related to the SQUIRE trial in their searches, which were published before the company conducted the searches for the main systematic review (in August 2015) and that were subsequently identified by the ERG's update searches conducted from January 2015 to present. The company's searches for the systematic review underpinning the NMA was conducted in January 2015, therefore any relevant studies published in the past year will have been missed. Eligibility criteria were reported for the main systematic review and NMA review, however inappropriate post-hoc exclusions were performed for the NMA review. The process for study selection, data extraction and quality assessment are not described in the CS, but details were provided in clarification response A3 and are appropriate. In addition, limited details of study characteristics are reported for those trials included in the NMA.

The submitted evidence generally reflects the decision problem. However, the only head-to-head comparison was with GCis. Comparisons with DCis, PCis, GCarbo and PCarbo were made through an NMA, although there were a number of limitations with this due to the sparse network. Comparisons could not be made with DCarbo or VCarbo, and comparison with VCis

could only be made for OS analyses based on median data (which could not be used in the economic model).

Table 8 Quality assessment (CRD criteria) of CS review

CRD Quality Item; score Yes/No/Uncertain with comments	
1. Are any inclusion/exclusion criteria reported relating to the primary studies which address the review question?	Partial. Criteria for the main systematic review are reported in CS Table 9 p. 39. Eligibility criteria for the NMA are reported in CS Table 21 p. 85, however some studies were excluded post hoc.
2. Is there evidence of a substantial effort to search for all relevant research? ie all studies identified	Partial. The search strategy for the systematic review of GCis + N was appropriate, but not all relevant publications relating to the SQUIRE trial were identified and searches were 5-6 months out of date. However, no relevant studies were missed by these searches. The search strategy for the NMA was appropriate but last updated in January 2015 therefore any studies published in the last year will have been missed.
3. Is the validity of included studies adequately assessed?	Partial. The CS uses the NICE recommended criteria and the ERG generally agrees with the company's assessment of the SQUIRE RCT (see section 3.1.4). Quality assessment of the trials included in the NMA was not provided in the CS but was provided on request from NICE and the ERG in clarification response Appendix 10. The company assessed trial validity using the PEDRO tool and Cochrane Risk of bias tool which are appropriate, although the company does not explicitly discuss whether or how their quality assessment results informed their NMA analyses and conclusions.
4. Is sufficient detail of the individual studies presented?	Partial. Details are provided for the SQUIRE RCT on methodology (CS p. 42-51), statistical analysis (CS p. 52-54), and participant flow (p. 54-55). Limited details were provided for the trials included in the NMA, but were provided on request from NICE and the ERG in clarification response Appendix 10.
5. Are the primary studies summarised appropriately?	Partial. Results of the SQUIRE RCT are presented in narrative form with tabulation of data. The synthesised results from the studies included in the NMA are highly uncertain (as discussed in section 3.1.7).

3.3 Summary of submitted evidence

Summary of evidence

During the appraisal, in response to clarification question A1, the company supplied results for the EGFR expressing subgroup of patients in the SQUIRE trial for all outcomes except HRQoL. Below, we present the EGFR expressing subgroup (n = 935) results first for each outcome, as this is the SmPC population, followed by the ITT population results. We have summarised the

results for all four patients populations included in the CS and company's clarification response in the subsection 'Summary of results for all populations', including the Western Europe subgroups results. We present results for the Western Europe subgroups these are used to inform the company's base case cost-effectiveness analysis. Where possible, the ERG has checked data with the published data. The CS also reports that pharmacokinetics and immunogenicity of necitumumab were secondary outcomes. These do not meet the decision problem and are not discussed further here.

Summary overall survival results

EGFR expressing subgroup

As shown in Table 9, in the EGFR expressing population, the median OS was 11.7 months [redacted] in the GCis + N group; and 10 months [redacted] in the GCis group: a modest improvement in OS of 1.74 months. The HR showed a statistically significant survival benefit in the GCis + N group compared with the GCis group (HR 0.79 (95% CI 0.69, 0.92), p=0.002). The number of deaths was [redacted] out of [redacted] participants in the GCis + N group and [redacted] out of [redacted] in the GCis group. Twelve month and 24 month survival rates were not reported in the clarification response.

Table 9 Overall survival in the EGFR expressing subgroup, as reported in the company's clarification response

	GCis + N, N = 462	GCis, N=473
OS results presented in the CS		
Number of deaths, n (%)	[redacted]	[redacted]
Number censored, n (%)	[redacted]	[redacted]
Median survival[a], months (95% CI)	11.73 [redacted]	9.99 [redacted]
Stratified Hazard Ratio (95% CI), p-value[b]	0.79 (0.69, 0.92) p=0.002	
12 month survival rate, % (95% CI)	NR	NR
24 month survival rate,% (95% CI)	NR	NR

CI = confidence interval; GCis: gemcitabine + cisplatin; GCis + N: necitumumab + gemcitabine + cisplatin; NR: not reported
[a]Kaplan-Meier estimated
[b]Stratified log-rank p-value

Table 10 shows the OS results from the SQUIRE trial for patients with and without detectable EGFR expression (H-scores of > 0 and = 0, respectively) reported in a FDA Briefing Document[14] identified by the ERG. [redacted]

[redacted]. A small proportion of participants in the SQUIRE trial (GCis + N = 24; GCis= 23) lacked detectable EGFR expression (H-score = 0). There was no statistically significant difference in OS (HR 1.86) between treatment arms for participants with an H-score of 0. The company commented in their clarification response A7 that results of a subgroup analysis of patients without detectable EGFR expression showed that these patients may not benefit from the addition of necitumumab to GCis.

Table 10 OS by % positive EGFR expression, as reported in a FDA briefing document identified by the ERG

	Percent positive >0		Percent positive = 0[a]	
	GCis + N n=462	GCis n=473	GCis + N n=24	GCis n=23
Overall survival				
Median, months	11.73	9.99	6.47	17.35
HR (95% CI)	0.81 (0.70, 0.93)		1.86 (0.94, 3.65)	
P value	0.004		0.072	
Interaction p value	0.018			

Source: FDA Briefing Document[14]
[a]0 % positive is equivalent to H-score=0 for EGFR staining

ITT population

Overall survival was longer with GCis + N than with GCis (Table 11). Median OS was 11.5 months (95% CI 10.4, 12.6) among 545 participants in the GCis + N group and 9.9 months (95% CI 8.9, 11.1) among 548 participants in the GCis group. The stratified hazard ratio (HR) was 0.84 (95% CI 0.74, 0.96) suggesting a 16% reduction in risk with GCis + N. At 1 year, the survival rate was 48% (95% CI 43, 52) with GCis + N versus 43% (95% CI 39, 47) with GCis. At 2 years the survival rate was 20% (95% CI 16, 24) with GCis + N compared with 17% (95% CI 13, 20) with GCis. The median follow-up time was 25.2 months (95% CI 23.7, 27.1) in the GCis + N group and 24.8 months (95% CI 22.8, 28.3) in the GCis group. Loss to follow-up and withdrawals of consent for follow-up for OS was low and similar across both groups (GCis + N 39 (7.2%); GCis 35 (6.4%)).

A conference abstract[16] from the SQUIRE trial identified by the ERG's searches that reported on the efficacy of treatment with necitumumab alone during the maintenance phase (i.e. following treatment with GCis + N) found a two-year survival rate of 27.5% among the 51% of the GCis + N patients receiving maintenance treatment.

Table 11 Overall survival (ITT population)

	GCis + N, N = 545	GCis, N=548
Number of deaths, n (%)	418 (77)	442 (81)
Number censored, n (%)	127 (23)	106 (19)
Median survival[a], months (95% CI)	11.5 (10.4, 12.6)	9.9 (8.9, 11.1)
Stratified Hazard Ratio (95% CI), p-value[b]	0.84 (0.74, 0.96) p=0.01	
12 month survival rate[a], % (95% CI)	48 (43, 52)	43 (39, 47)
24 month survival rate[a],% (95% CI)	20 (16, 24)	17 (13, 20)

CI = confidence interval; GCis: gemcitabine + cisplatin; GCis + N: necitumumab + gemcitabine + cisplatin
[a]Kaplan-Meier estimated
[b]Stratified log-rank p-value (stratified by ECOG PS and geographic region).

Summary of progression-free survival results

EGFR expressing subgroup

PFS in the EGFR expressing subgroup was slightly longer with GCis + N than with GCis (12). Median PFS was 5.7 months [redacted] in the GCis + N group and 5.5 months [redacted] in the GCis group. The HR for PFS was 0.84 (95% CI, 0.72, 0.97), p=0.018.

Table 12 Progression-free survival in the EGFR expressing subgroup, as reported in the company's clarification response

	GCis + N, N = 462	GCis, N=473
Number of events, n (%)	[redacted]	[redacted]
Number censored, n (%)	[redacted]	[redacted]
Median PFS[a], months (95% CI)	5.72 ([redacted]	5.49 [redacted]
Stratified Hazard ratio (95% CI), p-value[b]	0.84 (0.72, 0.97) p=0.018	
3 month PFS rate[a], % (95% CI)	NR	NR
6 month PFS rate[a],% (95% CI)	NR	NR

CI = confidence interval; GCis: gemcitabine + cisplatin; GCis + N: necitumumab + gemcitabine + cisplatin; NR: not reported
[a]Kaplan-Meier estimated
[b]Stratified log-rank p-value

Table 13 shows the PFS results from the SQUIRE trial for patients with and without detectable EGFR expression (H-scores of > 0 and = 0, respectively) reported in the FDA Briefing Document[14] identified by the ERG. [redacted]. The was no statistically significant difference in PFS (HR 1.19) between treatment arms for participants with an H-score of 0.

Table 13 PFS by % positive EGFR expression, as reported in a FDA briefing document identified by the ERG

	Percent positive >0		Percent positive = 0[a]	
	GCis + N n=462	GCis n=473	GCis + N n=24	GCis n=23
Progression-free survival				
Median, months	5.72	5.49	4.24	5.59
HR (95% CI)	0.83 (0.72, 0.97)		1.19 (0.61, 2.30)	
p-value	0.015		0.611	
Interaction p value	0.305			

Source: FDA Briefing Document[14]
[a]0 % positive is equivalent to H-score=0 for EGFR staining

ITT population

PFS was slightly longer with GCis + N than with GCis (Table 14). Median PFS was 5.7 months (95% CI, 5.6, 6.0) in the GCis + N group and 5.5 months (95% CI, 4.8, 5.6) in the GCis group (HR for progression or death 0.85; 95% CI 0.74, 0.98). At 3 months, the PFS rate was 79% (95% CI, 76, 83) with GCis + N versus 73% (95% CI, 68, 76) with GCis. At 6 months, the PFS rate was 45% (95% CI, 40, 49) with GCis + N versus 37% (95% CI, 33, 42) with GCis. It is not clear to the ERG what the median follow-up time was for the assessment of PFS, but the ERG notes that the number of events in the GCis group is lower than the number of deaths in this group, as presented in Table 11 above.

Table 14 Progression-free survival (ITT population)

	GCis + N, N = 545	GCis, N=548
Number of events, n (%)	431 (79)	417 (76)
Number censored, n (%)	114 (21)	131 (24)
Median PFS[a], months (95% CI)	5.7 (5.6, 6.0)	5.5 (4.8, 5.6)
Stratified Hazard ratio (95% CI), p-value[b]	0.85 (0.74, 0.98) p=0.02	
3 month PFS rate[a], % (95% CI)	79 (76, 83)	73 (68, 76)
6 month PFS rate[a],% (95% CI)	45 (40, 49)	37 (33, 42)

CI = confidence interval; GCis: gemcitabine + cisplatin; GCis + N: necitumumab + gemcitabine + cisplatin
[a]Kaplan-Meier estimated
[b]Stratified log-rank p-value (stratified by ECOG PS and geographic region).

EGFR expressing subgroup

In the EGFR expressing subgroup, the difference in ORR between the GCis + N and GCis groups was [redacted] (Table 15). The disease control rate in this subgroup was [redacted] in the GCis + N group [redacted] than in the GCis group [redacted].

Table 15 Objective response rate in the EGFR expressing subgroup

n (%)	GCis + N, N = 462	GCis, N=473
Objective response (CR+PR) rate, n (%), 95% CI	[redacted]	[redacted]
Difference (95% CI)	[redacted]	
OR (95% CI), p-value	[redacted]	
Disease control rate (CR+PR+SD) (95% CI)	[redacted]	[redacted]
Difference (95% CI)	[redacted]	
OR (95% CI), p-value	[redacted]	
Best overall response, n (%):		
Complete response (CR)	[redacted]	[redacted]
Partial response (PR)	[redacted]	[redacted]
Stable disease (SD)	[redacted]	[redacted]
Progressive disease (PD)	[redacted]	[redacted]
Not evaluable/No assessment[b]	[redacted]	[redacted]

CI = confidence interval; GCis: gemcitabine + cisplatin; GCis + N: necitumumab + gemcitabine + ci[redacted]

[b]Calculated by ERG from 'not evaluable' and 'no assessment'.

ITT population

ORR was higher with GCis + N than with GCis, however this was not statistically significa (p=0.40) (**Table** 16). The ORR was 31% (95% CI, 27, 35) in the GCis + N group and 29% CI, 25, 33) in the GCis group. The disease control rate was also reported in the CS, this v significantly higher in the GCis + N group than in the GCis group (**Table** 16).

Table 16 Objective response rate (ITT population)

n (%)	GCis + N, N = 545	GCis, N=548
Objective response (CR+PR) rate (95% CI)	170 (31) (27, 35)	158 (29) (25, 33)
p-value (stratified Cochran-Mantel-Haenszel[a])	0.40	
Disease control rate (CR+PR+SD) (95% CI)	446 (82) (78, 85)	422 (77) (73, 80)
p-value (stratified Cochran-Mantel-Haenszel[a])	0.043	
Best overall response, n (%):		
Complete response (CR)	0	3 (<1)
Partial response (PR)	170 (31)	155 (28)
Stable disease (SD)	276 (51)	264 (48)
Progressive disease (PD)	41 (8)	55 (10)
Not evaluable/No assessment[b]	58 (11)	71 (13)

CI = confidence interval; GCis: gemcitabine + cisplatin; GCis + N: necitumumab + gemcitabine + cisplatin
[a]stratified by ECOG PS and geographic region.
[b]calculated by ERG from not evaluable and no assessment in CS Table 16

Summary of Health related quality of life (ITT population only)

The CS presents data from the SQUIRE trial on the LCSS. For the LCSS the CS presents the time to deterioration for lung symptoms based on each of six symptom questions (loss of appetite, fatigue, cough, dyspnoea, haemoptysis, pain) together with the Average Symptom Burden Index (ASBI). The CS also reports three global items (overall symptoms, interference with normal activities, quality of life) together with a global 3-item composite score. Finally, a total LCSS score is presented. On the LCSS each item is assessed with a 100-mm visual analogue scale with higher ratings corresponding to poorer quality of life. The CS reports that 88.3% of participants in the GCis + N group and 88% of participants in the GCis group had a baseline and at least one post baseline assessment of the LCSS. The CS reports summary data for the nine individual items of the LCSS at baseline for both treatment groups (CS Table 20, p. 74) and data for the time to deterioration in a forest plot (CS Figure 14, p. 77). This includes the number of events for each item by treatment group, and the HR and 95% CI. No items on the LCSS were significantly different between treatment groups, with all confidence intervals from the reported HRs crossing 1.0. As stated in section 3.1.5, a number of other analyses of the LCSS were undertaken, but the results of these analyses are not reported in the CS.

The CS also presents data on the EQ-5D-3L. Data were reported for the EQ-5D index at baseline, cycles 2-6 and end of therapy in a series of figures illustrating the percentage of patient responses for each response option for each EQ-5D dimension (CS figures 15-17, pp. 79 to 81) and in a series of figures illustrating the percentage of patient responses at each visit (CS figure 21, p. 83). The CS reports that [redacted] of participants in the GCis + N group and [redacted] of participants in the GCis group had a baseline and at least one post baseline assessment of the EQ-5D. The CS reports on page 78 that most patients in both arms experienced no or some problems in each of the five dimensions and that fewer than 6.5% in each arm experienced extreme problems on any of the five dimensions.

The CS also reports ECOG PS time to deterioration, and an analysis of severity of LCSS as predictor of OS. These analyses are post hoc and are therefore not summarised by the ERG (see CS p. 77 to 78).

Pre-specified sub-group analysis results (ITT population only)

There were no subgroups noted as relevant in the NICE scope or the decision problem. Pre-specified subgroup analyses (described in CS Section 4.8 pp. 68 to 73) were planned of the OS and PFS outcomes by:

- geographical region (see section 3.1.6 for a list of the pre-planned analyses)
- age (<70 vs. ≥70 years; and <65 vs. ≥65 years).
- gender (female vs. male).
- race (White vs. non-White).
- ECOG PS (0 vs. 1 vs. 2 and 0-1 vs. 2).
- smoking history (never smoker [non-smoker and light ex-smoker combined] vs. smoker).

Except for the pre-specified analyses by geographical region (provided in clarification response A6c, Appendix 6 but in different groupings to those stated above), results were presented for these subgroups in Figure 9 in the CS (p. 70). These concur with results seen in the CSR with the exception of the age subgroup analyses. In the CS and clarification response Appendix 6, the subgroups (including those by geographical region) show a similar pattern on outcomes of OS and PFS, generally favouring treatment with GCis + N (please see Figure 1, Figure 2 and Figure 3).

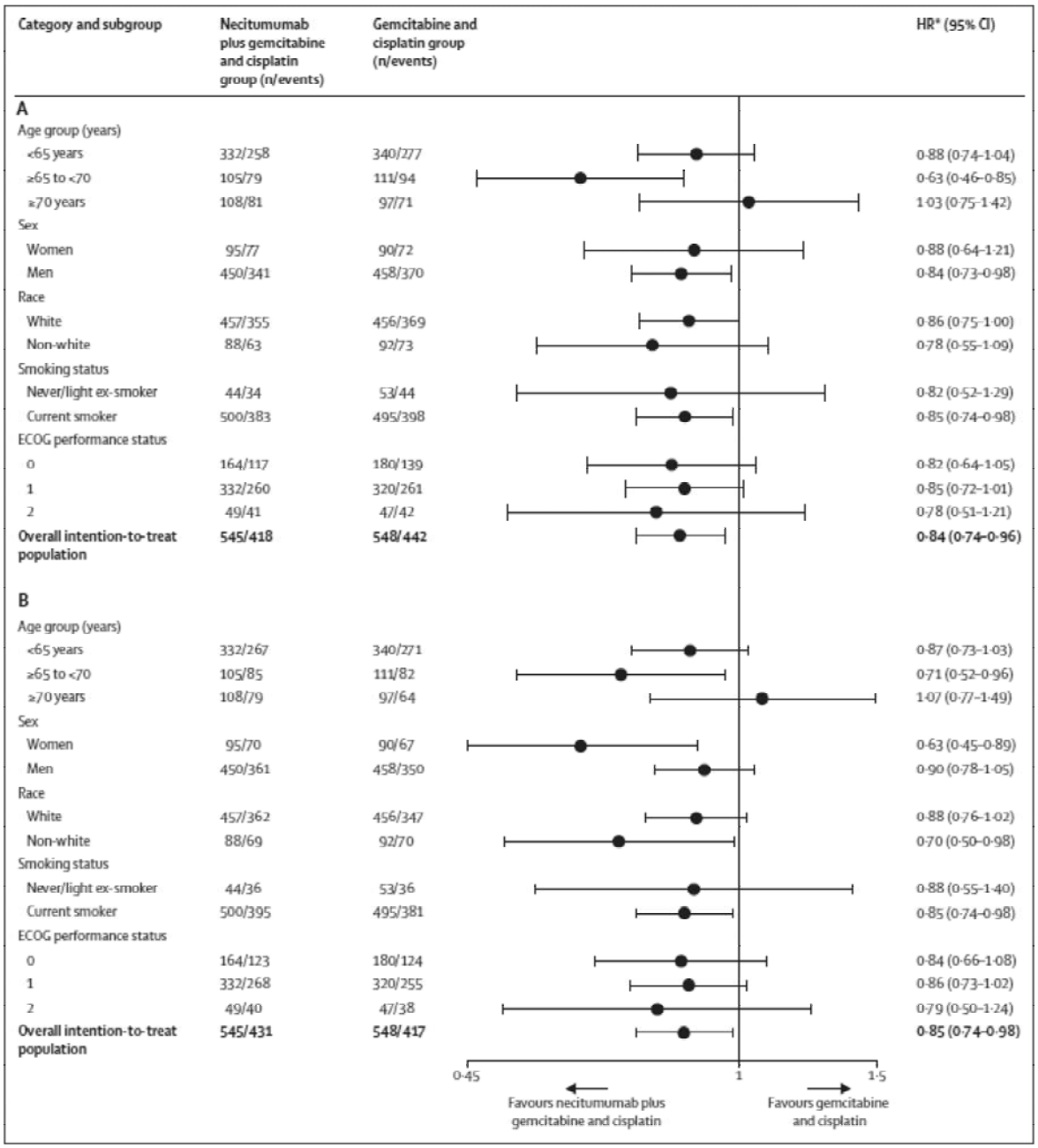


Figure 1 Forest plot of subgroup analyses results in ITT population Panel A shows OS results; panel B shows PFS results.

Panel A shows OS results; panel B shows PFS results. Figure reproduced from Appendix 6, p. 61, of the company's clarification response.

Source: Thatcher et al. 2015.

Figure 2 OS Subgroup analyses results by geographical region.

Figure reproduced from Appendix 6, p. 61 of the company's clarification response.

Figure 3 PFS Subgroup analyses results by geographical region

Figure reproduced from Appendix 6, p. 62, of the company's clarification response.

As discussed in section 2.3 above, the age and ECOG performance status results were not presented in the CS entirely in line with the pre-specified groupings of patients. The ERG therefore considers the age and performance status results presented in the CS to be at risk of selective reporting bias.

The ERG identified a conference abstract[15] through its searches that reports the results of analyses of patients with a performance status of 0-1; results that were not reported in the CS. This showed that patients treated with GCis + N who had an ECOG performance score of 0-1 experienced statistically significantly better OS [HR 0.85 (95% CI: 0.74 to 0.98; p = 0.026) and PFS [HR 0.86 (95% CI: 0.75 to 0.99; p = 0.035) than patients treated with GCis alone, while patients treated with GCis + N with a PS score of 2 did not [OS: HR 0.78 (95% CI: 0.51 to 1.21; p = 0.275); PFS: HR 0.79 (95% CI: 0.50, 1.24; p = 0.292)].

The CS (and trial publication) also report pre-specified exploratory analysis of tumour EGFR expression, categorising participants into high (H-score ≥200) and low (H-score <200) EGFR expression groups. This was undertaken to establish whether H-scores were predictive of a differential effect of the addition of necitumumab on OS and PFS. No significant differenc were seen in HRs for OS or PFS between the two H-score groups (see CS pp. 71 to 73).

Summary of results for all populations

In Table 17, the ERG has summarised the OS, PFS and ORR results reported for all four populations included in the CS and the company's clarifications response. As Table 17 sh [redacted]. These were population subgroups used in the company's economic model base case. The CS states although the improvement in OS in the Western Europe subgroup [redacted] is moderate clinically significant in the squamous NSCLC populations. The company did not comment clinical significance of the results for the other populations. The difference in median OS in the EGFR expressing subgroup (the subgroup the ERG considers to be most relevant to the SmPC indication and this appraisal) was 1.74 months, favouring GCis + N (HR 0.79 (0.69, 0.92).

There were statistically significant differences between the treatment arms in OS for all four populations. GCis + N resulted in statistically significantly better PSF outcomes than GCis in the ITT population and EGFR expressing population. There were, however, no statistically

significant differences in PFS between GCis + N and GCis in either of the Western Europe subgroups.

Table 17 Summary of results for all SQUIRE trial populations presented in the CS and the company's clarifications response

	GCis + N	GCis
Median survival, months (95% CI)		
ITT population	11.5 (10.4, 12.6)	9.9 (8.9, 11.1)
EGFR subgroup	11.73 [redacted]	9.99 [redacted] [a]
Western Europe subgroup	[redacted] [a]	[redacted]
EGFR Western Europe subgroup		
OS: stratified HR (95% CI) [b]		
ITT population	0.84 (0.74, 0.96); p=0.01	
EGFR subgroup	0.79 (0.69, 0.92); p=0.002	
Western Europe subgroup	[redacted]	
EGFR Western Europe subgroup		
Median PFS, months (95% CI)		
ITT population	5.7 (5.6, 6.0)	5.5 (4.8, 5.6)
EGFR subgroup	5.72 [redacted]	5.49 [redacted]
Western Europe subgroup	[redacted] [a]	[redacted] [a]
EGFR Western Europe subgroup		
PFS: stratified HR (95% CI) [b]		
ITT population	0.85 (0.74, 0.98); p=0.02	
EGFR subgroup	0.84 (0.72, 0.97); p=0.018	
Western Europe subgroup	[redacted]	
EGFR Western Europe subgroup		
ORR, % (95% CI)		
ITT population	31 (27, 35)	29 (25, 33)
EGFR subgroup	[redacted]	[redacted]
Western Europe subgroup	[redacted] [c]	[redacted] [c]
EGFR Western Europe subgroup		
ORR difference (95% CI)		
ITT population	Not reported	
EGFR subgroup	[redacted] [d]	
Western Europe subgroup	[redacted] [c]	
EGFR Western Europe subgroup		
ORR: odds ratio (95% CI)		
ITT population	Not reported	
Western Europe subgroup	[redacted]	
EGFR subgroup	[redacted]	
EGFR Western Europe subgroup	[redacted] [c]	

EGFR, epidermal growth factor receptor; GCis: gemcitabine + cisplatin; GCis + N: necitumumab + gemcitabine + cisplatin; ITT, intention-to-treat.

[a] CIs extracted by the ERG from the CSR

[b] unstratified analysis for EGFR Western Europe subgroup

[c] the company's clarification response Appendix 1 states that these results were for the Western Europe subgroup, but the ERG believes that this is a typo and that these results are for the EGFR expressing Western Europe subgroup.

[redacted]

[e] calculated by the ERG.

Mixed Treatment Comparison results

The CS presents results from the NMA as median HR and mean HR for OS and PFS HRs (Table 18). The means were utilised in the CS economic evaluation. The CS also presents results from the secondary analyses; however, these are not used in the economic evalua and are not reported here. As stated in section 3.1.7, the spare evidence base and lack o around the NMA indicates there is considerable uncertainty and the outcomes should be viewed with caution.

Overall Survival

The NMA for OS allows comparisons of GCis + N with five comparators in the decision problem (PCarbo; GCis; PCis; DCis; GCarbo). No evidence was identified that allowed pairwise comparisons with the remaining three comparators in the decision problem (DCarbo; VCarbo; VCis), although there is evidence for VCis in the secondary NMA analysis of median survival. The credible intervals for GCis + N with all comparators are overlapping, suggesting that this NMA does not provide evidence of a hierarchy of effectiveness between comparators.

Table 18 NMA overall survival estimates, fixed effect model

Intervention	Comparator				
	PCarbo	GCis	PCis	DCis	GCarbo
Median OS HR (from CS Table 25)					
GCis + N	[redacted]				
Mean OS HR[a] (from CS Table 26)					
GCis + N	[redacted]				

CrI: Credible interval; DCis: Docetaxel + cisplatin; GCarbo: Gemcitabine + carboplatin; GCis: Gemcitabine + cisplatin; GCis + N: Necitumumab + gemcitabine + cisplatin; PCarbo: Paclitaxel + carboplatin; PCis: Paclitaxel + cisplatin; VCis: Vinorelbine + cisplatin.
[a]95% credible intervals were obtained from the median HR analyses which the CS states is reasonable given they are taken from the same distribution

Progression-Free Survival

The NMA for PFS allows comparisons of GCis + N with the same comparators as in the OS analyses (PCarbo; GCis; PCis; DCis; GCarbo). There are no PFS HR comparisons available for DCarbo; VCarbo or VCis. In all analyses the HRs are favourable for GCis + N. For the comparison with GCis, PCis and GCarbo the 95% CrI do not include unity, indicative of a treatment effect for GCi + N on PFS (Table 19).

Table 19 NMA progression-free survival HR results, fixed effect model

Intervention	Comparator				
	PCarbo	GCis	PCis	DCis	GCarbo
Median PFS HR (from CS Table 28)					
GCis + N	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Mean PFS HR[a] (from CS Table 29)					
GCis + N	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

CrI: Credible interval; DCis: Docetaxel + cisplatin; GCarbo: Gemcitabine + carboplatin; GCis: Gemcitabine + cisplatin; GCis + N: Necitumumab + gemcitabine + cisplatin; PCarbo: Paclitaxel + carboplatin; PCis: Paclitaxel + cisplatin.
[a]95% credible intervals were obtained from the median HR analyses which the CS says is reasonable given they are taken from the same distribution, ERG notes upper bound CrI for GCis + N vs. PCarbo, and lower bound CrI for GCis + N vs. GCarbo are slightly different.

Summary of adverse events

Here we have summarised the AE results for the safety population from the SQUIRE trial, followed by a comment on how similar the AE results for the EGFR expressing subgroup, as provided in the company's clarification response Appendix 1, are to those observed in the safety population.

Safety population

For the safety population in the SQUIRE trial,[1] any participants who received at least one dose of therapy were included in the analyses. The CS states (p. 121) that there was a longer observation phase for those in the GCis + N group because the treatment phase was longer due to this group receiving maintenance therapy. To represent the difference in observation phases the CS states (page 121) that it presents safety data as overall for each group, and also separately for the chemotherapy and maintenance phases for the GCis + N group. The CS presents the chemotherapy phase and the maintenance phase and the overall safety set in different tables, although these are not always clearly identified. Where possible, the ERG has presented results from chemotherapy phases and maintenance phases separately unless these were not distinguished in the CS or CSR.

Treatment emergent adverse events (TEAEs) were defined in the SQUIRE trial as detaile in section 3.1.5. The frequency of deaths, including fatal cases of disease progression, that occurred as a result of TEAEs was similar between groups in the chemotherapy phase (GCis + N 9.3% vs GCis 10.5%) (Table 20). In the maintenance phase (i.e. necitumumab monotherapy) there were 5.8% TEAE related deaths. TEAEs with outcome of death, excluding fatal cas disease progression, were also similar between groups in the chemotherapy phase (see

20). Most participants reported at least one TEAE and the rates during the chemotherapy phases were similar between groups (>97%).

The proportion of participants reporting at least one treatment-emergent serious adverse event (SAE, defined in section 3.1.5) in the chemotherapy phase was 42.6% in the GCis + N group compared with 37.5% in the GCis group. These values concur with the CSR data; however, the publication for the SQUIRE trial[1] reports that serious adverse events were reported in 48% of participants in the GCis + N group. The CS also differs from the published data on the proportion of participants with grade 3 or worse TEAEs in the GCis + N group (67.7% in the CS, 72% in the publication). On checking the CSR, the data in the trial publication were found to correspond to the overall safety set, which includes the maintenance phase for the GCis + N treatment group. The proportion of participants discontinuing the allocated treatments due to at least one TEAE was similar between the two arms of the SQUIRE trial (13.8% and 14.8% respectively).

The CS does not report details about which specific treatment emergent SAEs the participants in the SQUIRE trial experienced. NICE and the ERG requested details about these from the company. The company referred to detailed information about SAEs available in the CSR in its response to this clarification question (clarification response A14). The ERG has reproduced the SAEs reported in the CSR that occurred in ≥2% of patients in the GCis + N arm and where there was a ≥1% difference between arms in Table 21. The GCis + N arm had higher rates of anemia, pulmonary embolism and vomiting than the GCis arm.

A number of adverse events were classified as being of special interest (not defined) in the CS and these are presented as composite categories in CS Tables 40, 41, 42 and 43. The ERG presents a summary of thromboembolic events for any grade and grade ≥3 in Table 22.

Table 20 Summary of treatment-emergent adverse events

	GCis + N	GCis

	Chemotherapy phase N = 538	Maintenance phase N = 275	Chemotherapy phase N = 541
At least one TEAE, all grades, %	99.1	77.5	97.8
TEAEs with outcome of death including fatal cases of disease progression,%	9.3	5.8	10.5
TEAEs with outcome of death excluding fatal cases of disease progression, %	5.9	3.6	6.8
Patients with ≥1 treatment emergent SAE, %	42.6	17.1	37.5
Patients with ≥1 Grade ≥3 TEAE	67.7	28.7	61.6
Discontinued study drug due to ≥1 TEAE, n (%)	74 (13.8)		80 (14.8)
Discontinued at least one of the combination treatments due to AEs, n (%) [a]	[redacted] (31)		[redacted] (25)

GCis: gemcitabine + cisplatin; GCis + N: necitumumab + gemcitabine + cisplatin
[a] from CS p. 123 and CSR section 12.3.3.1

Table 21 Treatment emergent serious adverse events

[redacted], n (%)	GCis + N N = 538	GCis N = 541
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

GCis: gemcitabine + cisplatin; GCis + N: necitumumab + gemcitabine + cisplatin

Venous thromboembolic events were experienced more frequently in the chemotherapy phase in those treated with GCis + N than GCis alone for any grade (8.2% vs [redacted] respectively) and ≥ grade 3 (4.3% vs [redacted] respectively). Rates of arterial thromboembolic events in the chemotherapy phase were more similar between groups. The CS states there were no differences between treatment groups with respect to fatal thromboembolism (arterial or

venous) (<1% in both groups). Other adverse events of special interest that were experienced more frequently in the GCis + N group than GCis group alone were rashes, hypomagnesaemia, and conjunctivitis (Table 22). Rates of haematological toxicities were similar between the groups.

Treatment emergent adverse events were also reported in CS Table 38 for grade 3, grade 4 and grade 5 events. These were for the overall safety sets (for the GCis + N group including the maintenance phase). The events presented included the haematological toxicities, rash, hypomagnesemia and fatigue (as presented in CS Tables 42 and 43) and other adverse events of asthenia, pulmonary embolism, nausea and vomiting. For the events that were also reported in CS Tables 42 (and CS Table 43 for the GCis group) the number of events of grade 3, 4 and 5 do not correspond. The ERG considers the data from CS Tables 42 and 43 as accurate as we have checked these data against the CSR.

EGFR expressing subgroup

Adverse events for the EGFR expressing subgroup were provided by the company in clarification response Appendix 1. The rates of AEs in the EGFR-expressing subgroup generally reflect those seen in the ITT population (reported above) and as such are not reproduced here. Exceptions are that in the GCis group the rates of any grade hypomagnesaemia are higher in the EGFR expressing population [redacted] than in the ITT population (15.7%); and rates of any grade rash for both treatment groups in the EGFR expressing group are lower than in the ITT population (GCis + N [redacted] versus 76.2%, GCis [redacted] versus 10.2%). The reasons for these discrepancies are unclear. The company also provided AE results for the EGFR expressing Western European population in clarification response Appendix 1 (not shown here).

Table 22 Adverse events of special interest

Adverse event of specicial interest experienced in at least one participant unless stated, n (%)	GCis + N				GCis	
	Chemotherapy phase N = 538		Maintenance phase N = 275		Chemotherapy phase N = 541	
	Any Grade	≥Grade 3	Any Grade	≥Grade 3	Any Grade	≥Grade 3
Thromboembolic events						
Venous thromboembolic events[a]	44 (8.2)	23 (4.3)	[redacted]	[redacted]	29 (5.4)	14 (2.6)
Arterial thromboembolic events[a]	23 (4.3)	16 (3.0)	[redacted]	[redacted]	21 (3.9)	11 (2.0)
Haematologic toxicity events						
Neutropenia	235 (43.7)	131 (24.3)	[redacted]	[redacted]	248 (45.8)	149 (27.5)
Febrile neutropenia	4 (0.7)	3 (0.6)	[redacted]	[redacted]	8 (1.5)	7 (1.3)
Anaemia	216 (40.1)	56 (10.4)	[redacted]	[redacted]	248 (45.8)	59 (10.9)
Thrombocytopenia	116 (21.6)	55 (10.2)	[redacted]	[redacted]	146 (27.0)	58 (10.7)
Fatigue	219 (40.7)	37 (6.9)	[redacted]	[redacted]	230 (42.5)	38 (7.0)
Skin reactions						
Rash[b]	405 (75.3)	30 (5.6)	[redacted]	[redacted]	55 (10.2)	2 (0.4)
Hypomagnesaemia	162 (30.1)	48 (8.9)	[redacted]	[redacted]	85 (15.7)	6 (1.1)
Hypersensitivity / infusion-related reactions	8 (1.5)	2 (0.4)	[redacted]	[redacted]	11 (2.0)	0
Conjunctivitis[c]	30 (5.6)	0	[redacted]	[redacted]	12 (2.2)	0
Interstitial lung disease (pneumonitis)	4 (0.7)	1 (0.2)	[redacted]	[redacted]	4 (0.7)	3 (0.6)

GCis: gemcitabine + cisplatin; GCis + N: necitumumab + gemcitabine + cisplatin
[a]Events experience in at least 2 participants
[b]The category of 'Rash' is a subset of the category 'Skin Reactions'.
[c]Conjunctivitis equates to 'eye disorders'

3.4 Summary

The systematic review of direct evidence in the CS identified one trial comparing GCis + N to GCis as a first-line treatment for patients with metastatic (stage IV) squamous NSCLC (the SQUIRE trial). The trial did not include patients with locally advanced disease (stage III). The CS reports OS and PFS results for both the ITT population and a post-hoc subgroup analysis of patients from Western Europe, as well as other results for the ITT population only. The company also provided results from the SQUIRE trial of post-hoc subgroup analyses of patients with EGFR expressing tumours from the ITT population and Western European subgroups in response to NICE and the ERG's clarifications questions, to reflect the population specified in the indication for necitumumab in the SmPC (clarification response A1). In the CS, the company argues that the Western Europe subgroup is more generalisable to patients in England than the ITT population. The company used data from the Western Europe EGFR expressing subgroup in their updated economic model submitted with the clarifications response. The SQUIRE trial was of a reasonable quality, although there is a risk of performance and detection bias due to lack of blinding of participants, care providers and outcome assessors.

The CS also presents an NMA comparing GCis + N with some of the scoped comparators.

The SQUIRE trial showed that GCis + N resulted in statistically significant greater improvements than GCis in OS and PFS in [redacted] in the total ITT population. Objective response rates did not differ significantly between the trial arms in the ITT population and [redacted]. The CS states that HRQoL was similar between treatment arms over time during the trial in the ITT population, although limited HRQoL data are presented. HRQoL results were not provided for the EGFR expressing subgroup. In the ITT population, the proportion of patients experiencing at least one serious adverse event was marginally higher during the treatment phase with GCis + N than during treatment with GCis. Venous thromboembolic events of any grade were experienced more frequently in those treated with GCis + N than GCis alone. The GCis + N group also experienced rashes, hypomagnesaemia, and conjunctivitis more frequently than the GCis group. In the EGFR expressing subgroup from the ITT population, rates of AEs were similar to those reported for the total ITT population, although, in the GCis group rates of any grade hypomagnesaemia were higher in the EGFR expressing group than in the ITT population. In addition, rates of any grade rash for both treatment groups in the EGFR expressing group appeared to be lower than in the ITT population. The reasons for this are unclear. Subgroup

analyses suggest that GCis + N has little benefit for people without EGFR expressing NSCLC (H-score = 0).

The ERG considers that the main systematic review in the CS based on the SQUIRE trial matches the decision problem, but that the treatment effect estimates are subject to some bias. The company's interpretation of the evidence is not fully appropriate and justified. The ERG has identified the following concerns and uncertainties:

- The patient population in the SQUIRE trial does not fully match the SmPC for GCis + N, which states GCis + N is indicated only for patients with EGFR expressing squamous NSCLC. The company supplied a post-hoc subgroup analysis for this population in response to NICE's and the ERG's clarification questions.
- The results for the EGFR expressing subgroup do not match those reported in an FDA document[14] for this subgroup. [redacted].
- The OS and PFS results from the EGFR expressing, Western Europe and EGFR expressing Western Europe subgroup analyses are at risk of bias, as these were post-hoc analyses.
- The company does not present a clear rationale for why the Western Europe subgroups are considered more relevant to the UK than the ITT population. The company did not demonstrate a significant treatment interaction for this subgroup. The company's rationale for excluding patients from Eastern Europe from the subgroup considered relevant to the UK is not convincing. The company has also not provided a clear rationale for why patients from other regions that may be similar to England (e.g. North America) were not included in the subgroup considered relevant to England. Additionally, there was an imbalance between the trial arms in ECOG performance status in the Western Europe subgroup at baseline which may have marginally favoured GCis + N. The ERG considers that the EGFR expressing subgroup from the ITT population is the most appropriate population on which to base efficacy conclusions, as this is the licensed indication. Participant baseline characteristics were balanced in the EGFR expressing subgroup between treatment arms.
- It is uncertain if the OS benefits associated with GCis + N compared to GCis in the SQUIRE trial are clinically meaningful, as the company did not define what a clinically meaningful improvement would be. The company only commented that the OS benefit in

the Western Europe group was clinically significant. Clinical expert advice to the ERG is that the improvement in OS in the EGFR expressing subgroup (the population most relevant to the licensed indication) was clinically meaningful.

- Direct efficacy data (i.e. from the SQUIRE trial) were only available for patients with stage IV disease. It is unclear whether the relative effectiveness of GCis + N and GCis would differ for patients with stage III and IV disease.

The company's systematic review conducted for the NMA of OS and PFS identified evidence to enable comparisons of GCis + N against PCarbo, GCis, PCis, DCis and GCarbo, but not against other comparators specified in the final scope and decision problem (DCarbo and VCarbo). A comparison with VCis could only be made for secondary OS median data analyses. All the studies included in the NMA were assessed by the company as having a high risk of bias on at least one quality assessment domain. The NMA showed that median OS was improved when patients were treated with GCis + N compared with GCis and PCis, but not PCarbo, DCis or GCarbo. PFS was longer with GCis + N compared with GCis, PCis and GCarbo, but not the other included comparators.

The ERG also considers that the NMA treatment effect estimates are highly uncertain for the reasons that follow:

- The company appears to have made some inappropriate post-hoc exclusions of studies.
- As acknowledged in the CS, the OS and PFS efficacy estimates from the NMA (which are used in the economic model) are uncertain due to the sparse evidence available to inform the network.
- The limited evidence base meant that: only fixed effects models could be analysed, resulting in narrow credible intervals; only unadjusted analyses could be estimated; only two of the eight pre-planned sensitivity analyses could be undertaken; and the appropriateness of the proportional hazards assumptions could not be determined for all studies in the network.
- Based on limited participant baseline characteristics provided by the company, the ERG does not agree with the company's statement on CS p. 89 that covariates were similar across studies in all but two studies.
- The company did not supply information about the length of follow-up in the studies included in the NMA, so it is unclear if this is comparable across the studies (i.e. that they are similar enough to combine).

- Most of the comparisons were based on indirect evidence, so consistency with direct evidence could not be assessed.
- The NMA was based on mainly potentially underpowered subgroup analyses.

4 COST-EFFECTIVENESS

4.1 Overview of company's economic evaluation

The company's submission to NICE includes:

1. a review of published economic evaluations of GCis + N compared with GCis for previously untreated patients with locally advanced or metastatic squamous NSCLC. (Section 5.1 of the CS, p. 139).
2. a report of an economic evaluation undertaken for the NICE STA process. The de novo model estimates the cost-effectiveness of GCis + N compared with GCis, GCarbo, PCarbo and DCis for previously untreated patients with locally advanced or metastatic squamous NSCLC eligible for first-line treatment. (CS, p. 142).

At the clarification stage, the company submitted a revised version of the de novo economic evaluation in order to be consistent with the SmPC indication (patients with advanced or metastatic EGFR expressing squamous NSCLC tumours). The clarification response included revised tables and figures relating to model inputs and results for the Western European population with EGFR expressing tumours (clarification response Appendix 1). A revised version of the executable Excel model was also submitted. The description, critique and analysis presented below is based on this revised version of the CS model.

4.2 Company's review of published economic evaluations

A systematic literature review was conducted by the company to identify studies that assessed the cost-effectiveness of GCis + N compared to GCis. The search was conducted using Embase, Medline (including Medline-R In-Process), EconLit and the NHS EED databases. Inclusion and exclusion criteria are reported in Table 47 (p. 139) in the CS, which is adapted below in Table 23 with modified inclusion criteria consistent with the text on p. 139 of the CS. The company conducted initial searches that included a wide population of all NSCLC patients, but at full text screening this population was narrowed to studies having <80% of the popu having adenocarcinoma or non-squamous histology. This narrowing of the population was included in the company's Table 47 (p. 139) where inclusion and exclusion criteria are listed.

Table 23 Cost-effectiveness review eligibility criteria

Parameter	Inclusion Criteria	Exclusion Criteria
Population	Previously untreated NSCLC patients with less than 80% combined adenocarcinoma and non-squamous histology.	Small cell lung cancer patients, non-lung cancer patients (mesothelioma), previously treated patients
Intervention	GCis + N	
Comparator	GCis	
Outcome	Cost per QALY gained, Cost per LY gained	
Study Design	Economic Evaluations (cost-effectiveness analyses, cost-utility analyses, cost-benefit analyses and cost-minimisation analyses)	RCTs, observational data, Budget Impact Assessments

GCis, gemcitabine plus cisplatin; GCis + N, necitumumab in combination with gemcitabine plus cisplatin.
Note: Table adapted from CS Table 47, p. 139.

After de-duplication, 718 potentially relevant studies were identified by title and abstract screening. Of these, 44 were identified for full-text screening. Full-text screening was conducted in two iterative passes. In the first pass, studies assessing any NSCLC population were included, whilst in the second pass only studies with less than 80% of patients who had adenocarcinoma or non-squamous histology were included. At first pass 17 studies were included. The second pass reduced this number to 10 studies.

There were 503 studies excluded for study design (19 at full text-screening), 107 for 'not first-line treatment' (6 at full-text screening), and 96 for study population (7 at full-text screening). There were also two papers that were excluded due to being duplicate studies of those found in a grey literature search. No methods or results were reported for this grey literature search. One of the excluded studies, Brown et al. 2013[19] was frequently cited in the model for resource use and cost data.

The 10 papers that were identified for extraction were assessed using the Drummond checklist (reported in Appendix 10 of the CS). None of the studies was deemed suitable for the NICE decision problem and a *de novo* cost-utility analysis model was constructed.

After completion of the systematic review, an additional economic evaluation by Goldstein et al. 2015[20] was identified. This paper reported an economic evaluation of GCis + N compared to GCis using USA Medicare reimbursement rates. The company expressed concern over the generalisability of Goldstein et al. 2015 to a UK NHS context, and did not discuss or critique it

further. The ERG agrees that the cost and cost-utility results from Goldstein et al. are unlikely to be applicable in the UK. However, Goldstein et al. was a well-conducted economic evaluation relevant to the necitumumab scope, and the modelled estimates of clinical benefits (life years gained and QALYs gained) for GCis + N compared with GCis do provide a means of cross-validating the results of the company economic model: see section 4.3.10.4 below for further discussion.

The primary limitation of the systematic review of cost-effectiveness evidence is the potential restrictiveness of the exclusion criteria with regards to study design. The inclusion criteria do not explicitly include cost-consequence analyses, a type of economic analysis where the results are disaggregated. The exclusion criteria also specifically exclude RCTs, observational studies, and budget impact analyses. Economic evaluations that are conducted alongside trials and observational studies are often cost-consequence analyses. Additionally, many budget impact analyses are equivalent to cost-minimisation analyses. This lack of clarity on inclusion and exclusion criteria for study design leaves the 503 studies excluded for study design open to question.

4.3 Critical appraisal of the company's submitted economic evaluation

4.3.1 NICE reference case

The NICE reference case requirements have been considered for critical appraisal of the submitted economic evaluation in Table 24.

The CS has a number of inconsistencies and potential issues with regard to the NICE reference case. The model population in the CS is consistent with the NICE Reference case, but not with the SmPC. However, the company submitted a revised model and results for the licensed population as part of their clarification response (Clarification Response A1 Appendix 1). I comparators were included in the NMA, and therefore could not be included in the model see section 3.1.7 for discussion of the NMA). Studies reporting results for vinorelbine platinum doublets were excluded, and data were not available for docetaxel plus carboplatin. The r also excluded PCis, which was included in the NMA, as the company argued that this combination is used rarely in NHS practice (Clarification Response B3, p. 11).

Table 24 NICE reference case requirements

NICE reference case requirements:	Included in submission	Comment
Decision problem: As per the scope developed by NICE	Yes	The SmPC indication is narrower than the NICE scope or CS model, but the revised model submitted with the company's clarification response is consistent with the SmPC.
Comparator: As listed in the scope developed by NICE	?	Vinorelbine plus cisplatin, a comparator listed in the scope, was only included in the NMA for median OS, and therefore could not be included in the economic model (*Discussed in section 4.3.4*). PCis was also excluded from the model, as the company argued that it is little used in UK practice.
Perspective on costs: NHS and PSS	Yes	
Evidence on resource use and costs: Costs should relate to NHS and PSS resources and should be valued using the prices relevant to the NHS and PSS	Yes	
Perspective on outcomes: All direct health effects, whether for patients or, when relevant, carers	Yes	
Type of economic evaluation: Cost utility analysis with fully incremental analysis	Yes	A fully incremental analysis was not reported, but the ERG has created this analysis from the model.
Synthesis of evidence on outcomes: Based on a systematic review	Yes	
Time horizon: Long enough to reflect all important differences in costs or outcomes between the technologies being compared	Yes	
Measuring and valuing health effects: Health effect should be expressed in QALYs. The EQ-5D is the preferred measure of health related quality of life.	?	Not all utility values were derived from EQ-5D, all adverse events were derived from standard gamble methods. *(Discussed in section 4.3.6)*
Source of data for measurement of health related quality of life: Reported directly by patients and/or carers.	?	Utility decrements for adverse events were derived from members of the general public. Health status measured by EQ-5D was derived from patients. *(Discussed in section 4.3.6)*
Source of preference data: Representative sample of the UK population	?	EQ-5D data used the UK tariff. It is unclear whether studies for adverse events have representative samples. *(Discussed in section 4.3.6)*

NICE reference case requirements:	Included in submission	Comment
Equity considerations: An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit.	No	Treatment assessed for end of life criteria.
Discount rate: 3.5% pa for costs and health effects	Yes	
Notes: ? = uncertain; N/A=not applicable		

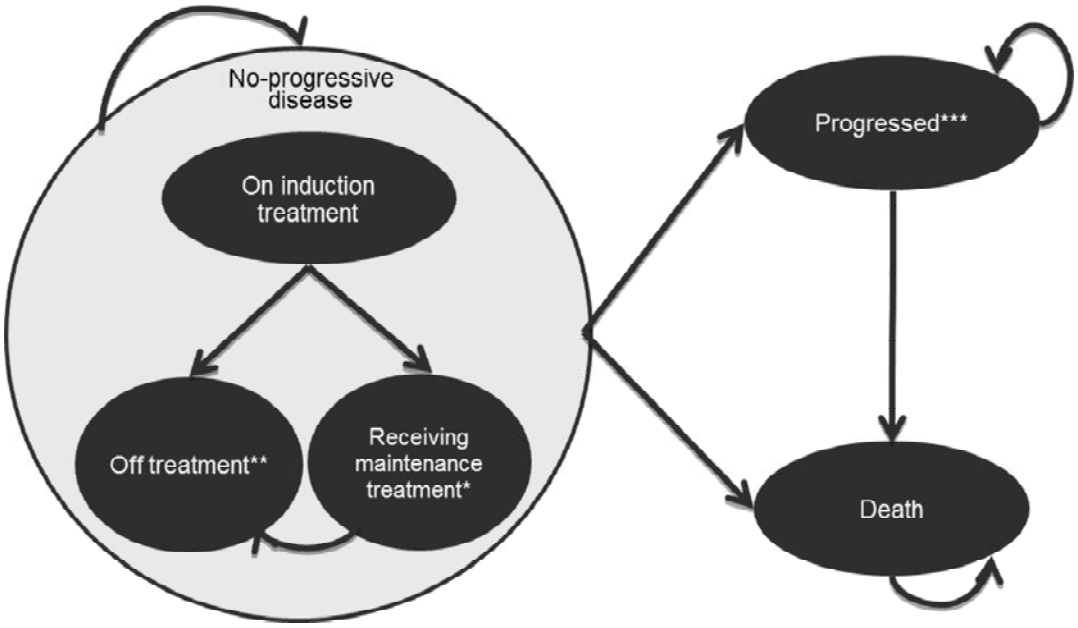
A common thread throughout Table 24 relates to quality of life measurement. The current NICE Guide to the Methods of Technology Appraisal (Methods Guidance)[21] states that utility in cost-utility analyses should be measured using EQ-5D, with patients submitting health state scores and valuation done by the general UK public. The utility values for modelled health states use EQ-5D data measured in accordance with NICE preferences, but the utility decrements used for adverse events are not in accordance with NICE preferred methods.

The company does not present a standard incremental analysis. Instead they present a series of pairwise comparisons between GCis + N and each included comparator. However, the company does report disaggregated costs and QALYs for most interventions, and the model contains full disaggregated results. We have therefore calculated fully incremental results tables (see section 4.3.8 below).

The company has used mostly appropriate methods but their analysis has a number of limitations: due to lack of available data vinorelbine doublets and PCis were excluded from the modelling; utility decrements for adverse events are inconsistent with NICE methodological guidance; and analyses are presented in a pairwise manner that obfuscates cost-effectiveness conclusions.

4.3.2 Model structure and methodological approach

The company model is a state transition model, which reflects the progress of a cohort of patients through the stages of first-line treatment and disease progression to death. The structure is illustrated in Figure 4 below.



* Patients who completed up to six cycles of first-line treatment and are receiving maintenance treatment.

** Patients who discontinued induction treatment or maintenance treatment due to AEs, or physician or patient preference.

*** At least a 20% increase in the sum of the longest diameter of target lesions or unequivocal increase in the size of non-target lesions or the appearance of one or more new lesions.

Figure 4 Company model structure

Note: This figure is a direct reproduction of CS Figure 36, p. 144.

4.3.2.1 Choice of health states and transitions

There are three main health states in the model: no progressive disease (NPD); progressed disease (PD); and death. Patients start in the NPD state at initiation of first-line treatment. Each week, patients may remain in NPD, move to PD as they develop progressive disease, or die. After progression, patients may remain in the PD state for some time but will eventually die.

Within the NPD state, there are three sub-states that reflect the process of first-line treatment: on induction treatment (NPD-induction); maintenance treatment (NPD-maintenance); and off treatment (NPD-discontinued). Patients start in NPD-induction, receiving chemotherapy in three-week cycles. With conventional chemotherapy, patients remain on induction until completion (usually after 4-6 treatment cycles), or they may discontinue early due to unacceptable adverse

effects, patient choice, because of disease progression or death. Patients who complete induction treatment with a platinum doublet move into the NPD-discontinued state, where they remain until progression or death. However, after completion of induction with GCis + N patients move into NPD-maintenance, where they continue to receive necitumumab every three weeks until discontinuation, progression or death.

This broad model structure is appropriate for the decision problem. The three main states and transitions between them reflect the progressive and usually terminal nature of advanced NSCLC, and the three sub-states and transitions are consistent with current and recommended practice for first-line chemotherapy, and with the draft SmPC use of necitumumab.

4.3.2.2 Method for estimating transitions between states

A partitioned survival (or area under the curve) approach was used to estimate the proportion of the cohort in each of the five states at each weekly cycle. The distribution of the cohort between the NPD, PD and Dead states is illustrated in Figure 5. Here the distribution is governed by two survival curves for each treatment: PFS and OS. At each time point (t), the proportion of the cohort who are dead is 1 - OS(t); the proportion in the PD state is OS(t) – PFS(t); and the proportion in the NPD state is just PFS(t).

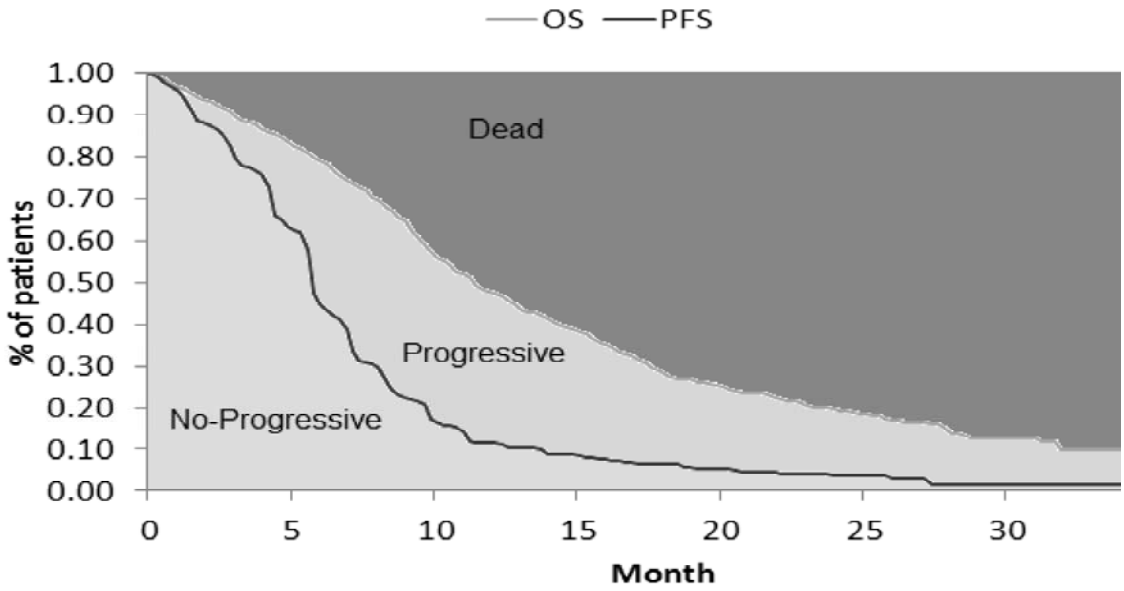


Figure 5 Illustration of health states derived from OS and PFS curves

Note: This figure is a direct reproduction of CS Figure 51, p. 164.

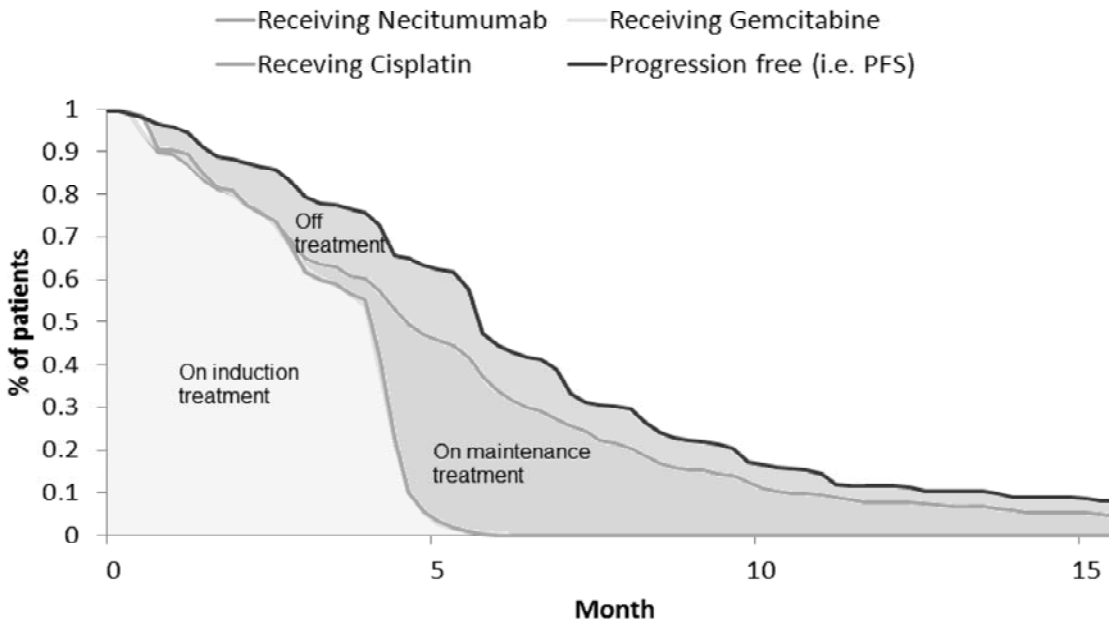


Figure 6 Illustration of sub-states for patients without progressive disease

Note: This figure is a direct reproduction of CS Figure 52, p. 164.

A similar approach was used to split the patients in the NPD state between the induction, maintenance and discontinued sub-states, as illustrated in Figure 6. In this case, the calculations are a little more complicated:

- For the GCis treatment arm, there are three survival curves to consider: PFS and the times to discontinuation of gemcitabine (TTD_g) and cisplatin (TTD_c). At any time (t), the proportion of patients in the induction phase is the proportion who have not yet discontinued both gemcitabine and cisplatin: the maximum of $TTD_g(t)$ and $TTD_c(t)$. The proportion of patients who have stopped treatment but not yet progressed is then the difference between PFS(t) and the proportion on induction.
- For the GCis + N treatment arm, there is an additional survival curve to consider: the time to discontinuation of necitumumab (TTD_n). The proportion of patients in the induction phase is still the maximum of $TTD_g(t)$ and $TTD_c(t)$. The proportion of patients on maintenance treatment is the difference between the proportion who have not discontinued necitumumab and the proportion who have discontinued both gemcitabine and cisplatin: $TTD_n(t)$ minus the minimum of $TTD_g(t)$ and $TTD_c(t)$. The proportion of patients who are off treatment but not yet progressed is the difference between PFS(t) and the proportion on induction or maintenance.

Table 25 Distribution of cohort between states

Health State	Methodology
NPD-induction	Derived from the survival curves for TTD for each treatment arm and treatment compound. Defined as the maximum of the proportion of patients receiving gemcitabine or cisplatin and less or equal to the proportion of patients who were progression-free.
NPD - discontinued	Patients that remain progression-free and are not receiving maintenance or induction treatment. This is calculated by subtracting the proportion of patients on treatment from the proportion of patients in the progression-free disease state.
NPD - maintenance	The proportion of patients on maintenance treatment was estimated as the proportion of patients on treatment minus those on induction treatment. By definition, the proportion on maintenance treatment was zero for the platinum doublet treatments.
PD	All patients surviving (OS) minus those who remain progression-free (OS-PFS).
Death	(1-OS)

Note: Adapted from CS Table 53, p. 165.

This partitioned survival method is convenient for conventional three-state cancer models, as it uses survival results that are usually reported anyway for clinical purposes (PFS and OS). As seen above (Table 25), the method can be extended for models with more than three states,

provided they are acyclic (patients cannot return to a state that they have left). There is some controversy over the validity of Partitioned Survival compared with the more usual Markov method, in which direct estimates of transition probabilities are used to derive the flow of a cohort between states. [22 23] However, a recently published economic evaluation of nivolumab for the treatment of second-line advanced squamous NSCLC found very similar results with Partitioned Survival and Markov models.[24]

The ERG concludes that the Partitioned Survival method is theoretically reasonable, and that there is no evidence to believe that different results would have been found with an equiv Markov model. Of course the ability of the model to reflect real-life patient flows depends data and methods used to fit and extrapolate the various survival curves. These data and methods are described and critiqued below: for OS see section 4.3.5.1, PFS section 4.3.5 and treatment discontinuation section 4.3.5.3. We also assess the consistency of the mod times on treatment, PFS and OS in comparison with SQUIRE results and external eviden (section 4.3.10).

4.3.2.3 Treatment duration

The protocol for the SQUIRE trial specified up to six, three-week cycles of treatment for induction therapy with GCis and with GCis + N. The duration of treatment was modelled using KM survival curves from SQUIRE. In this study, some patients stopped before six cycles, spending less than 18 weeks on induction therapy, while others spent longer than 18 wee induction, due to delays in administration and 'treatment holidays'. To avoid overestimati costs for the latter group, the model included an adjustment for the intensity of treatment incurred within each three-week period. See section 4.3.5.4 for a discussion of how this adjustment factor was calculated. For indirect comparators, the HR of treatment discontinuation was set equal to the HR of PFS. PFS was considered the most suitable proxy for continu therapy as only progression-free patients can remain on treatment. Duration of necitumumab maintenance therapy was also modelled using KM survival curves from SQUIRE.

4.3.2.4 Adverse events associated with first-line treatment

TEAEs were not modelled explicitly as health states. This is acceptable if the model captures impacts of the TEAEs on treatment discontinuation, costs and utilities in a way that is reflective

of current NHS practice and any differences between the comparators. The impact of TEAEs on first-line treatment discontinuation was modelled via the TTD survival curves, which regulate transitions out of the NPD-induction and NPD-maintenance states. The model also incluc direct estimates of costs and disutilities for TEAEs that occurred in at least 2.5% of the tot patient population in the SQUIRE trial, and grade 3 and 4 febrile neutropenia. The freque these events for GCis and GCis + N comparators was estimated from the observed rates SQUIRE (see 4.3.5.5, p.102). For other comparators, it was assumed that the relative ris TEAEs would be the same as for GCis versus GCis + N. Assumptions were also made a the costs of treating the included TEAEs (section 4.3.7) and associated utility decrements (section 4.3.6).

4.3.2.5 Second-line treatment and palliative care

The company model does not explicitly map out subsequent treatment or palliative care a progression. One would not expect such treatments to have an effect on survival, althouç may impact on quality of life and costs. The model includes estimates of second-line treatment and palliative care costs in the PD state. Utility impacts of any treatments received after progression are implicitly incorporated in the post-progression utility. This approach is reasonable, provided that the evidence used to underpin the cost and utility estimates in the PD state are reflective of current practice in the NHS, and of any differential impacts of first-line treatment on the use or effects of treatment that patients receive after progression (see section 4.3.6 for a discussion of the PD utility data and section 4.3.7 for resource use and costs).

4.3.2.6 Other assumptions

The model includes a half-cycle correction, assuming that utilities and costs were spread throughout each week cycle. It is stated in the model ('Model design' sheet) that the half-correction was not applied to first-line treatment and administration costs, since the drugs usually administered at the beginning of the cycle, but in practice it was.

The company provided a summary of methodological and structural assumptions in their r (CS Table 74, p. 212). This is reproduced below, together with comments from the ERG 26). A summary of model parameters is provided in CS Table 73 (pp. 206 to 211). This

includes the mean values used in the base case analysis, 95% confidence intervals and distributions used in the Probabilistic Sensitivity Analysis (PSA).

Table 26 Summary of assumptions in the company model

Assumption	Description	Company justification	ERG position
Patient Population	The Western European subgroup of the SQUIRE trial is representative of the UK NSCLC patient population.	The SQUIRE trial has reported a difference in the clinical efficacy of necitumumab across regions. Statistical analysis has determined that this is not due to a difference in baseline characteristics or treatment received during the SQUIRE trial, but is likely due to potential unobserved treatment effect modifiers including difference in the disease burden associated with NSCLC and environmental causes of cancer including social and cultural practices such as heavy smoking. The literature suggests that this is likely to have resulted in higher incidence and mortality rates for lung cancer patients in Eastern Europe than in Western Europe. The unobserved treatment effect modifiers in the SQUIRE trial may have contributed to an overall varying impact on health outcomes geographically for necitumumab. Therefore, it is considered appropriate to employ data which has been generated from a patient population reflective of the disease burden of NSCLC patients in England. As a result, the economic evidence presented in this submission is reflective of the Western Europe subgroup of the SQUIRE trial as it is considered the most appropriate population for decision making regarding the NHS. The Western European subgroup of patients consists of patients from Austria, Belgium, France, Germany, Greece, Italy, Portugal, Spain and UK.	We disagree. There is no evidence of significant differences in treatment effects (OS or PFS) between geographical regions in the subgroup analyses or in the post-hoc Western Europe analysis (see Appendix 6 p. 61-2 of the CS clarification response). There was no statistically significant interaction between Western Europe and the remaining patients in the SQUIRE trial (CS p. 229).
Maximum of 6 cycles of induction treatment	It is assumed that patients receive induction treatment for a maximum of 6 cycles.	The SQUIRE clinical trial data allowed patients to continue induction treatment for a maximum of 6 cycles before initiating necitumumab maintenance treatment. While this varies from UK clinical practice in which patients typically only receive 4 cycles, it has been assumed that this discrepancy has no impact on the outcomes associated with treatment.	We agree. Up to 6 cycles is consistent with the SmPC, and expert advice suggests that 4-6 cycles of treatment is common in the UK.
BSA	The average body surface area (BSA) was considered to be 1.85 m^2. This was used to calculate the cost for all comparators.	The BSA from the trial was slightly lower than the average UK patient found in Sacco et al. 2010 for NSCLC patients.[25] Therefore, it was determined to use the BSA of the average UK NSCLC patient rather than the average BSA from the trial.	We agree with the company approach.

Assumption	Description	Company justification	ERG position
Second-line therapies	It is assumed that the second-line therapies reported in the SQUIRE trial have the same efficacy and utility as those assumed to be routinely used in UK clinical practice.	A number of the treatments used in the SQUIRE trial are not representative of UK clinical practice. However, the rates of use are similar in both arms therefore OS should not differ by arm due to second-line therapy.	This is a reasonable assumption, in the absence of evidence to the contrary. Effects of second-line treatments were not modelled explicitly, but are implicitly included in the SQUIRE results. It would therefore be difficult to model an alternative second-line treatment regimen.
Choice of second-line therapy	Second-line therapies in the model are docetaxel and erlotinib	Brown et al. 2013 assumed that docetaxel and erlotinib are used equally in the second-line setting for NSCLC patients. Positive NICE appraisals for each of these treatments support their choice as second-line therapy in the UK.	NICE Guidance has changed since Brown et al. 2013. Erlotinib is now only recommended for patients who have the EGFR-TK mutation. Expert opinion also indicated that nivolumab is used in second-line therapy.
Duration of subsequent treatment	The duration of subsequent treatment is the median duration of each therapy received in SQUIRE	To determine the number of infusions of each subsequent treatment received, the median duration was used in combination with the cycle length and number of administrations per cycle. Median duration was preferred over the mean, as many patients had not completed subsequent treatment which would result in an underestimated duration if the mean was used. Data was not available to inform use and duration of subsequent treatments for the indirect comparators. Therefore the use and duration of subsequent treatments for indirect comparators was assumed to be equivalent to those observed in the GCis arm. This assumption was tested in scenario analysis.	Although the mean would be preferred for costing, it is reasonable to use the median in this case, due to truncated follow up of second-line treatment. It is reasonable to assume similar second-line treatment patterns following other first-line platinum doublets as for GCis.
End-of-life (EOL) Care	All patients are assumed to receive 2 weeks of EOL care	EOL care is assumed to occur for 2 weeks and is can be provided at home by Macmillan Nurse, in Hospice or a Hospital according to Brown et al. 2013	We agree.

Assumption	Description	Company justification	ERG position
Utility following progression	The utility value for patients that have progressed following first-line treatment is based on Khan et al. 2015	Post-progression health state utilities were obtained from the literature as the SQUIRE trial only conducted EQ-5D assessments until disease progression. For the post-progression health state, the values by Khan et al. 2015 were used because they are values obtained during a RCT of patients that had an active treatment until progression and valued based on UK weights applied to the EQ-5D-3L.	We agree that this is the best available estimate. However, there are questions over its applicability, since the Khan et al. 2015 study included only elderly patients who were considered unfit for chemotherapy.
NMA AE	The rate of AEs for indirect comparators is set equal to AEs observed in the GCis arm.	In the base-case analysis the relative safety profile of indirect comparators versus GCis + N was assumed to be equivalent to the relative safety profile of GCis versus GCis + N. This was because the systematic literature review did not identify AE data specific to the squamous population for these comparators. To examine the impact of this assumption two extreme scenarios were tested, where the risk of AEs for all indirect comparators was set to 0 or to double that associated with the GCis + N arm.	This is a reasonable approach.

Note: Summary adapted from CS Table 74 pp. 212 to 213.

4.3.3 Population

The model submitted with the CS assesses a patient population that is generally consiste the decision problem. However, the SmPC restricts the indication to first-line treatment of advanced or metastatic EGFR expressing squamous NSCLC. The revised version of the is appropriate for this indication.

The company model relies on the SQUIRE trial for estimates of OS, PFS and TTD for GC N and GCis. As discussed in clinical effectiveness section 3.1.3 above, the population in S(was somewhat narrower than that in the decision problem, since this study only included patients with stage IV squamous NSCLC, excluding patients with stage IIIB disease. It is unclear whether the relative effectiveness of GCis + N and GCis would differ for patients v stage IIIB and IV disease.

The results of the economic model reported in the CS, and the revisions in the clarificatio response were based on analysis of the Western Europe subgroup of patients in the SQL trial. Results for the whole ITT sample were only reported in summary form as a scenaric analysis (Table 38 of the clarification response). As noted in 2.3 above, the rationale for t of this subgroup is not robust. Statistical tests did not show a significant difference in trea effects for this post-hoc subgroup or for the other pre-specified subgroups. The results fo Western Europe subgroup are more favourable to GCis + N than those for the ITT popula so it is likely to have biased the cost-effectiveness results reported by the company. Both the original and revised versions of the submitted models include data for the SQUIRE ITT population and EGFR expressing subgroup from the ITT population, respectively, and so we have conducted further analyses to estimate cost-effectiveness using these data (see section 4.4 below).

Effectiveness for comparator treatments was primarily derived from the NMA. Systematic searches were conducted to populate the NMA (see section 3.1.7) and in these searches studies were excluded based on whether results were reported separately for squamous histology. This approach is consistent with the scope.

Bottom-line summary of ERG view on patient group

The patient population for the economic model generally reflects the scope, but some studies were excluded from the NMA using post-hoc exclusion criteria (see section 3.1.7). Additional analyses including data from the patients in these studies (where studies could have connected to the network) would have been appropriate.

4.3.4 Interventions and comparators

The model includes GCis + N as the intervention, and four of the eight comparators specified in the decision problem: GCis, GCarbo, PCarbo and DCis. The modelled doses and adminstration schedules are broadly in line with UK practice and relevant SmPC criteria.

The CS omits vinorelbine platinum doublets, DCarbo and PCis from their economic evaluation. The company did not identify any evidence relating to the vinorelbine combinations to include in their NMA of OS HR or PFS OR (the NMA data used in the model). DCarbo was also omitted from the economic model, as HR estimates were not available from the NMA. As discussed in 3.1.7, the ERG considers the results from the company's NMA highly uncertain; please see that section for a summary of the key caveats regarding the NMA. The model omitted the PCis combination, for which evidence was available from the NMA. In response to a clarificatic question (B3), the company noted that market share data suggests that PCis is not typically used in clinical practice in England. The ERG has included PCis against other comparators in the sensitivity analysis in section 4.4.

4.3.5 Treatment effectiveness and extrapolation

4.3.5.1 Overall survival

In the company's base-case analysis, only GCis + N and GCis are compared. For this analysis KM data were used until end of follow-up (36 months), and then parametric regression was used to extrapolate from the last point. This approach gives greater weight to the tails of the KM curves, which are estimated from small numbers of patients, and may introduce error into the survival estimates. In the Western European patients with EGFR expressing tumours, the population used for the base case analysis, there were four patients remaining at the end of follow up.

The company reported that they had used a selection process, based on the NICE Decision Support Unit (DSU) technical support document 14,[26] to select appropriate parametric survival functions (CS p.149). They tested six functional forms: exponential, Weibull, Gompertz, log-normal, log-logistic and generalised gamma, although it was stated that the Gompertz curve was not included as it did not converge. Functions were fitted separately to the trial arms, without predictors. The use of a joint distribution, with a treatment interaction term was also considered. Methods for model selection included: assessment of goodness of fit within the observed period with Akaike information criterion (AIC) and Bayesian information criterion (BIC) statistics and plots of Cox-Snell residuals; inspection of hazard functions to assess the plausibility of the proportional hazards assumption; visual inspection of the fit against the KM function; and assessment of the tail of the parametric function in comparison with long term survival estimates.The specific methods used to assess suitability of parametric survival curves are given in Table 27.

Table 27 Methods for assessing parametric survival model suitability

Criteria	Method	Description
Observed trial period	AIC & BIC statistics	Assess the relative fit of parametric models whilst accounting for the number of parameters
	Cox-Snell residuals	Assess how closely a parametric function follows the KM function
	Kernel-smoothed hazard function	Assess the behaviour of the hazard function and the plausibility of the proportional hazards assumption
	Visual inspection	Assess how closely a parametric function follows the KM function and the clinical plausibility of the prediction in relation to other endpoints
Extrapolation period	Visual inspection	Assess how closely the tail of the parametric function fitted to the active treatment arm(s) concurs with any available external longer term data or clinically expected outcomes

Note: This table is a direct reproduction of CS Table 49, p. 149. KM, Kaplan-Meier.

The company provided diagnostic plots for the Cox-Snell residuals and kernel-smoothed hazard functions for the EGFR subgroup in their clarification response Appendix 1. The OS KM survival curve for the EGFR subgroup (ITT population) is shown in Figure 7.

Figure 7 Kaplan- Meier OS survival for EGFR expressing tumours (ITT Population)

Note: This is a direct reproduction of clarification response, Appendix 1, Figure 1

Parametric model fit was assessed by the company using AIC and BIC statistics for the EGFR expressing Western European subgroup of the SQUIRE trial (base case) (clarification response A1, Appendix 1), reproduced below (Table 28). The company also produced a variety of diagnostic plots for this subgroup, which were evaluated in the CS by visual inspection. They concluded that the proportional hazards assumption was not valid, and that the log-logistic curve was the best-fitting distribution.

Table 28 Estimates of model fit for OS in the EGFR expressing Western European subgroup

	AIC	BIC
GCis + N		
Weibull	391.893	397.847
Log-normal	392.795	398.748
Log-logistic	**385.977**	**391.931**
Exponential	395.275	398.252
Generalized Gamma	390.397	399.327
GCis		
Weibull	**416.062**	**422.149**
Log-normal	433.06	439.147
Log-logistic	**416.941**	**423.028**
Exponential	426.513	429.556
Generalized Gamma	417.003	426.134

Note: This is a direct reproduction of Table 18, clarification response Appendix 1

Diagnostic fit was not assessed in the CS for the ITT population. The ERG considers this to be inappropriate, in the absence of evidence supporting the Western European subgroup.

Based on the available diagnostic assessments for the Western European subgroup, the log-logistic curve has the lowest AIC and BIC for the GCis + N group (indicating a better fit), and it has a good visual fit. The log-logistic also provides a reasonable fit for the GCis group, although the AIC and BIC were slightly lower than for the Weibull curve, which also has a good visual fit. The hazard function plots presented in the CS to justify the rejection of the proportional hazards assumption are difficult to assess, due to the small numbers of patients remaining in the unstable portions of the graphs. Statistical tests for proportional hazards were not presented in the CS, and the analysis for the larger ITT population might have been informative.

The diagnostic statistics and curves presented are not definitive, and the visual fit was similar between the log-logistic, Weibull, and generalised gamma distributions. The choice of curve should also be predicated on clinical plausibility. The log-logistic curve has a heavy tail, so predicts that a proportion of patients survive for a long time. This may be questionable for the stage IV NSCLC population in the SQUIRE trial. Relative expected survival from Cancer Research UK shows stage IIIB patients having a 5-year survival rate of around 6.32%,[27] whilst

extrapolation using the company methods estimates 5-year survival of 7% for GCis + N and 2% for GCis. Estimates from United States SEER data indicated that patients with stage IV disease have expected 5-year survival of approximately 1%.[28] Comparator trials used a mixed population of stage IIIB and stage IV patients, so it is likely that expected survival for thes comparators at five years is between 1% and 7%.

Trial KM estimates of OS with alternative extrapolations (log-logistic and Weibull curves) are shown in Figure 8**Error! Reference source not found.** for the Western European subgrc with EGFR expressing tumours. It can be seen that the area between the curves (the esti life years gained from GCis + N compared with GCis) is greater when log-logistic curves are used for extrapolation than when Weibull curves are used. Note that the start point of the parametric extrapolations (which are fitted on the whole KM dataset) are adjusted to meet the final KM endpoints, which are estimated from a small number of patients remaining in the analysis at that time.

In the analysis that included the remaining treatment regimens (PCarbo, PCis, GCarbo and DCis) OS was modelled using HR derived from the fixed-effects NMA with GCis + N used as the baseline survival curve. To enable comparison via NMA, proportional hazards were assumed – indicating that non proportional hazards survival functions (such as the log-logistic) would not be appropriate. Of the candidate proportional hazards survival curves tested, as described above, the company selected the Weibull curve, as it is the best fitting curve that can be used in a proportional hazards model, but with reservations on whether proportional hazards assumptions were justified. The ERG does not find fault with assuming proportional hazards. Additionally, since GCis + N is the new treatment, it is easier to interpret its comparative effectiveness if GCis is used as the baseline survival curve.

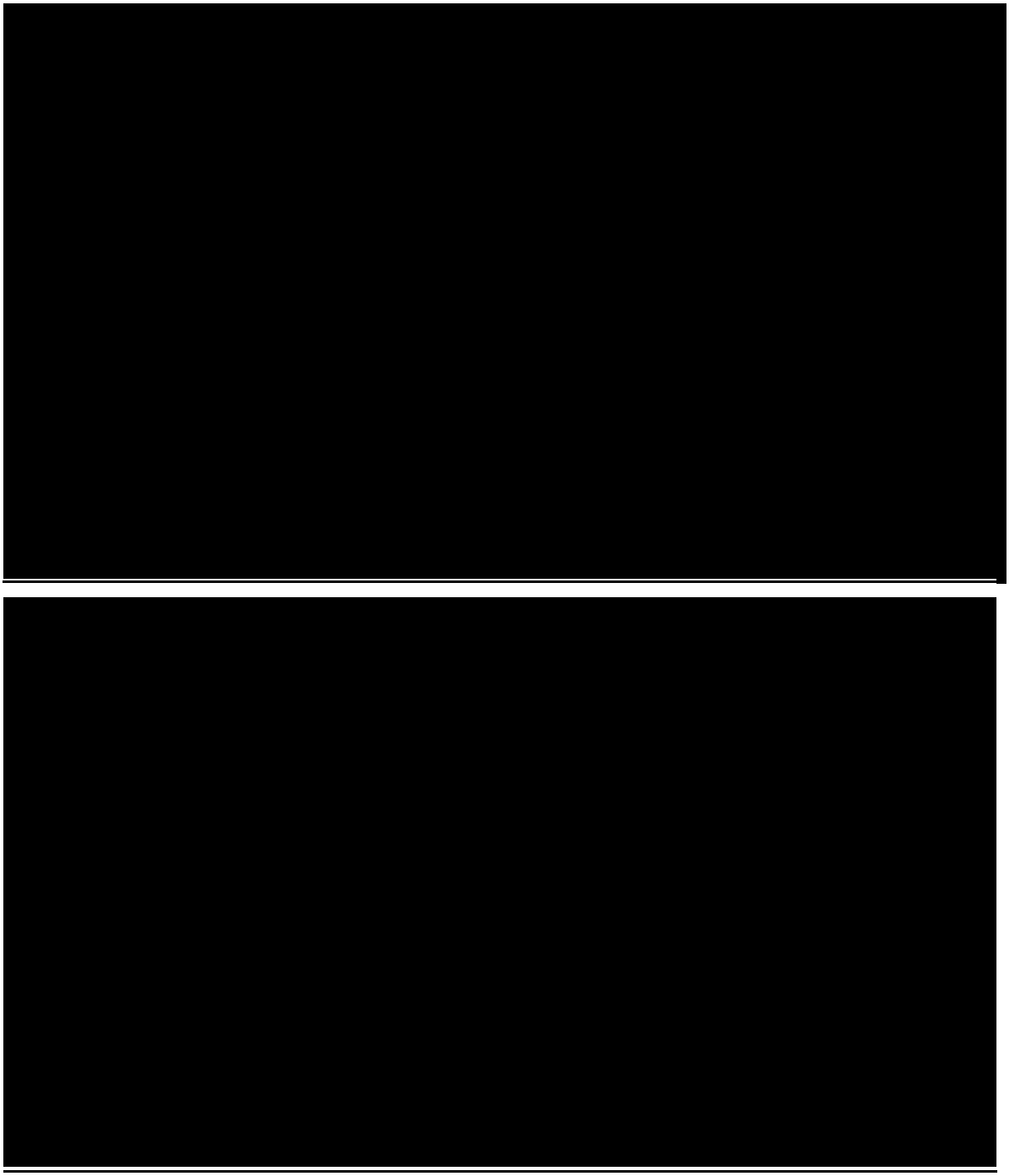


Figure 8 SQUIRE OS KM and fitted curves; Western European subgroup with EGFR expressing tumours

4.3.5.2 Progression free survival

Disease progression was based on PFS, which was defined by the time from randomisation until first radiographic documentation of objective progression, or death from any cause. PFS for GCis + N and GCis was estimated based on KM data from SQUIRE. Extrapolation using a log-logistic parametric survival curve was used for the remainder of patients who had not progressed at the end of the trial (4% of patients). The log-logistic curve was chosen in the same manner as for OS (Table 27), diagnostic plots for model fit in EGFR expressing tumour patients were reported in the Clarification Response (Appendix 1). The PFS survival curve for the EGFR subgroup (ITT population) is provided in Figure 9.

The PFS curve used in the base case of the company EGFR expressing model is given ir Figure 10. The ITT curves for the EGFR expressing population appear to converge earlie the curves and extrapolation for the EGFR expression Western European subgroup used company model, and the overall PFS benefit for GCis + N compared to GCis appears sm PFS in the original model submitted by the company (which includes patients without EGFR expression reported) was externally validated against Hoang et al. 2013.[29] In Hoang et al. 2013 GCis patients had an expected PFS time of 4.3 months (3.3 - 6.6 months), which is similar to the estimate of 4.37 months from the separately fitted log-logistic curve for GCis using SQUIRE data.

4.3.5.3 Time to treatment discontinuation

A KM analysis was conducted on the safety population of the SQUIRE trial (all patients who received at least one dose of the study drug) to estimate time to treatment discontinuation. No parametric estimation was necessary as >99% of patients had discontinued treatment at the end of the trial.

As none of the trials included in the NMA reported treatment discontinuation specific to the squamous population, the HR of treatment discontinuation was assumed to be the same as the HR of PFS. The company states that clinical experts were consulted to validate this assumption.

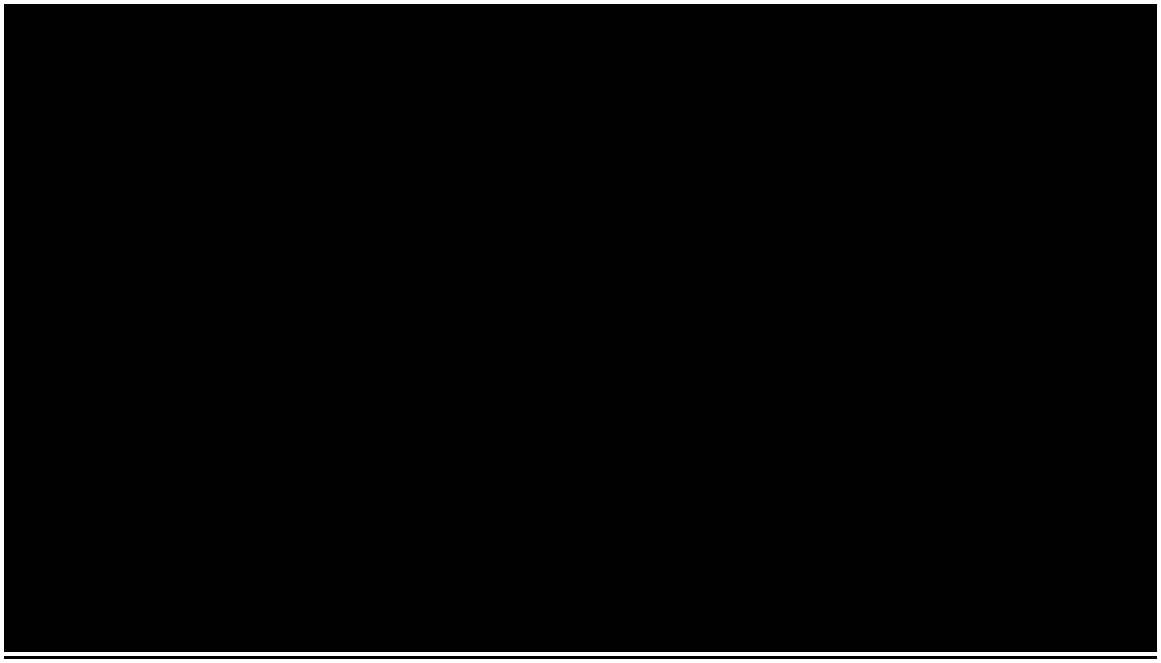


Figure 9 Kaplan-Meier PFS for EGFR expressing tumours (ITT population)

Note: This figure is a direct reproduction of clarification response, Appendix 1, Figure 3

Figure 10 Kaplan-Meier and extrapolation of PFS in patients with EGFR expressing tumours (Western European subgroup)

4.3.5.4 Treatment duration adjustment

Because patients experienced delays in treatment due to AEs or personal reasons, a proportion of patients remained on induction therapy beyond 18 weeks. Without adjustment the time on treatment is overestimated: 15.5 days of treatment for GCis + N compared to 13.6 observed, and 8.96 days of treatment for GCis vs 7.4 observed. To adjust for this, ratios were applied to the costs for the drugs in the model, 0.88 for GCis + N and 0.83 for GCis. Treatment intensity was assumed identical to GCis for all other comparators. Where treatment holidays are due to adverse events, treatments with lower adverse event rates could be made less expensive by this assumption. Expert advice indicated that carboplatin based regimens have less toxicity.

4.3.5.5 Adverse events

The incidence and duration of AEs for GCis + N and GCis were derived from the SQUIRE trial. Treatment emergent adverse events were included if they occurred in 2.5% or more of the total patient population. Because febrile neutropenia is considered to have important cost and utility consequences, it was also included in the analysis. Probilities of adverse event occurrence were converted into one week probabilities, adjusted for when they occurred (induction or maintenance). Induction corresponded to 13.6 weeks for GCis + N and 12.8 weeks for GCis. Maintenance continued until 21.7 weeks for GCis + N.

Apart from the SQUIRE trial, none of the studies included in the NMA reported AEs specifically for the squamous population. Therefore, the company assumed that the relative risk of adverse events for comparators was the same as the relative risk for GCis versus GCis + N. The company stated that this assumption was validated with clinical experts. Given that no studies included in the NMA compared AEs in squamous and non-squamous patients, there is insufficient data to conclude that histology drives AEs. It might therefore have been more appropriate to consider all evidence on AEs, independent of histotype. The RCT evidence base for AEs for the comparator treatments is substantial, as nearly every trial reports safety data.

As the company assumes that adverse event rates are identical for GCis, GCarbo, PCarbo, and DCis, the per cycle calculations in the model should produce identical results. They do not. There are small discrepancies in the rates of adverse events, and their utility decrements and costs. Between GCis and the other comparators, the rate of hyponatraemia is lower in GCarbo, PCarbo and DCis. However, the utility decrements per cycle are higher for GCarbo, PCarbo, and DCis. Similarly, AE costs are higher in GCarbo, PCarbo, and DCis

Bottom-line summary of ERG view on clinical effectiveness

Primary clinical effectiveness data for GCis + N and GCis was derived from the good quality SQUIRE RCT.[8] The approach taken by the company was generally appropriate, with the important caveat of the use of the Western European subgroup, rather than the ITT dataset.

In order to extend the model to a lifetime horizon some extrapolation was required. The method of extrapolation, modelling from the last follow-up observation, emphasises the tail of the KM curve based on sparse data, and appears to favour GCis + N. The choice of the log-logistic curve for extrapolation had good fit, but may not be the most clinically plausible curve for OS, as it predicts 7% OS at five years in a stage IV NSCLC population, which is higher than estimated 6.32% 5-year survival for stage IIIB patients from Cancer Research UK using data from the East Anglia Cancer Network.[27] Data for stage IV patients was not available in the East Anglia Cancer Network from Cancer Research UK, but SEER data from the United States estimated five year survival for stage IV at approximately 1%.[28] However, trials for other comparators included mixed stage IIIB and stage IV populations, therefore the 5-year OS in the stage IIIB and stage IV NSCLC patients should be expected to be between 1% and 7%.

Evidence for the effectiveness of GCarbo, PCarbo and DCis were derived from the NMA via HRs. For these analyses a proportional hazards assumption was used, with Weibull curves fitted to GCis + N and GCis survival data from SQUIRE, and hazard ratios applied in comparison to GCis + N.

4.3.6 Health related quality of life

Quality of life data enters the model as utility scores assigned to the following health states: NPD-induction, NPD-maintenance (this health state only applies to GCis + N), NPD-discontinued, PD, and death. Death has a utility score of 0 by definition. Utility decrements were used to represent the effects AEs have on patient quality of life. The company derived quality of life data from the SQUIRE trial and from a systematic review of the literature. SQUIRE collected EQ-5D-3L utility data from patients at baseline, at each chemotherapy session (approximately every three weeks) and then every six weeks after discontinuation of treatment for each treatment arm. The systematic review identified further quality of life data.

Utility scores for the NPD-induction, NPD-maintenance (this health state only applies to GCis + N), and NPD-discontinued health states were derived from the SQUIRE trial.{Thatcher, 2015 #29} Utility data was pooled between the treatment arms during the induction treatment phase, and after discontinuation. All utilities from the SQUIRE trial are in accord with NICE preferred methods for utility measurement in economic evaluations: they use EQ-5D, health status reported by patients, and valuation from the UK general public (UK EQ-5D-3L tariff).[21]

Table 29 Criteria for company systematic review of HRQoL

	Inclusion criteria	Exclusion criteria
Population	Adult patients with metastatic or advanced NSCLC	Small-cell lung cancer; not advanced or metastatic; stage I, II, III only
Intervention	Not restricted	
Comparator	Not restricted	
Outcomes	EQ-5D SF-36 SF-6D SF-12 HUI2 HUI3	Any measurement of health-related quality of life not converted to utility values
Study Design	Interventional and observational studies	Non-human, pre-clinical studies; case reports; studies exclusively sourcing secondary data (i.e. review articles, meta-analyses, economic models)
Language	English	Non-English
Date	2000 onwards [b]	Prior to 2000 [b]

Notes: This tabe is a direct reproduction of CS Table 55, p. 168. SF-36:Short Form 36 Health Survey; SF-6D: Abbreviated Short Form 36 Health Survey; SF-12: 12-Item Short Form Health Survey; HUI2: McMaster Health Utilities Indexes Mark 2; HUI3: McMaster Health Utilities Indexes Mark 3

[a] After January 2010 in accordance with the release of the American Joint Committee on Cancer (AJCC) Staging Manual, 7th edition, stage IIIb with pleural effusion was upgraded to stage IV cancer.[30] Therefore, articles published prior to January 2010, or articles published after January 2010 but with reference to earlier data and/or methodologies will be included if referencing Stage IIIb with pleural effusion and Stage IV. Later articles will only be included if referencing Stage IV.

[b]Abstracts published prior to the year 2013 were excluded.

The utility score for progressive disease and all utility decrements for adverse events in the CS were identified through a systematic review of the literature. The systematic review searched the following databases: PubMed, EMBASE, MEDLINE, Cochrane Library, NHS Economic Evaluation Database (NHS EED), Cost-effectiveness Analysis Registry from the Centre for the Evaluation of Value and Risk in Health, and EconLit. The inclusion and exclusion criteria are

reported in Table 29. Health state utility scores derived from SQUIRE and the systematic review of HRQoL are presented in Table 30.

Table 30 Utility values used in the CS economic model EGFR expressing (Western European subgroup)

Health state or adverse event	Mean	Lower CI	Upper CI	Distribution in PSA	Source
Utility scores for health states in the economic model					
Pre-progression and on induction treatment	[redacted]	[redacted]	[redacted]	Beta	SQUIRE
Pre-progression and off treatment	[redacted]	[redacted]	[redacted]	Beta	SQUIRE
Pre-progression and receiving maintenance treatment	[redacted]	[redacted]	[redacted]	Beta	SQUIRE
Relapsed progressive disease	0.55	0.52	0.58	Beta	Khan et al. 2015

Note: Data in this table were derived from Table 21 in clarification response Appendix 1.

The systematic review of HRQoL publications identified 833 references, of which 27 were selected for data extraction. Of these 27 studies, three studies provided utility scores for a squamous NSCLC population.[31-33] All three of these studies used EQ-5D-3L data. The three studies were quality assessed according to the NICE Technical Support Document for assessment of health state specific utility studies.[34]

Four of the studies identified in the systematic review of HRQoL included information on disutilities related to adverse events. The company assessed the disutilities for these studies for inclusion in the economic model. Utility decrements were selected from two of the four studies, Nafees et al. 2008[35] and Doyle et al. 2008.[36] Neither of these studies follow NICE Reference Case preferred methods. In the current NICE Guide to the Methods of Technology Appraisal[21] preferred methods are provided for measuring HRQoL: the preferred preference based utility questionnaire is the EQ-5D, health status data should be collected from patients, and health state valuations should be from the UK general public. Where EQ-5D data is not available, and validated preference-based utility scores are not used, NICE prefers that utility data be collected using the time trade-off method. Where these methods are not available, NICE recommends the

use of mapping algorithms. Any other methods of measuring utility may be considered appropriate, but require further justification, and should be considered for sensitivity analyses, according to NICE.[21] Nafees et al. 2008[35] and Doyle et al. 2008[36] do not use EQ-5D, health status is not collected from patients, and the standard gamble method of utility measurement is used. To account for the effect of pulmonary embolisms on HRQoL, an additional study, Locadia et al. 2004[37] was identified. Similar to adverse event utility decrements identified in other studies, the methods in Locadia et al. 2004[37], do not follow NICE Methods Guidance preferences for utility measurement. Locadia et al. 2004[37] did not use EQ-5D, and did not value states using the general public, but did derive utility scores from patients using the time trade-off method. Utility decrements for adverse events reported in the CS are presented in Table 31.

The utility data collection schedule in the SQUIRE trial means that patients who received necitumumab have, on average, a greater number of data points due to more frequent observation. They also have a health state that corresponds to the period directly after discontinuation of platinum doublets (NPD-maintenance). Patient utilities were averaged across the length of time spent in the respective treatment states. This means that patients in the off-treatment health state in the GCis + N arm are further removed from discontinuing platinum doublet therapy and further along in their disease, on average, than patients in the GCis arm. The difference in measurement timing has the potential to bias the utility estimates in favour of GCis + N. In order to address this potential bias, the ERG requested utility scores reported by arm, for each time point that was recorded with adjustment for baseline imbalances. This data would allow the comparison of HRQoL in the two treatment arms by time, rather than by potentially biased health states. This data was not supplied as requested in the company's response to clarification questions.

Table 31 Utility decrements and duration of effect for AEs in company model

Adverse event	Mean	Lower CI	Upper CI	PSA distribution	Source
Neutropenia	0.09	0.05	0.13	Beta	Nafees et al. 2008
Anaemia	0.07	0.04	0.12	Beta	Nafees et al. 2008
Thrombocytopenia	0.09	0.05	0.13	Beta	Nafees et al. 2008
Hypomagnesaemia	0.09	0.06	0.12	Beta	Nafees et al. 2008
Pulmonary Embolism	0.32	0.12	0.57	Beta	Locadia et al. 2004
Asthenia	0.07	0.04	0.12	Beta	Assumption
Leukopenia	0.09	0.05	0.13	Beta	Assumption
Rash	0.03	0.02	0.05	Beta	Nafees et al. 2008
Fatigue	0.07	0.04	0.12	Beta	Nafees et al. 2008
Nausea1	0.05	0.02	0.09	Beta	Nafees et al. 2008
Vomiting1	0.05	0.02	0.09	Beta	Nafees et al. 2008
Hypokalaemia2	-	0.00	0.00	Beta	Assumption
Hyponatraemia2	-	0.00	0.00	Beta	Assumption
Febrile neutropenia	0.09	0.06	0.12	Beta	Nafees et al. 2008
Dyspnoea	0.05	0.02	0.10	Beta	Doyle 2008
Pneumonia	0.07	0.04	0.12	Beta	Assumption
Utility decrement duration for AEs grade 3/4 (days)					
Neutropenia	7.00				
Anaemia	7.00				
Thrombocytopenia	7.00				
Hypomagnesaemia	12.30				
Pulmonary Embolism	30.44				
Asthenia	7.00				
Leukopenia	7.00				
Rash	12.30				
Fatigue	32.00				
Nausea[a]	2.50				
Vomiting[a]	2.50				
Hypokalaemia[b]	7.00				
Hyponatraemia[b]	7.00				
Febrile neutropenia	7.00				
Dyspnoea	7.00				
Pneumonia	7.00				
[a]Nausea and vomiting combined into one state in the model [b]No adverse utility in the model, decrement listed in CS Table 60					

Note: Table derived from CS Table 73, pp. 208 to 209.

No imputation methods were reported for the analysis of SQUIRE EQ-5D data. Approximately [redacted] across both arms had baseline utility data and at least one completed post-baseline assessment in the ITT population as reported in section 4.9 of the CS. However, the data used in the model is not from the ITT population in the base case, it is from the Western European population, so the amount of missing data is unclear. The ERG requested data on utility scores collected at each follow-up with the number of patients missing. This data was not provided by the company, so the amount of data gathered and the amount of data missing over the length of the SQUIRE trial is unclear.

Using the Western European population for utilities was not justified in the CS. It is possible within the model to use utility values from the ITT population. Table 32 reports the SQUIRE trial ITT pre-progression health state utility values from the company's model. The utility value for the progressed state is derived from Khan et al. 2015[33] and remains unchanged from the Western European (base case) analysis. The utility scores for the ITT population are lower than those found in the Western European (base case) analysis in general, but have the greatest difference for maintenance therapy. Because of this, they adversely affect the cost-effectiveness of GCis + N, as the maintenance state is exclusive to GCis + N.

Table 32 Pre-progression utility scores from the SQUIRE EGFR expressing ITT population (company model)

Health State	Utility Score (mean)	Standard Error
On induction treatment	[redacted]	[redacted]
Off treatment	[redacted]	[redacted]
Receiving maintenance treatment	[redacted]	[redacted]

There are several limitations in the company's approach to measuring quality of life. They went through considerable effort to add utility decrements for adverse events to the economic model. However, it might be argued that this was unnecessary, as the patients in the SQUIRE trial will have included adverse events in their assessment of their own quality of life during the trial.

The assumption that the treatment arms produce the same utility for the progression free on induction and progressed states is unnecessary. It might have been more appropriate to adjust the arms for baseline imbalances and allow them to have different utilities over time. Another

method of handling utility scores would be to assign them to cycles in the model instead of to health states. This would allow more granular utility data and eliminate potential biases due to differential measurement timings.

Bottom-line summary of ERG view on patient outcomes

The data from the SQUIRE trial appears to have been collected in an appropriate manner, but has not been used to its full potential. There is little reporting on missing data and the base case analysis used utility scores from the Western European subgroup. AEs in the model were derived from sources that have poor compliance with NICE's preferred methods of HRQoL measurement methods[21] and using this data is a form of double counting.

4.3.7 Resource use and costs

4.3.7.1 Resource Use

The model included resource use for drug administration, disease monitoring and management, adverse events, and for active treatment and palliative care following disease progression. Resource use for drug administration, disease monitoring and supportive care was obtained from a retrospective medical chart review (Table 33) and validated by expert opinion. In addition to the retrospective chart review, a systematic review and consultations with experts were undertaken to identify further resource use. Clinical experts were consulted to confirm resource use assumptions.

Table 33 Retrospective medical chart review methodology

objective	**To assess treatment patterns among patients with a diagnosis of metastatic squamous NSCLC and who are receiving first-line treatment with a platinum-based doublet regimen in the UK.**
Study design	This study was carried out using a retrospective, non-interventional observational review of medical records for patients with a confirmed diagnosis of metastatic squamous NSCLC (i.e., Stage IIIB with pleural effusions according to the 6th edition of the AJCC guidelines or Stage IV according to the 6th or 7th edition of the AJCC guidelines; or initially diagnosed with a more limited stage and progressed to metastatic disease). In this study, physicians served as the direct data abstractors, allowing for efficient and accurate interpretation of their own notes and records. Participating physicians selected patients who met the screening criteria and abstracted the requested data elements which existed in the patient's chart at the time of abstraction. Physicians then entered the abstracted data into a web-based DCF, which was compiled into a patient-level analytic file. As patient chart data may contain highly sensitive and private personal health information, only anonymous data were collected for use in this study. The patient's physician was the only entity who had access to potentially sensitive patient data.
Sample size	Due to the retrospective, descriptive nature of this study, the study size was not based on formal statistical considerations. A sample of 54 physicians in the UK participated in the study with 203 patients in the UK.
Physician Selection	Physicians recruited to perform the patient-level medical record abstractions must have been located in the UK, with a case load of at least 6 patients with metastatic squamous NSCLC treated in the past 12 month. They must have also been in practice for 5 to 30 years after completion of formal training or board certification and a medical specialty of medical oncology, clinical oncology, haematology-oncology, pulmonology, or internal medicine specialized in pulmonology.
Patient Selection	The patients must have a confirmed diagnosis of metastatic squamous NSCLC, aged at least 18 years on the date of diagnosis of metastatic squamous NSCLC, initiated systemic treatment after diagnosis of metastatic disease with a platinum-based doublet regimen (i.e., cisplatin or carboplatin in combination with another agent) and stopped first-line systemic treatment and stopped maintenance therapy (if any maintenance therapy was received after first-line treatment). Maintenance therapy was defined as either the continuation of one first-line therapy agent or a switch to another single agent before any disease progression occurred.
Outcomes Measured	Overall treatment patterns, systematic therapy and supportive care

Note: This table is a direct reproduction of CS Table 65, p. 191.

The systematic review search strategy expanded search strategies conducted as part of the erlotinib STA (TA 258) and crizotinib STA (TA 296) related to advanced or metastatic lung cancer. Medline, Medline In Process and EMBASE were searched using the OVID platform. An

expanded strategy was adopted for the NHS EED and NICE technology appraisals. For the update of the resource use systematic literature review, the date limits were restricted to 2012 onwards to account for the time elapsed since the searches performed for erlotinib and crizotinib. The inclusion and exclusion criteria for the resource use systematic review are reported in Table 34.

Table 34 Resource use systematic review inclusion/exclusion criteria

Population	Adult patients with metastatic or advanced lung cancer
Interventions	Any
Comparators	Any
Outcomes	Resource use from a UK NHS perspective
Study Design	Any
Exclusion Criteria	Not in metastatic/advanced lung cancer Not UK specific Not regarding resource use Publications prior to 2012

Note: This table is a direct reproduction of CS Table 61, p. 245.

In addition to studies identified by the update search, the studies identified in the original TA 258 and TA296 searches were evaluated. In total 19 studies, including 10 NICE STAs and 9 publications, were identified. Among these studies, only Brown et al. 2013[19] was utilised. Brown et al. 2013[19] provided data on AE costs and palliative care costs.

The identification of resource use data appeared to be well conducted and comprehensive. The restrospective chart review had a thorough description and was conducted in a squamous NSCLC population, which is directly relevant to this STA. Brown et al. 2013[19] and other studies identified in the systematic review of resource use did not directly compare resource use in squamous populations. The justification for selecting Brown et al. 2013[19] for resource use was not provided, but it appears to be appropriate.

4.3.7.2 Unit costs

Unit costs were derived from the drugs and pharmaceutical electronic market information tool (eMit) (December 2014), the British National Formulary (BNF) 68 (2015), NHS Reference Costs (2013/14), and Personal and Social Services Research Unit (PSSRU) costs (2014).[38-41] EMit

costs were used for all pharmacological treatments with the exception of erlotinib and dietary supplements, which were obtained from the BNF. NHS Reference Costs were used for outpatient administration of chemotherapy, medical oncology outpatient appointments, palliative care specialists, biochemistry tests, CT scans, chest x-rays, radiation therapy, red blood cell transfusions, hospitalisations and Accident and Emergency outpatient visits. PSSRU costs were used for GP visits, clinical nurse specialists, GP home visits and community nurse visits. In general, the most recent unit cost data was used, but the use of more recent eMit data (which was published in November 2015[42]) would allow the analysis to be conducted in 2015 GBP (£) instead of 2014 GBP (£).

Drug costs

The acquisition costs for treatment are reported in Table 35 and Table 36. Drug costs were derived from the company for GCis + N, whilst eMit prices were used for comparators, and the BNF was used for erlotinib second-line therapy. The cost of administration was derived from NHS Reference Costs for delivering complex chemotherapy at first attendance, and for subsequent administrations. The cycle length and number of administrations for each treatment were obtained from the appropriate SmPC. After discontinuing first-line treatment, patients were either off treatment or receiving second-line therapy of docetaxel or erlotinib.

Proportions of patients receiving second-line therapy were based on the SQUIRE trial. Patients on GCis + N and GCis received different mixes of second-line therapies. The CS states that 47.3% of patients in the GCis + N arm and 44.7% of patients in the GCis arm received second-line therapy. For GCis + N, 30.6% of patients received second-line docetaxel and 10.5% received erlotinib. For GCis, 23.2% received second-line docetaxel and 13.7% received erlotinib. The company's estimated cost for each daily erlotinib administration is £67.78, £1,423.46 per 3 week cycle. The cost per docetaxel administration is £54.53 and is given every three weeks. The company indicated that, according to their market data, 70% of lung cancer patients receiving second-line therapy receive docetaxel, and 30% receive erlotinib.

Table 35 Drug costs

Regimen	Technology	Recommended Dose	Cycle Length	Unit Cost	Source
First-line treatment					
GCis + N	Necitumumab	800 mg on Days 1 and 8	3 weeks	800mg vial=£1,450 [redacted]	
	Gemcitabine	1250 mg/m^2 on Days 1 and 8	3 weeks	200mg/2ml=£5.11; 1g/10ml=£12.71; 2g/20ml=£29.03	eMit December 2014
	Cisplatin	75 mg/m^2 on day 1	3 weeks	100mg/100ml=£15.60; 10mg/10ml=£3.71; 50mg/50ml=£8.09	eMit December 2014
GCis	Gemcitabine	1250 mg/m^2 on Days 1 and 8	3 weeks	200mg/2ml=£5.11; 1g/10ml=£12.71; 2g/20ml=£29.03	eMit December 2014
	Cisplatin	75 mg/m^2 on day 1	3 weeks	100mg/100ml=£15.60; 10mg/10ml=£3.71; 50mg/50ml=£8.09	eMit December 2014
GCarbo	Gemcitabine	1250 mg/m^2 on Days 1 and 8	3 weeks	200mg/2ml=£5.11; 1g/10ml=£12.71; 2g/20ml=£29.03	eMit December 2014
	Carboplatin	400 mg/m^2 on day 1	3 weeks	450mg/45ml=£19.07; 150mg/15ml=£7.71; 50mg/5ml=£3.51	eMit December 2014
PCarbo	Paclitaxel	175 mg/m^2	3 weeks	150mg/25ml=£12.71; 30mg/5ml=£3.78	eMit December 2014
	Carboplatin	400 mg/m^2 on day 1	3 weeks	450mg/45ml=£19.07; 150mg/15ml=£7.71; 50mg/5ml=£3.51	eMit December 2014
DCis	Docetaxel	75 mg/m^2	3 weeks	80mg/ml=£25.73; 20mg/1ml=£7.45; 140mg/7ml=£54.60	eMit December 2014
	Cisplatin	75 mg/m^2 on day 1	3 weeks	100mg/100ml=£15.60; 10mg/10ml=£3.71; 50mg/50ml=£8.09	eMit December 2014
Second-line treatment costs					
D	Docetaxel monotherapy	75 mg/m^2	3 weeks	80mg/ml=£25.73; 20mg/1ml=£7.45; 140mg/7ml=£54.60	eMit December 2014
E	Erlotinib	150 mg	3	25mg=£378; 100mg=£1,324;	BNF 68

monotherapy	weeks	150mg=£1,632 *All strengths are provided as 30 tablets	March 2015

Note: This table was adapted from Table 66 and Table 67 in the CS.

Table 36 Cost of chemotherapy administration

Description	Unit Cost	Reference
Deliver Complex Chemotherapy, at First Attendance	£401	NHS Reference Cost 2013-2014. Chemotherapy administration. Day Case. Currency Code SB14Z.
Deliver subsequent elements of a chemotherapy cycle	£328	NHS Reference Cost 2013-2014. Chemotherapy administration. Day Case. Currency Code SB15Z.

Note: This is a direct reproduction of CS Table 68, p. 195.

It would have been more appropriate to use this marketing data in the model. As currently modelled, the make-up of second-line therapy favours GCis + N, and it is unclear if this would be replicated in UK general practice. The ERG believes that the quantity of patients receiving second-line treatment and the duration of that treatment are appropriately derived from SQUIRE, but that the make-up of therapies received by patients on second-line therapy should be equivalent between arms. However, it should be noted that whilst the methods used in the company model are appropriate, NICE guidance on second-line erlotinib treatment has changed. Erlotinib is now only recommended for second-line treatment in patients with EGFR-TK mutation. Additionally, expert advice indicated that nivolumab is also used in second-line therapy.

Treatment intensity was adjusted using SQUIRE data for GCis + N and GCis and assuming comparators were equivalent to GCis. Drug wastage was included by default.

Health state costs

Beyond chemotherapy, patients in the model receive supportive and palliative care. Table 37 reports resource use and costs for patients on active therapy, whilst Table 38 reports resource use and costs for patients receiving supportive care.

The retrospective medical chart review determined that patients receiving active therapy require the following resource use: medical oncologist outpatient visits, GP visits, clinical nurse specialists, biochemistry tests, full blood count tests, CT scans, Chest X-rays, red blood cell

transfusions, opiate analgesics, antiemetics, accident and emergency (A&E) visits and oral dietary supplements. This resource use was determined to be appropriate for all patients receiving active treatment, even if they had previously progressed (includes patients on second-line therapy).

For patients who have progressed and are only receiving supportive care, the retrospective medical chart review determined the following resource use: medical oncologist outpatient appointments, GP home visits, district nurse visits, clinical nurse specialist home visits, chest x-rays opiate analgesics, antibiotics and A&E visits.

For patients who are within the last two weeks of life and receiving palliative care, the model assumes that 55.8% of patients receive palliative care in hospital, 16.9% will receive palliative care in hospice and 27.3% will receive palliative care at home with the aid of a Macmillan nurse, community nurse and GP home visits based on Brown et al. 2013.[19] Table 39 reports resource use for patients on palliative care at the end of life.

Table 37 Health state unit costs for patients receiving active therapy

Resource Required	Frequency	Unit Cost	Reference
Outpatient Visit with medical oncologist	100%; Once every 3 weeks	£147	Resource use: Clinical Expert Opinion Cost: NHS Reference Cost 2013/2014. Weighted average of Consultant Led Outpatient appointments- Non-admitted face to face, first time appointments with service code 370 (medical oncology) and Non-admitted face to face, follow-up appointments with service code 370 (medical oncology).
GP Visit	100%; Once monthly	£35	Resource use: Brown (2013), Appendix 1 NICE CG81 Cost: PSSRU 2014. 10.8b General practitioner — unit costs. Excluding qualification costs and direct care staff costs. Surgery consultation lasting 11.7 minutes.
Clinical Nurse Specialist	100%; Once every 3 weeks	£22	Resource use: Clinical Expert Opinion Cost: PSSRU 2014. 10.7 Nurse advanced (includes lead specialist, clinical nurse specialist, and senior specialist). £22 per surgery consultation (excluding qualification cost).
Complete	100%;	£3	Resource use:

Resource Required	Frequency	Unit Cost	Reference
Blood Count	Once per week		Clinical Expert Opinion Cost: NHS Reference Cost 2013/2014. Haematology. Currency Code DAPS05.
Biochemistry (Renal and Liver Function)	100%; Once per week	£2	Resource use: Clinical Expert Opinion Cost: NHS Reference Cost 2013/2014. Clinical Biochemistry. Currency Code DAPS04.
CT-Scan(Chest)	100%; once every 6 weeks	£110	Resource use: Retrospective Medical Chart Review Cost: NHS Reference Cost 2013/2014. Weighted average of computerised tomography scan with outpatient service description. Currency Code (RA08A, RA09A, RA10Z-RA14Z, RA50Z).
Chest X-ray	100%; once every 3 weeks	£30	Resource use: Retrospective Medical Chart Review Unit Cost: NHS Reference Cost 2013/2014. Directly Accessed Diagnostic Services. Direct Access Plain Film (Currency Code DAPF).
Opiate analgesics (30mg of codeine 4 times daily)	30%; daily	£0.11 per day	Resource use: Retrospective Medical Chart Review Unit Cost: EMIT December 2014. Codeine 30mg tablets/Pack size 100=£2.86
Antiemetic's (16mg ondansetron daily)	100%; 3 day of every cycle	£0.36 per day	Resource use: Clinical Expert Opinion Cost: EMIT December 2014. Ondansetron 8mg tablets / Pack size 10=£1.82
Red blood cell transfusion	21%; two units every 3 months	£195 per transfusion	Resource use: Retrospective Medical Chart Review Cost: NHS Reference Cost 2013/2014. Single Plasma Exchange, Leucophoresis or Red Cell Exchange, 19 years and over. Procedures in Outpatients (Currency Code SA13A)
Accident & Emergency visit	11%; once every 12 weeks	£88	Resource use: Retrospective Medical Chart Review Cost: Non-Consultant led Outpatient Attendances. Non-admitted Face to Face Attendance, First. Currency Code (WF01B). Accident & Emergency.

Resource Required	Frequency	Unit Cost	Reference
Antibiotics	25%; 7 days in every cycle	£1.71 per day	Resource use: Retrospective Medical Chart Review Cost: 500 mg of levofloxacin once daily for 7 days. Levofloxacin 500mg/pack size 10=£17.19
Oral dietary supplement (200 ml Ensure daily)	11%; daily while on	£2.02 per day	Resource use: Retrospective Medical Chart Review Cost: Ensure Plus. Liquid. Bottle, 200ml=£2.02

Note: Adapted from CS Table 69, pp. 197 to 199.

Table 38 Resource use and costs for patients receiving supportive care

Resource Required	Frequency	Unit Cost	Reference
Outpatient Visit with medical oncologist	100%; Once every 3 weeks	£147	Resource use: Clinical Expert Opinion Cost: NHS Reference Cost 2013/2014. Weighted average of Consultant Led Outpatient appointments- Non-admitted face to face, first time appointments with service code 370 (medical oncology) and Non-admitted face to face, follow-up appointments with service code 370 (medical oncology).
District Nurse	100%; Twice monthly	£19	Resource Use: Brown (2013) Cost: PSSRU 2014. 10.1 Community nurse (includes district nursing sister, district nurse.) £57 per hour of patient related work. Excluding qualifications costs. (Assuming each visit has a 20 minute duration according to Brown et al. 2013)
GP home Visit	100%; Twice monthly	£35.00	Resource use: Brown (2013), Appendix 1 NICE CG81 Cost: PSSRU 2014. 10.8b General practitioner — unit costs. Excluding qualification costs and direct care staff costs. Surgery consultation lasting 11.7 minutes.

Note: Adapted from CS Table 70.

Table 39 Unit costs for end-of-life palliative care

Description	Resource Use	Unit Cost	Source
Palliative Care- Hospital	55.8%	£4,153	Resource use: Brown (2013) Cost: NHS Reference Cost 2013/2014: Non-elective inpatient (long stay). Respiratory Neoplasms with CC Score 11+. Currency Code DZ17E.
Palliative Care- Hospice	16.9%	£5,191	Resource use: Brown (2013) Cost: Brown (2013) assumed hospice was a 25% increase in hospital inpatient cost. NHS Reference Cost 2013/2014: Non-elective inpatient (long stay). Respiratory Neoplasms with CC Score 11+. Currency Code DZ17E.
Palliative Care- Home	27.3%	Community Nurse Visit: £266	Resource use: Brown (2013) Cost:

Description	Resource Use	Unit Cost	Source
		GP Home visit: £70 Macmillan Nurse: £1901	Community Nurse Visit- PSSRU 2014. 10.1 Community nurse (includes district nursing sister, district nurse.) £57 per hour of patient related work. Excluding qualifications costs. (Assuming each visit has a 20 minute duration and occurs for 14 days during terminal care according to Brown et al. 2013) GP Home Visit- PSSRU 2014. 10.8b General practitioner — unit costs. Excluding qualification costs and direct care staff costs. Home visit lasting 11.4 minutes. (Assuming occurs weekly- twice in 14 days) Macmillan Nurse-PSSRU 2014. Brown (2013) assumed a Macmillan nurse was 66.7% of the cost of a community nurse (£57 per hour). Also assumed would be required for 50 hours for terminal care.

Note: This is a direct reproduction of CS Table 71, p. 202

Adverse Events

AE costs were also included for grade three and four AEs related to treatment. The resource use associated with AEs were derived from Brown et al. 2013,[19] and the NICE DSU document on febrile neutropenia. Costs were obtained from NHS Reference Costs, eMit and the NICE DSU document on febrile neutropenia. The company indicated that resource use estimates were validated with clinical opinion. Table 40 reports the resource use and costs used for adverse events in the model.

Bottom-line summary of ERG view on resource use and costs

The company did a comprehensive search for appropriate costs, updating systematic reviews in previous NSCLC STAs and performing a comprehensive and well-reported chart review of relevant squamous patients. The cost categories included were reasonable and well-reported. EMit cost data used for the price of comparator drugs was not up to date, and the variance data reported with eMit data was not used to inform the PSA.

Overall, the company's analysis of resource use and costs was appropriate and comprehensive.

Table 40 Calculation of resource use and cost of adverse events

Description	Unit Cost	Reference
Neutropenia	£349.34	NHS Reference Cost 2013/2014. Weighted average of mean costs for HRG code WA02W (disorders of immunity without HIV/AIDS with complicating condition) across non-elective long- and short-stay episodes and day-case admissions.
Anaemia	£755.53	NHS Reference Cost 2013/2014. Weighted average of Iron Deficiency Anaemia with CC Score 0-14+ (Currency Code SA04G, H, J, K, L) non-elective inpatient long stay, non-elective inpatient short stay and day case.
Thrombocytopenia	£195	NHS Reference Cost 2013/2014. Single Plasma Exchange, Leucophoresis or Red Cell Exchange, 19 years and over. Procedures in Outpatients (Currency Code SA13A)
Hypomagnesaemia	£590.00	EMIT December 2014. Magnesium sulphate 10% solution for infusion 20mmol IV over a 6 hour period for a maximum of 5 days. (£9.68/day, totalling £48.42). NHS Reference Cost 2013/2014. Assumed the infusion given as a non-elective inpatient short stay as a Neoplasm Related Admission with CC Score 0-3+ (Currency Code WA17A-WA17D) (£541.80 (£319.70, £612.30)
Pulmonary embolism	£654.84	NHS Reference Cost 2013/2014: Weighted average of Deep Vein Thrombosis with CC Score 0-12+ (Currency Code YQ51A-YQ51E) inpatient long stay, short stay and day case.
Skin rash	£147.39	NHS Reference Cost 2013/2014. Weighted average of Consultant Led Outpatient appointments- Non-admitted face to face, first time appointments with service code 370 (medical oncology) and Non-admitted face to face, follow-up appointments with service code 370 (medical oncology).
Fatigue	£3008.41	NHS Reference Cost 2013/2014. Weighted average of the non-elective long-stay. Neoplasm Related Admission with CC Score 0 to 3+). Currency code WA17A-WA17D.
Nausea	£1494.00	NHS Reference Cost 2013/2014. Minor Therapeutic or Diagnostic, General Abdominal Procedures, 19 years and over as a non-elective short-stay episode. Currency Code FZ13C. Each hospitalisation cost £747 (£459, £833) with two hospital admissions typically required during chemotherapy.
Febrile Neutropenia	£4,402.87	The NICE Decision Support Unit report (2007) on the cost of febrile neutropenia as an inpatient estimated the cost to be £2572.44 (89). The cost from 2007 has been inflated to 2015 using the CPI from the ONS to £3144.19 Brown (2013) assume 1.4 episodes per patient during the four cycles (12 weeks) of chemotherapy.

Note: This table is a direct reproduction of CS Table 72, p. 204.

4.3.8 Cost-effectiveness results

Deterministic results from the company's base case economic model are presented in section *5.7* (pp. 214 to 220) of the CS. These relate to the Western European subgroup, including patients without EGFR expressing tumours. The company's clarification response included revised tables for patients with EGFR expressing tumours within the Western European subgroup (CS clarification response Appendix 1). In this section we report the revised results from the clarification response. Results from the company's sensitivity analyses are discussed below in section 4.3.9. This includes probabilistic results, one-way sensitivity analyses, and scenario analyses (which include estimates for the ITT population).

The company base case estimates for GCis + N compared with GCis in the Western European population with EGFR expressing tumours are shown below (Table 41). These are based on a deterministic analysis for the direct comparison only, using data from the SQUIRE trial (with KM survival functions for OS and PFS, extrapolated from the KM endpoint with log-logistic survival functions, fitted separately to each study arm). The company noted the modelled benefits of GCis + N: a mean gain in OS of 0.54 years (6.5 months), corresponding to a mean gain of 0.34 QALYs per patient. Given the higher estimated cost (a mean increase of £19,516 per patient), the ICER was £57,725 per QALY gained.

Table 41 Base case cost-effectiveness results – Direct comparison (deterministic) (Western European EGFR expressing population)

Technologies	Total costs	Total LYG	Total QALYs	Incremental costs	Incremental LYG	Incremental QALYs	ICER
GCis	[redacted]	[redacted]	[redacted]				
GCis + N	[redacted]	[redacted]	[redacted]	£19,516	0.544	0.338	£57,725

Note: this table is directly reproduced from Table 22 in the clarification response Appendix 1, p. 36.

Results presented by the company for the other comparisons are shown in Table 42. The deterministic results, with survival functions for OS and PFS estimated from SQUIRE data GCis + N and GCis (KM survival functions extrapolated from the endpoint using separatel Weibull models); and hazard ratios from the NMA for other comparators versus GCis + N. results for GCis + N are slightly different to those reported for the above direct comparisor (Table 41) because of the use of different parametric survival curves for extrapolation – Weibull instead of log-logistic. Note that the ICERs in this table are for GCis + N compared separately

with each comparator – it is not a fully incremental analysis – and the company did not report results for GCis in this table.

Table 42 Base case cost-effectiveness results – Indirect comparisons (deterministic) (Western European EGFR expressing population)

Technologies	Total costs	Total LYG	Total QALYs	Incremental costs	Incremental LYG	Incremental QALYs	ICER
GCis + N	[redacted]	[redacted]	[redacted]				
GCarbo	[redacted]	[redacted]	[redacted]	£20,316	0.523	0.344	£59,031
DCis	[redacted]	[redacted]	[redacted]	£19,948	0.482	0.312	£63,982
PCarbo	[redacted]	[redacted]	[redacted]	£20,036	0.236	0.172	£116,344

Note: this table is directly reproduced from Table 23 in the clarification response Appendix, p.37. LYG, life years gained.

To aid interpretation, we have extracted the results for GCis from the submitted company model and conducted an incremental analysis. Table 43 suggests PCarbo would have a lower mean cost and greater mean QALY than DCis or GCis. When an intervention is less effective and more expensive as DCis and GCis are in the base case incremental analysis, they are referred to as dominated. Comparing GCis + N with the next best, non-dominated option (PCarbo), the estimated ICER is over £116,000. However, we note that the absolute differences in costs and QALYs between the four included platinum doublets are all small and that there is a question over the robustness of the NMA. Based on the direct evidence from the SQUIRE trial with Weibull endpoint extrapolations, the results from this analysis suggest an ICER of £80,912 per QALY for GCis + N compared with GCis.

Table 43 Incremental analysis – direct and indirect comparisons (deterministic) (Western European EGFR expressing population)

Technologies	Total costs	Total LYG	Total QALYs	Incremental costs	Incremental LYG	Incremental QALYs	ICER
GCarbo	[redacted]	[redacted]	[redacted]				
PCarbo	[redacted]	[redacted]	[redacted]	£280	0.287	0.172	£1,628
DCis	[redacted]	[redacted]	[redacted]	Dominated			
GCis	[redacted]	[redacted]	[redacted]	Dominated			
GCis + N	[redacted]	[redacted]	[redacted]	£20,036	0.237	0.172	£116,488

LYG, life years gained.

4.3.9 Assessment of uncertainty

The company used probabilistic and one-way deterministic analyses to explore the impact of uncertainties around input parameters, and scenario analyses to examine structural uncertainties.

4.3.9.1 Probabilistic sensitivity analysis

The company reported the results of their probabilistic base case analysis in section 5.8 (pp. 220 to 223) of the CS. They noted that these results were based on 2,000 PSA iterations, and that they had checked the stability of the cost-effectiveness results by running up to 10,000 iterations. The PSA included the following sets of parameters that were subject to uncertainty:

Nonparametric survival curves (KM)

A random process was used in the PSA to introduce variation to the KM curves to reflect sampling uncertainty over the SQUIRE data. The process used a single random number per curve, per PSA iteration, to shift the curves vertically in proportion to the reported standard errors for the KM estimates at each time point. The PSA value for the proportion of the cohort surviving in the first week $S^*(t_1)$ was drawn randomly from a lognormal distribution based on the KM mean μ_1 and point standard error σ_1 at time t_1: $S^*(t_1) \sim LN(\mu_1, \sigma_1)$. Survival at the next time point t_2 was then calculated by adjusting the KM estimate : $S^*(t_2) = [S^*(t_1) - \mu_1]^* (\sigma_2 / \sigma_1)$. And so on for successive time points.

Parametric survival curves (log-logistic, Weibull, etc)

Random sampling was also used to draw PSA values for the parametric survival models which were used to extrapolate OS and PFS beyond SQUIRE follow up. The related sets of parameters, such as the scale and the shape parameters for the Weibull distribution, were drawn from mulitvariate normal distributions with mean and variance-covariance matrices from the regressions used to fit the models to the SQUIRE data. This process introduces variation in the position and shape of the sampled curves between PSA iterations.

Hazard ratios for indirect comparators

The effectiveness of the indirect comparators entered the model in the form of hazard ratios for OS and PFS compared with GCis + N. Values for these hazard ratios were sampled from independent lognormal distributions, based on means and 95% confidence intervals estimated

from the NMA. This approach does not account for correlations between estimates for the different comparators within the NMA.

Incidence of adverse events

The proportions of patients experiencing the included AEs for the GCis + N and GCis arms were sampled from beta distributions, according to the observed proportions reporting these events in the SQUIRE trial and the numbers of patients within the relevant treatment arm.

Relative risks of adverse events

For the indirect comparators, it was assumed that the relative risks of TEAEs compared with GCis + N would equal those for GCis. For the PSA, these relative risks were sampled from lognormal distributions, based on mean and 95% confidence intervals estimated from SQUIRE. Values for each comparator (GCarb, GCarbo and DCis) and for each included TEAE were sampled independently.

Health state utilities and adverse event utility decrements

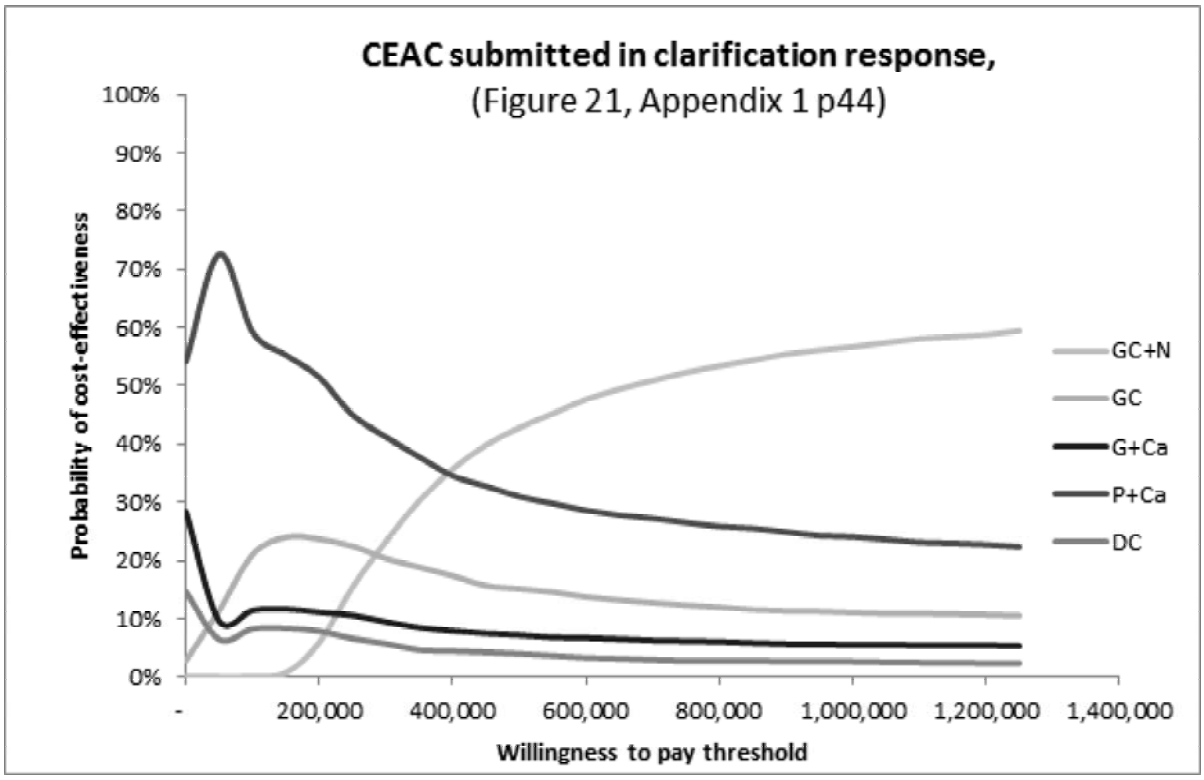
PSA values for utilities and utility decrements were sampled from beta distributions. Means and standard errors for the health state utilities prior to progression were taken from the SQUIRE trial, and those for the post-progression state from literature. Means and standard errors for disutilities associated with the included adverse events were taken from the literature.

Resource use and costs

Unit costs of drugs (£ per mg or ml) were not varied in the PSA. However, the cost of drugs per administration was variable, due to variation in the treatment duration adjustment and wastage rates (for drugs other than necitumumab): the treatment duration adjustment was sampled from a beta distribution (see section 4.3.5 above); and wastage depended on body surface area, which was sampled from a normal distribution. Uncertainty over the cost of drugs after progression was included by sampling the proportions of patients assumed to receive erlotinib and docetaxel (from a beta distribution) and the duration of that treatment (from a gamma distribution). Other costs, including the costs of drug administration, other resource use, and one-off costs for palliative care and death were sampled directly from gamma distributions, with subjective estimates of variation (standard errors set at 30% of the mean).

The methods used in the company model PSA were generally of a good standard. Uncertainty over the key effectiveness parameters of overall and progression-free survival was captured appropriately, based on variance in the SQUIRE trial. Uncertainty related to the outputs from the NMA (hazard ratios for OS and PFS) was included, but did not account for correlations of estimates between comparators. Conventional methods were used to incorporate uncertainty around other input parameters, with variance based on empirical sources or reasonable subjective estimates.

The CS summarises the results of the PSA in the form of Cost-effectiveness Acceptability Curves (CEAC) and cost-effectiveness scatterplots. No numerical results are provided in the CS. A revised version of the CEAC based on the Western European subgroup with EGFR expressing tumours was provided in the company's clarification response (CS Appendix 1, Figure 21 p. 44), this is reproduced below (Figure 11, top). The ERG notes that the company's submitted CEAC suggests that GCis + N does not have the greatest probability of being the most cost-effective option unless the willingness to pay threshold is above about £400,00 QALY. This appears inconsistent with the deterministic results summarised above (e.g. s Table 43). It also appears inconsistent with the equivalent PSA results calculated from th submitted model by the ERG (Figure 11, bottom).



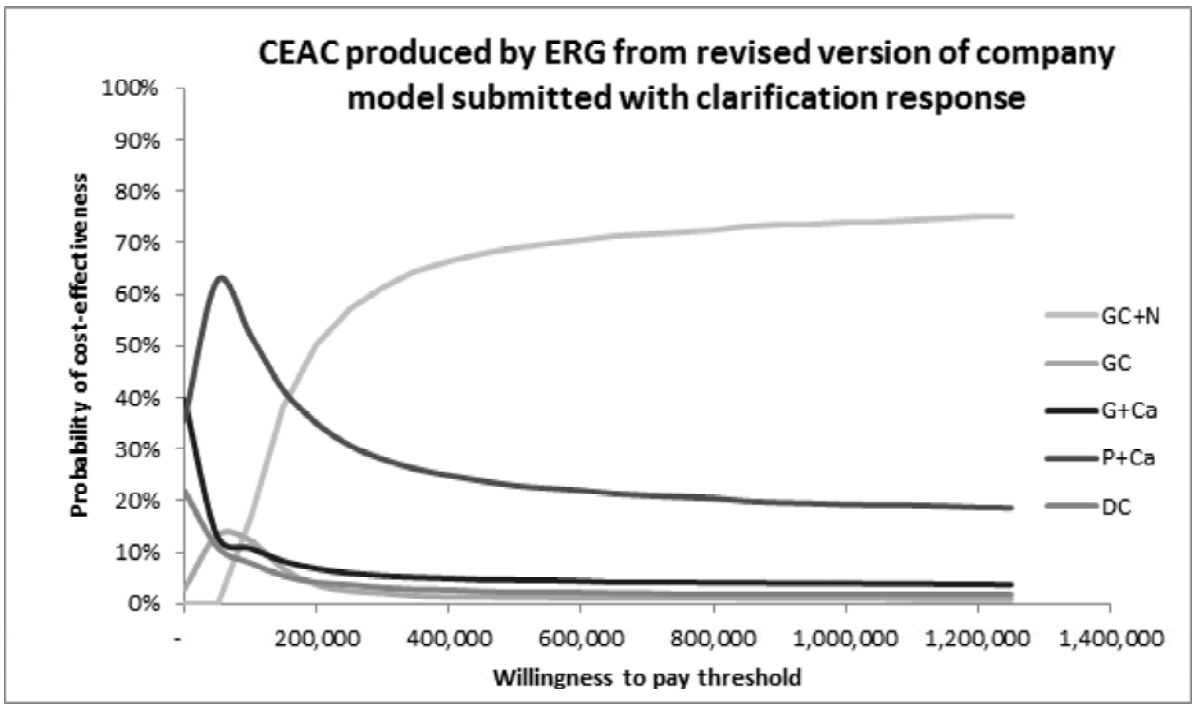


Figure 11 Cost-Effectiveness Acceptability Curve in patients with EGFR expressing tumours (Western European subgroup)

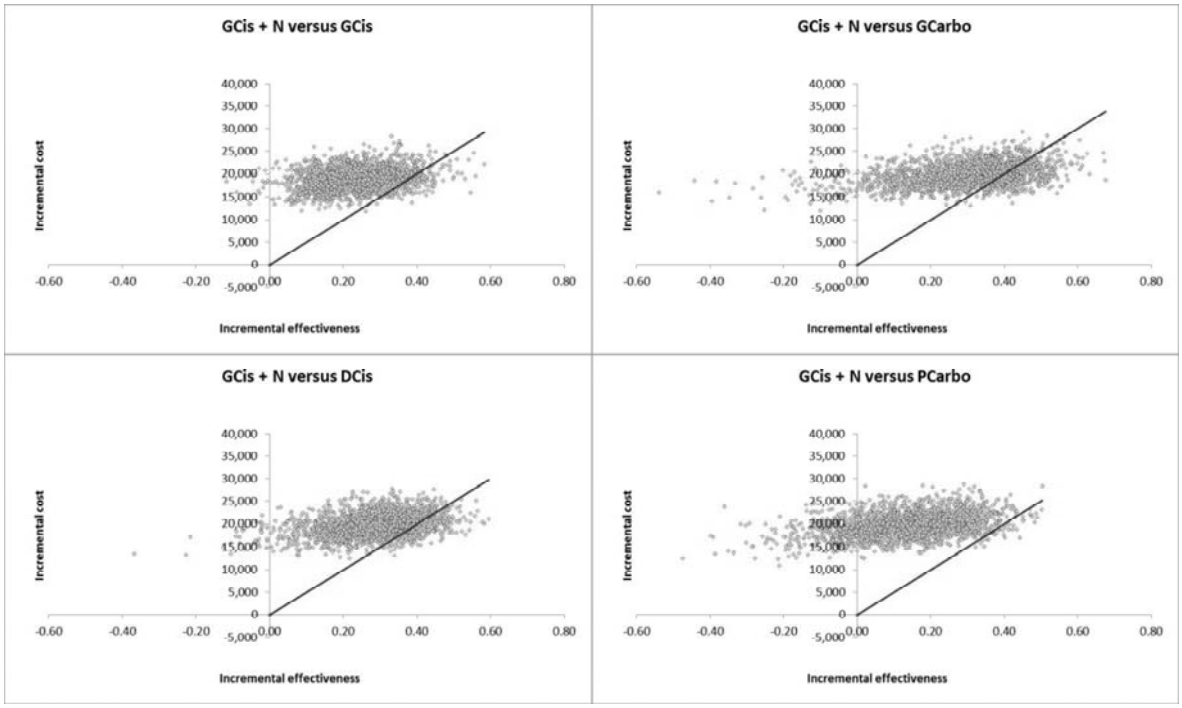
Numerical results from the revised model submitted as part of the clarification response are presented below in Table 44. These relate to the Western European subgroup of patients with EGFR expressing tumours, and are based on KM estimates of OS and PFS extrapolated from the endpoint with separately fitted Weibull curves for GCis and GCis + N, and adjusted by NMA hazard ratios for other comparators. This estimated ICER for GCis + N versus the next best non-dominated alternative (PCarbo) is £144,737 per QALY. If we restrict the analysis to the direct comparison between GCis + N and GCis, the ICER is £80,634.

Table 44 Incremental analysis – direct and indirect comparisons (PSA results) (Wes
European EGFR expressing subgroup)

Technologies	Total costs	Total LYG	Total QALYs	Incremental costs	Incremental LYG	Incremental QALYs	ICE
GCarbo	[redacted]	[redacted]	[redacted]				
PCarbo	[redacted]	[redacted]	[redacted]	£134	0.265	0.160	
DCis	[redacted]	[redacted]	[redacted]	Dominated			
GCis	[redacted]	[redacted]	[redacted]	Dominated			
GCis + N	[redacted]	[redacted]	[redacted]	£19,868	0.181	0.137	£1

LYG, life years gained.

Cost-effectiveness scatterplots for the included comparisons in this analysis are shown be (Figure 12). The spread of dots illustrates the extent of uncertainty over the relative costs *effects of GCis + N in relation to the included comparators*. It can be seen that uncertainty the incremental effects (QALYs gained) is lower for the direct comparison with GCis than other, indirect comparisons estimated from the NMA. The slope of the red lines in these g shows a cost-effectiveness threshold of £50,000 per QALY gained. It can be seen that the probability that GCis + C is cost-effective (the proportion of dots below and to the right of t diagonal line) is low for all comparators.



The slope of the red lines indicates a cost-effectiveness threshold of £50,000 per QALY

Figure 12 Cost-effectiveness scatterplots– direct and indirect comparisons (PSA results) (Western European EGFR expressing subgroup)

4.3.9.2 Deterministic sensitivity analyses

The company conducted deterministic sensitivity analyses to test the impact of uncertainty around key model parameters on the cost-effectiveness results. This analysis was only conducted on the deterministic version of the model for the direct comparison (GCis + N versus GCis), and for the Western European subgroup. Parameters were varied one at a time, between defined lower and upper limits, and the resulting ICERs were recorded. The results are reported on pp. 223 to 226 of the CS, and revised results for Western European patients with EGFR expressing tumours were provided in the clarification response [clarification Appendix 1: Table 37 (p. 45) and Figure 22 (p. 47)]. The parameters included in this analysis were:

- **Survival curves** - OS, PFS and TTD estimates for GCis + N and GCis were varied (separately) at all time points between 95% confidence limits.
- **Adverse event risks, costs and utility decrements** – varied by 30% around the base case
- **Utilities for health states** – varied between 95% confidence limits.

- **Drug costs** – varied by 20% for necitumumab and 30% for other drugs.
- **Other costs** – varied by 50% around the base case.

The variables included and ranges of variation are reasonable.

Results are presented in the tornado diagram below (Figure 13). The company noted that the key drivers (among variables tested) for cost-effectiveness were: OS and PFS estimates for GCis +N and GCis, TTD for GCis + N, and the acquisition cost for necitumumab.

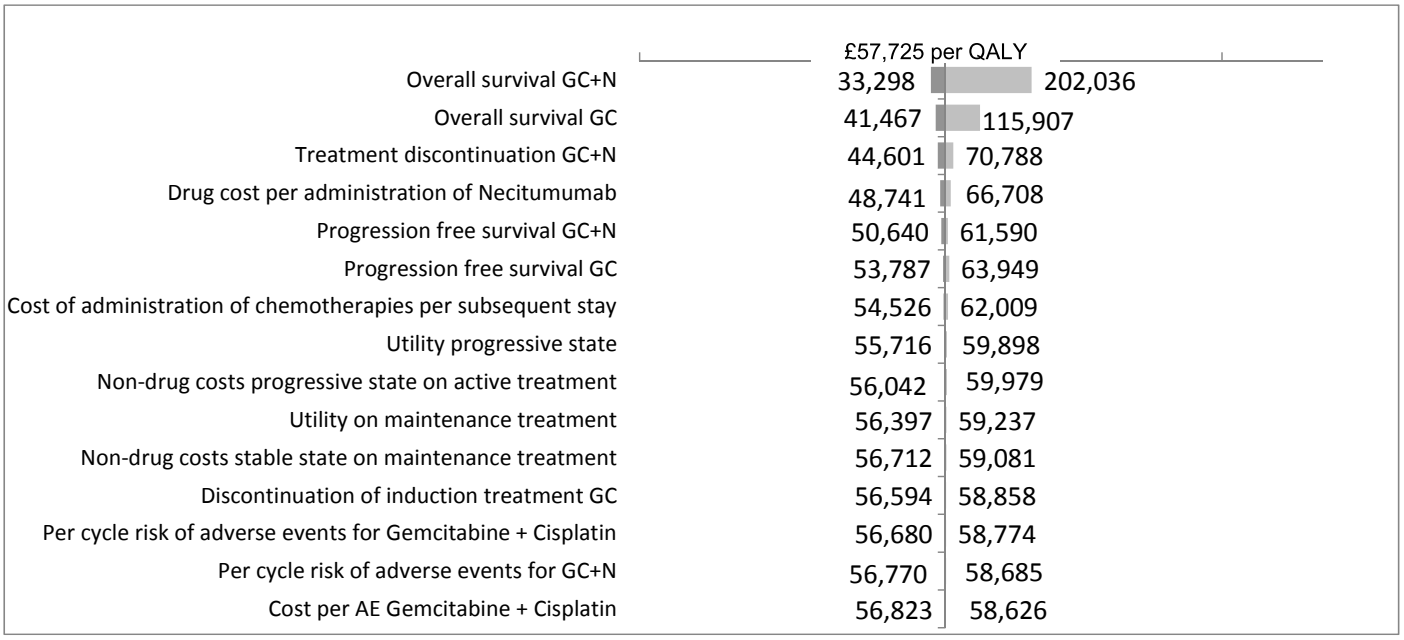


Figure 13 Tornado diagram for ICER (£ per QALY) – Western European patients with EGFR expressing tumours

Note: Reproduced for revised version of model submitted with clarification responses.

4.3.9.3 Scenario Analysis

The company conducted scenario analyses around key issues of structural uncertainty, see CS pp. 226 to 228. The analyses were deterministic, and only performed for the direct comparison (GCis + N versus GCis). The issues considered were:

- **Population** – the company's base case analysis relates to Western European patients with EGFR expressing tumours. A scenario analysis was conducted including results for all patients randomised in the SQUIRE trial with EGFR expressing tumours from the ITT population.
- **Functional forms for OS** – the company's base case analysis used KM estimates for OS, with separately fitted log-logistic curves used to extrapolate beyond the last OS observation for the GCis + N and GCis groups. They tested three different methods for extrapolation beyond the last OS observation: separately fitted Weibull functions for both groups; log-logistic for GCis + N and Weibull for GCis; and separately fitted exponential distributions.
- **Definition of PFS** – in the company's base, PFS was defined as radiographic documentation of progression (RECIST 1.0 criteria) or death. A scenario was conducted including symptomatic deterioration as well as radiographic progression and death.
- **Time horizon** – the base case assumed a time horizon of 18.5 years (effectively lifetime for this population. A shorter time horizon (5 years) was tested in order to assess the impact of the long tail of the log-logistic distribution, used to extrapolate OS and PFS in the base case model.
- **Time to treatment discontinuation** – due to the lack of reporting on discontinuation rates for studies in the NMA, assumptions were required to model TTD for comparators other than GCis. In the base case, it was assumed that hazard ratios for treatment discontinuation were equal to hazard ratios for PFS for each of the indirect comparators. A scenario analysis was conducted in which hazard ratios for treatment discontinuation for all comparators were assumed to be equal to those for GCis.
- **Source of utility estimates** – in the base case, utilities for the three pre-progression substates were estimated from the SQUIRE EQ-5D data (both arms pooled), post-

progression utility was taken from Khan et al 2015[33], and adverse event disutilites were taken from a variety of sources.[35-37] Two other senarios were tested: 1) state utilities from Chouaid et al. 2013[31] and AE decrements from Nafees et al. 2008; and 2) post-progression utility from Chouaid et al. 2013.[31]

The results of the company's scenario analyses for the Western European subgroup with EGFR expressing tumours were presented in the clarification response (Appendix 1, Table 38, p.48) and are reproduced below. The company noted that the model results were most affected by uncertainty in the methods used to extrapolate the OS and PFS survival curves. This is demonstrated by the scenarios testing alternative functional forms for OS (scenarios 5, 6 and 7), as well as the use of a truncated, 5-year time horizon (scenario 8).

The company did not comment on the impact of using the ITT patient population (scenario 4), but argued that the Western European subgroup is most applicable for the NHS.

Table 45, below, is derived from the company clarification response, and contains errors. For analyses 4-9 the company model assumes that time to treatment discontinuation is the same as GCis instead of the assumption from the base case, that time to treatment discontinuatior indirect comparators is equivalent to PFS.

Table 45 Scenario analysis results (GCis + N vs. GCis) in patients with EGFR expressing tumours (Western European subgroup)

Scenario	Description	Company ICER	Corrected ICER
Base-case	Base Case results	£57,725	£57,725
1)	Utilities from Chouaid et al. and AE decrements from Nafees et al.	£57,788	£57,788
2)	Utility post-progression from Chouaid et al.	£55,751	£55,751
3)	Time to treatment discontinuation assumed same as GCis for all indirect comparators	£64,713	£64,713
4)	Using ITT as patient population	£151,152	£110,248
5)	Using separate Weibull for OS	£87,543	£79,412
6)	Using Log-logistic for OS in GCis + N and Weibull in GCis arm	£53,433	£49,802
7)	Using separate Exponential distributions for OS	£78,868	£73,194
8)	5 year time horizon	£83,205	£76,744
9)	Symptomatic deterioration considered progression	£64,251	£57,354

Note: Adapted from clarification response, Appendix 1, Table 38, p.48.

The above scenario analyses are quite a narrow selection of the potential scenario analyses, and quite limited. Some important structural assumptions were not investigated in the CS. In particular, other methods for modelling the overall and PFS curves are appropriate for testing: other functional forms; use of purely parametric survival functions (rather than relying on the KM curves for SQUIRE follow up); use of jointly-fitted curves for GCis+ N and GCis (with a treatment interaction term); and methods for attaching the parametric extrapolation to the last KM estimates. It is also important to analyse the impact of influential scenarios and parameters under PSA (not just with the deterministic version of the model). These options are considered further in the ERG's additional analyses (section 4.4).

4.3.10 Model validation

The company noted (CS p.229-230) that as recommended by international methodological guidelines,[43] they used the following methods to test the validity of their model.

4.3.10.1 Face validity

The company noted that external clinical and economic advisors in the UK were consulted to validate that the model structure and modelling assumptions reflect the clinical pathway of patients with locally advanced or metastatic squamous NSCLC in England. The advisory panel comprised four NHS consultant oncologists, three UK academic health economists and two UK academic statisticians (CS p.165). The discussion guide for consultations is provided in the CS (Appendix 17, p.193). The company state that all of the recommendations from their experts have been addressed, although no further details are given.

The ERG also discussed some key assumptions in the model with a clinical expert, who agreed that they were generally plausible for a UK population and context.

4.3.10.2 Verification

Methods used to verify the model are reported on pp. 229 to 230 of the CS. The structure and programming of the model was checked by two modelling experts who were not involved in developing the model, but worked in the organisation commissioned by the company to develop the model. It is reported that they conducted stress tests – to check that the model behaved as expected when inputs were changed – and that debugging was conducted where necessary.

The ERG has also conducted a range of independent verification checks on the model:

- Reviewing the model structure and formulae;
- Checking that the model assumptions and inputs are consistent with those reported in the CS, and with the cited data sources (where available);
- Checking that the results and sensitivity analyses reported in the CS and clarification report are consistent with model outputs;

- Replicating the model in a separate Excel file to check that the calculations and macros yielded the same intermediate and final results;
- And investigating the impact of our own sensitivity and scenario analyses, and checking that the changes in results are consistent with expectations (see section 4.4 below).

We found a small number of minor errors and inconsistencies (Table 46), although none of these led to big changes in the model results.

Table 46 Inconsistencies identified by ERG in company model

Issue	Location in model	Effect on results	Action taken by ERG
1) Sum of incident cases of progression is greater than population size.	'Comp1Model' and 'Comp2Model' sheets, columns AG	Small increase in costs (about £200) for treatment after progression, for all comparators.	Corrected in ERG version of model used for additional analysis, although this formula might have related to the company's assumption that all patients entered the PD state before death.
2) Wrong denominator used for duration of second-line erlotinib use (whole SQUIRE population, rather than selected subgroup).	'Wastage' sheet cells D156 and J156.	Very small increase in costs (by about £20) for all comparators.	Corrected.
3) Denominator for calculation of AE risks differs between GCis and other chemotherapies, despite assumption of same relative risk.	'Per Cycle calculations' cells E45-G45.	Very small change in number of estimated AEs.	Corrected.
4) Cost for administration of chemotherapy in week 2 for DCis included, although costs for drugs are not included.	'Drug Costs' sheet cell K77, and Table 66 (p194 CS).	Increases cost of DCis arm by about £1,000.	Corrected.

4.3.10.3 External validity

The company compared their modelled estimates of median overall and progression-free survival with results from the two RCTs that had been used to inform model efficacy parameters:

- SQUIRE trial[8]: see CS Table 77, and revised Table 24 in the clarification response (Appendix 1 p. 37), and
- Eastern Cooperative Oncology Group study E1594[29 44]: see CS Table 92 (p. 230).

These results are reproduced in Table 47 below. They show a good level of concordance between the model estimates and SQUIRE results. This is not surprising since these data were used as inputs for the model, and so this might more properly be seen as a form of 'internal validity', a check of correct model assumptions and coding. Concordance with E1594 results was also reasonable, although somewhat less good than with the SQUIRE data. Again, this is not surprising, since these data were included in the NMA, but alongside other studies.

Table 47 Comparison of model results with other sources of evidence

Outcome	Time (months)		Source
	Model Result	External estimate	
GCis + N			
PFS (median)	5.52	5.6 (5.4, 6.2)	SQUIRE
OS (median)	11.73	11.7 (9.6, 13.6)	SQUIRE
GCis			
PFS (median)	4.37	4.5 (4.2, 5.3)	SQUIRE
OS (median)	8.74	8.9 (8.1, 11.1)	SQUIRE
PFS (median)	4.60	4.3	E1594 (1)
OS (median)	10.35	9.4	E1594 (1)
PCarbo			
PFS (median)	4.6	3.7	E1594 (1)
OS (median)	10.35	9.3	E1594 (1)
DCis			
PFS (median)	4.14	3.1	E1594 (1)
OS (median)	8.97	8.1	E1594 (1)
US Markov model			
Life years gained	0.544	0.154	Goldstein et al. 2015
QALYs gained	0.338	0.111	

Note: This table is adapted from CS Table 92 and clarification response Appendix 1, Table 24, p. 37.

4.3.10.4 Cross validity

The CS does not report on cross validity. An independent NIH-funded economic evaluation has recently been published by Goldstein and colleagues.[20] They used a Markov model to estimate the cost-effectiveness of GCis + N compared with GCis for first-line treatment of patients with metastatic squamous NSCLC in a US context. Although the costs are not likely to be reflective of UK practice, and absolute rates of survival may differ between the US and UK, it is not unreasonable to compare the estimates of incremental effects (life years and QALYs gained), between the CS and Goldstein models (see section 4.3.10.4 above).

4.3.10.5 Predictive validity

As stated in the CS, assessment of predictive validity is not possible at this time.

4.4 Additional work undertaken by the ERG

We conducted a range of additional analyses to further test the robustness of the company model to changes in structural assumptions. This included an alternative 'base case' that reflects the ERG's judgement about the most plausible set of assumptions on which to ba cost-effectiveness estimates for the decision problem (see section 4.4.1 below). We then this base case to explore other possible scenarios and uncertainty over key parameters (s 4.4.2).

4.4.1 SHTAC base case analysis

Differences between our preferred set of assumptions and the company's are summarise Table 48. Our base case analysis was conducted using our replication of the company m This version of the model included corrections for the minor inconsistencies noted in Tabl above, and avoided the need to edit the complicated code and many data input sheets in submitted model.

Table 48 Base case specifications

	Company base case	SHTAC base case
Population	Western European (WE) EGFR expressing The company believes that results for the WE subgroup are more applicable for UK patients.	ITT EGFR expressing We do not believe that use of the WE subgroup is justified. It is a post hoc analysis, a significant treatment-subgroup interaction has not been shown for this (or for other prespecified regional sugroups), and there is no clear explanation for why the relative effects of treatment should differ in this selection of countries in particular.

	Company base case	**SHTAC base case**
Comparators	Direct comparison preferred (GCis+N vs. GCis only), due to weakness of NMA and robustness of SQUIRE dataset. Indirect comparison was also conducted, including GCis, PCarbo, GCarbo and DCis. The company explained in their response to clarification questions from NICE and the ERG that they had excluded PCis as marketing data suggested that it is used infrequently in the UK.	There is considerable uncertainty over the robustness of the NMA. However, we prefer to include the full range of comparators specified in the decision problem, and to incorporate uncertainty through probabilistic sensitivity analysis. We included all comparators for which data were available: GCis, GCarbo, PCarbo, DCis and PCis. We have received clinical advice that PCis is rarely used, but still consider that it should be included for completeness.
OS / PFS	KM curves extrapolated from last observation. For the direct comparson, the extrapolation was based on log-logistic survival curves, fitted separately to the SQUIRE arms. The indirect comparison used separately fitted Weibull curves, which were the best-fitting distribution subject to the proportional hazards assumption required for integration of NMA results.	KM extrapolated from the last timepoint with more than 20 patients remaining in each arm. This avoids undue weight on survival estimates based on very small numbers of patients. In our base case, we used separately fitted Weibull distributions, but also tested other parametric functions, including log-logistic curves for a direct comparsion of GCis+N and GCis.
TTD	KM curves from SQUIRE for GCis + N and GCis. Follow up was complete, so extrapolation was not needed. For indirect comparisions, the company assumed that HRs for TTD would equal those for PFS. And they tested the effect of assuming that TTD HRs for other comparators would equal those for GCis.	We used the same assumptions as in the company base case.
Adverse events	Risks estimated from SQUIRE for GC+N and GCis. For other comparators, it was assumed that the relative risks of AEs (versus those for GCis+N) would equal the GCis relative risks from SQUIRE.	Same approach used. Although the absolute numbers of some events were small, SQUIRE represents the best available data. Uncertainty over absolute and relative AE risks are reflected in the PSA.

	Company base case	SHTAC base case
2nd line therapy	For costing purposes, it was assumed that a proportion of patients would receive active treatment after progression with either erlotinib or docetaxel. These treatments were the most frequently used at second-line in SQUIRE. The proportions of patients on erlotonib and docetaxel were based on SQUIRE data, estimated separately by trial arm.	For our base case we followed the company's approach, but tested the assumption that both arms would have the same proportions of patients receiving erlotonib and docetaxel. We have received advice that use of erlotinib in this patient population would be rare in the UK. However, the effects of second-line treatment is implicitly built into the SQUIRE effectiveness data. It is therefore consistent to estimate the costs of this treatment.
Costs	Detailed costings based on data from SQUIRE on utilisation and appropriate unit cost data.	The company's approach to costing was thorough and appropriate. The only main exception is that they did not include a cost for a test for EGFR expression, that would be required to be consistent with the SmPC indication. We have not been able to identify a cost for this test, as it is not currently routinely available.
Utilities	SQUIRE data for the pre-progression health states. EQ-5D tariff scores were pooled across arms, but estimated separately for ITT EGFR and Western Europe EGFR subgroups. Khan et al. for post-progression utility. Various sources for AE disutilities.	For our base case analysis we followed the company approach. SQUIRE does represent the best available source of utility data that is consistent with the NICE reference case, and applicable to the patient population. Data presented in the CS and response to clarification questions does show similar utility scores between GCis and GCis+N groups. The utility estimates for the post-progression and adverse events are less robust for the current decision problem and are tested in sensitivity analysis.
Time horizon	Lifetime	Lifetime
Discounting	3.5% per year for costs and effects	3.5% per year for costs and effects
Analysis	Deterministic	Probabilistic

Key outputs from the company and SHTAC versions of the model are presented in Table 49. To aid comparison, results are presented for both versions based on the ITT analysis of the SQUIRE data for patients with EGFR expressing tumours. In other respects, the company and SHTAC results follow their respective base cases. The results are broadly similar. Both produce equivalent estimates of median, one year and two year OS and PFS – since they use the same KM data for this duration of follow up. The differences in modelled five year and mean OS and PFS, and hence in estimated QALYs relate to different assumptions about extrapolation beyond the KM data. Cost estimates are very similar for the two versions of the model.

Table 49 Model outputs: ITT EGFR expressing population

	Company model			SHTAC base case		
	GCis + N	GCis	Difference	GCis + N	GCis	Difference
Modelled OS						
Median (months)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Mean (months)	[redacted]	[redacted]	3.69	[redacted]	[redacted]	2.25
One year OS	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Two year OS	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Five year OS	4.9%	2.4%	2.5%	0.5%	0.3%	0.3%
Modelled PFS						
Median (months)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Mean (months)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
One year PFS	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Two year PFS	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Five year PFS	0.5%	0.4%	0.1%	0.0%	0.0%	0.0%
QALYs (undiscounted)						
Pre-progression	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Post-progression	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total QALYs	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Costs (undiscounted)						
Pre-progression (£)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Post-progression (£)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total (£)	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Mean costs and effects for the six interventions from the SHTAC base case model are shown in

Figure 14. In terms of effectiveness estimates, the interventions fell into three groups: GC
N, which had the highest estimated QALY; GCis and PCarbo, which had similar, intermediate QALY estimates; and PCis, DCis and GCarbo, with the lowest estimated QALYs. This ranking reflects the more favourable estimates from the NMA for PCarbo and GCis, than for PCis, DCis and DCarbo. However, the confidence intervals around the modelled QALY estimates were broad, and overlapped for all interventions. Further, although confidence intervals around the NMA HRs were incorporated in the PSA, correlations between NMA estimates and structural uncertainties about the conduct of the NMA and biases in the literature are not reflected in the results below.

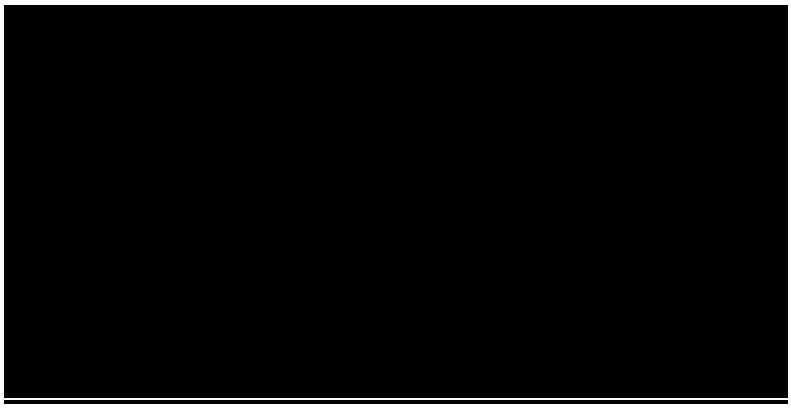


Figure 14 Estimated costs and effects: SHTAC base case analysis

Estimated costs were similar for all options, except for GCis + N which was a lot more expensive. In all of the incremental analyses conducted by the ERG, the relevant comparator for GCis + N was either GCarbo or GCis. The third set of comparators (PCis, DCis and GCarbo) offered fewer estimated QALYs for no or very small cost savings.

Cost-effectiveness results for the SHTAC base case analyses are shown in Table 50. These suggest an ICER of £169,612 per QALY, in comparison with the next best non-dominated comparator (in this case GCis). The related CEAC in Figure 15 illustrates the very high level of uncertainty over which treatment is most likely to be cost-effective. However, it does suggest a negligible probability that GCis + N is more cost-effective than the comparators at cost-effectiveness thresholds below £100,000 per QALY.

Table 50 Cost-effectiveness: SHTAC base case (ITT EGFR subgroup)

Technologies	Total		Incremental		ICER (£ per QALY)	Comparison
	Costs	QALYs	Costs	QALYs		
PCis	[redacted]	[redacted]			-	
DCis	[redacted]	[redacted]	_	_	-	Dominated
PCarb	[redacted]	[redacted]	£1,001	0.135	£7,429	vs PCis
GCarb	[redacted]	[redacted]			-	Dominated
GCis	[redacted]	[redacted]	£1,579	0.013	£124,663	vs DCis
GCis + N	[redacted]	[redacted]	£19,993	0.118	£169,612	vs GCis

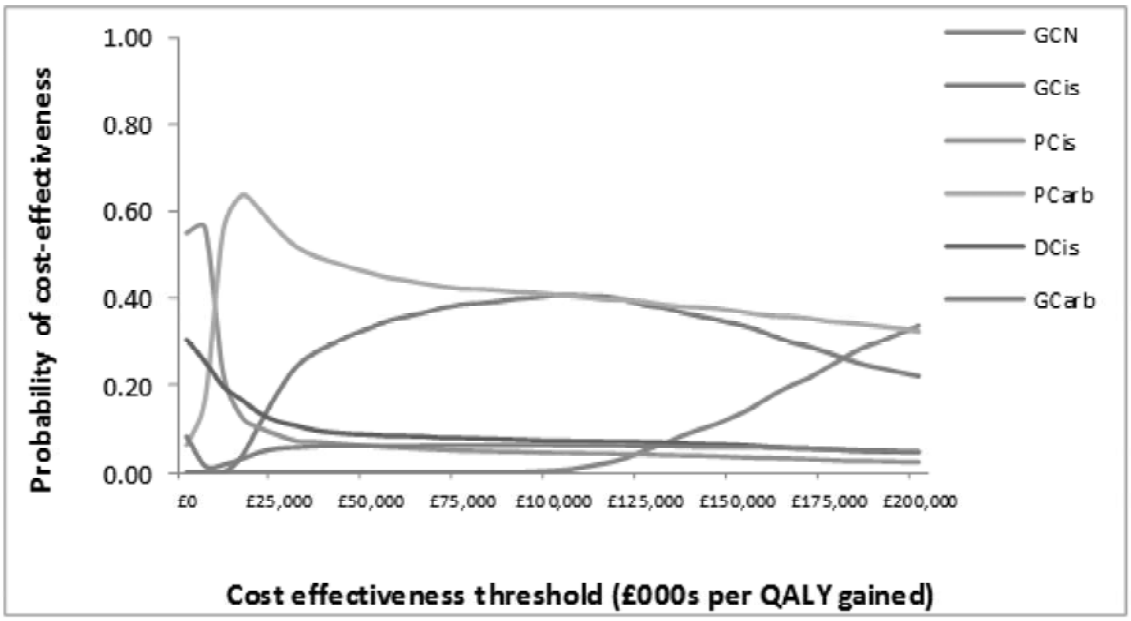


Figure 15 Cost-effectiveness acceptability curve: SHTAC base case analysis

4.4.2 Scenario and sensitivity analyses

The ERG conducted a range of sensitivity and scenario analyses on our base case version of the model (Table 51). We focussed on assumptions and parameter values to which the company model had been shown to be sensitive (CS pp. 223 to 232, and replacement Table 38 and Figure 22 in the clarification response A1):

- The SQUIRE subgroup used: the company model yielded a much higher estimated ICER for the ITT population than for their preferred Western Europe subgroup.
- The survival functions used to extrapolate OS beyond the KM data: Weibull curves have a shorter 'tail' than the log-logistic curves, producing a smaller QALY gain after the end of follow up at around three years, and hence a higher ICER. The use of exponential curves, or truncating the model after a time horizon of five years had a similar effect. Conversely, the use of a log-logistic model for GCis + N and Weibull for GCis had the effect of separating the tails of the survival curves, reducing the ICER.
- The results were very sensitive to parameter uncertainty over the OS, and to a lesser extent PFS, estimates: with higher ICERs at the upper 95% confidence limits for GCis + N, and lower ICERs at the lower 95% limits for GCis.
- Results were also somewhat sensitive to the time to treatment discontinuation in the GCis+N arm: in the model, a shorter duration of treatment had the effect of reducing costs with little impact on QALYs, resulting in a lower ICER.

In addition to these factors, the ERG was keen to test the timing of the transition from the KM curves to the parametric extensions. In particular, we were worried about basing very long extensions of OS and PFS (from about 3 to over 18 years) on the very small numbers of patients remaining in the SQUIRE trial at the end of follow up. As noted above, we set a prior criterion that the KM curves for OS and PFS would be used up to the time when more than 20 patients remained in both groups. However, this is an arbitrary cut-off, so we tested the effect of using the KM data directly to the final observation (as in the company model).

Changes to model assumptions relating to adverse event risks, utilities and resource use had minimal impact on cost-effectiveness results.

Table 51 shows summary results of additional analyses conducted by the ERG. The corr preferred analysis (scenario **C0**) yields an ICER of £57,725 per QALY gained for GCis + N compared with GCis. Introducing indirect comparisons, and changing the survival functio from log-logistic to Weibull to fit the proportional hazards assumption required for use of H from the NMA (scenario **C1**), increases the estimated ICERs for GCis + N compared with both

GCis and PCarbo. Changing from the Western European to ITT population (scenario **C2**) further increases the ICERs. The probabilistic version of this analysis (scenario **C3**) is similar to the SHTAC base case analysis (**S0**), although the latter does include further changes relating to the transition from KM to parametric survival curves, an additional comparator (PCis) and some minor corrections to the company version of the model.

The SHTAC base case (**S0**) yields an estimated ICER of £169,612 per QALY gained compared with GCis, which is the appropriate incremental comparator in this case. Scenarios **S1 to S8** illustrate the effect of different approaches to extrapolating the tail of the OS and PFS curves (or cutting the tail off by adopting a shorter time horizon in **S1**). The only scenario that makes an appreciable difference to the estimated ICER is **S8**, which adopts log-logistic tails for the GCis + N arm and Weibull for GCis. This reduces the ICER to £84,188 per QALY gained compared with GCis, or £109,214 with respect to the (correct incremental) comparator, PCarbo. The company have argued that this combination of survival curves presents a good fit for the two arms of the SQUIRE trial, but that it is inconsistent to use different functions. We are less concerned about the appropriateness of using different functions, given the decision to fit the parametric survival functions separately for the two arms, but are not convinced that this extrapolation provides a realistic estimate of longer-term survival within this population.

The final set of scenarios (**S9 to S16**), illustrate again the sensitivity of the ICER to changes in the OS, PFS and TTD estimates – re-iterating the conclusions of the company's tornado diagram (company clarification response A1 Figure 22).

Table 51 Summary of ERG additional analyses

Analysis		GCis+N versus GCis			GCis+N versus next best comparator			
		Incremental			Incremental			
		Cost	QALYs	ICER	Cost	QALYs	ICER	*Comp-arator*
C0	**Company - base case direct analysis**	**£19,516**	**0.3381**	**£57,725**				
C1	Company – direct and indirect analysis	£18,918	0.234	£80,912	£20,036	0.172	£116,344	*PCarbo*
C2	Company – C1 with ITT EGFR subgroup	£20,584	0.134	£153,947	£22,148	0.142	£155,654	*PCarbo*
C3	Company - C2 probabilistic	£20,591	0.134	£154,024	£21,999	0.116	£189,679	*PCarbo*
S0	**SHTAC - base case**	**£19,993**	**0.118**	**£169,612**	**£19,993**	**0.118**	**£169,612**	*GCis*
S1	SHTAC – five year time horizon	£19,976	0.117	£170,755	£19,976	0.117	£170,755	*GCis*
S2	SHTAC – KM for OS and PFS to final endpoint	£20,474	0.134	£153,085	£22,018	0.142	£154,569	*PCarbo*
S3	SHTAC – KM + Weibull OS and PFS (joint)	£20,037	0.123	£163,154	£21,596	0.132	£163,340	*PCarbo*
S4	SHTAC – KM + log-logistic OS and PFS (separate)	£20,571	0.149	£138,018	£20,571	0.149	£138,018	*GCis*
S5	SHTAC – KM + log-logistic OS and PSS (joint)	£20,608	0.156	£132,263	£20,608	0.156	£132,263	*GCis*
S6	SHTAC – Weibull OS & PFS (separate), no KM	£19,903	0.119	£167,233	£19,903	0.119	£167,233	*GCis*
S7	SHTAC – Log-logistic OS & PFS (separate), no KM	£20,514	0.142	£144,432	£20,514	0.142	£144,432	*GCis*
S8	SHTAC – Log-logistic GCis+N and Weibull GCis	£21,152	0.251	£84,188	£22,368	0.205	£109,214	*PCarbo*
S9	SHTAC – GCis OS at lower 95% limit	£20,427	0.185	£110,177	£21,572	0.131	£165,250	*PCarbo*
S10	SHTAC – GCis OS at upper 95% limit	£19,516	0.043	£457,474	£19,516	0.043	£457,474	*GCis*
S11	SHTAC – GCis+N OS at lower 95% limit	£19,337	0.039	£493,999	£19,337	0.039	£493,999	*GCis*
S12	SHTAC – GCis+N OS at upper 95% limit	£20,666	0.200	£103,574	£21,816	0.145	£150,426	*PCarbo*
S13	SHTAC – GCis+N PFS at lower 95% limit	£19,805	0.110	£180,194	£19,805	0.110	£180,194	*GCis*
S14	SHTAC – GCis+N PFS at upper 95% limit	£20,106	0.126	£159,862	£21,690	0.132	£163,922	*PCarbo*
S15	SHTAC – GCis+N TTD at lower 95% limit	£17,591	0.116	£151,908	£17,591	0.116	£151,908	*GCis*
S16	SHTAC – GCis+N TTD at upper 95% limit	£21,943	0.120	£183,597	£21,943	0.120	£183,597	*GCis*

4.5 Conclusions of cost-effectiveness

Overall, the company model provides a good foundation for assessing the cost-effectiveness of necitumumab for first line treatment of advanced squamous NSCLC. The model structure is appropriate and well implemented, with no major coding errors or inconsistencies that we could find. The key effectiveness estimates are based on a good quality RCT, and most other parameter estimates were of a good quality – or at least of best available quality. We agree with most of the assumptions in the company's base case analysis, with two notable exceptions. Firstly, we believe that the best available estimates of OS and PFS are from the ITT population in the SQUIRE trial. The company has argued for a narrower subgroup of Western European patients, based on post hoc analysis, but has not provided statistical evidence or a plausible explanation for whether and why necitumumab was more effective when added to conventional treatment for patients from this particular selection of countries.

The second set of assumptions in the company's base case analysis that we question, are those around the methods for extrapolating overall and progression-free survival beyond the observed RCT data. In particular, we question the particular choice of survival functions, and the way in which these have been attached to the sparse data at the end of follow-up. We believe that the methods used by the company exaggerate the proportion of patients likely to survive in the longer term, and therefore overestimate the projected estimates of QALYs gained. The company did provide alternative methods for extrapolation, and examined these in scenario analyses, and we have further explored these and some other plausible scenarios.

Our final, best estimate of the ICER for necitumumab compared with conventional platinum based chemotherapy for first line treatment of advanced squamous NSCLC is £169,612 per QALY gained. Although considerably higher than the company's base case estimate of £57,725, the company model yielded an ICER of £110,248 for the ITT (EGFR) population, with higher estimates based on less favourable assumptions about long-term survival.

Probabilistic analysis highlighted considerable uncertainty over the optimum choice of treatment for this group of patients, but the estimated probability of necitumumab being cost-effective was negligible below a cost-effectiveness threshold of less than £100,000 per QALY gained. Other uncertainties could not be quantified, including uncertainty over the completeness and robustness of the NMA, and the omission of relevant comparators. We are also concerned that

none of the estimated ICERs include the cost of testing patients for EGFR expression, as would be required to meet the SmPC indication. This cost was not included in the company model, and we have been unable to identify an estimate of the cost of this test, as it is not currently in routine use.

5 End of life

The company argues that necitumumab meets NICE's criteria in the 'Supplementary Advice for Appraising life-extending, end of life treatments'. The company states that expected survival in this patient population is less than 24 months (6.5 to 9.4 months, depending on the treatment used). The company further states that in the Western European subgroup, the modelled mean OS benefit for the Western Europe subgroup was 5.75 months for GCis + N compared with GCis alone. The company also indicates the treatment is indicated for an estimated small population of 2,575 patients in England with locally advanced or metastatic squamous NSCLC.

The ERG agrees that the company's estimate of the population size (CS Table 93, p. 233) appears reasonable, except that it includes all squamous NSCLC patients, not restricted to the EGFR expressing group (which is the SmPC indication). Therefore the patient population may be smaller than estimated.

The ERG agrees with the company that expected average survival in this population is less than 24 months. The ERG, however, does not agree with the company that GCis + N confers an additional survival benefit of 5.75 months compared with GCis alone. The company has used data from the Western European subgroup to support its argument, and the ERG believes that the company's rationale for basing efficacy conclusions on the Western European subgroup is unjustified. The ERG believes that the EGFR expressing subgroup from the ITT population is the most relevant population to the SmPC indication and this appraisal. Using the EGFR expressing subgroup data in the company's model (submitted in the clarifications response), the company's model showed a mean survival difference between GCis + N and GCis of 3.69 months, favouring GCis + N. The SHTAC base case analysis resulted in a mean survival difference of 2.25 months, favouring GCis + N. Therefore, when using the SHTAC base case, GCis + N does not meet this criterion in NICE's end of life criteria.

6 DISCUSSION

6.1 Summary of clinical effectiveness issues

The company identified one phase III trial in its review that was relevant to the decision problem. This was a well conducted, open-label, large trial including patients that are representative of those seen in clinical practice, based on clinical expert advice to the ERG. It provides direct evidence of the efficacy of GCis + N compared with GCis alone, the current gold standard treatment used in practice (for the fittest patients). The company argued that the results from a post-hoc Western European subgroup are the most generalisable to patients in England and used results from the SQUIRE trial from this population in the submitted economic model. The ERG, however, considers the company's rationale for this unjustified. The ERG considers the results from the EGFR expressing patient subgroup from the total population to be the most relevant population for this appraisal, as this is the SmPC indication. The company supplied additional analyses for this population during the appraisal.

The company also provided an NMA including 10 RCTs, which provided indirect and direct evidence for GCis + N versus PCarbo, GCis, PCis, DCis and GCarbo. Insufficient evidence was available to compare GCis + N with DCarbo, VCarbo and VCis. The NMA included patients with squamous NSCLC and was not limited by region or EGFR expression. Enough evidence was available for the company to conduct analyses of OS and PFS, but not HRQoL or AEs (which were also specified as outcomes in its inclusion criteria). The results from the OS and PFS analyses were used in the economic model. The ERG considers the treatment effects derived from the NMA are highly uncertain due to the lack of details provided about the studies included in the analysis and a number of methodological issues with how the NMA was conducted.

6.2 Summary of cost effectiveness issues

The company developed a model to estimate the cost-effectiveness of necitumumab in the context of the decision problem. This used a conventional model structure for cancer, with three main health states: pre-progression; post-progression and death. Before progression, patients were divided into three groups: those on induction treatment; those on maintenance with necitumumab (for the GCis + N arm); and those who have discontinued treatment but not yet progressed. The model included costs and utility impacts of adverse events, and costs for

active treatment and palliative care after progression. We believe this to be an appropriate model structure, and the model was implemented in a robust fashion. The effectiveness of GCis + N compared with GCis was based on good quality data from the SQUIRE trial, including: OS, PFS, TTD, AE risks and pre-progression utilities. Relative effects on OS and PFS for the other comparators were drawn from the NMA, and assumptions over TTD and AEs. Other parameters were estimated from the literature, and case note review.

Although the methods used for the company's economic evaluation were generally of a good quality, we do have some important reservations. In particular, the survival estimates that drive the cost-effectiveness results are based on a post-hoc subgroup analysis of SQUIRE, which we believe to be inappropriate. The company cite a much lower ICER (£57,725 per QALY gained) for this Western European subgroup than for the ITT population (£110,248 per QALY). We have further reservations about the methods that the company has used to extrapolate survival beyond the three years of follow-up that is available from SQUIRE. First, they extrapolate from the final endpoint of the KM data, which is based on a small number of patients and hence is subject to a high level of uncertainty. The company also choose a functional form for the extrapolation that gives a relatively wide separation between the tails of the survival curves (log-logistic), rather than another functional form (Weibull) that has a similar fit to the available data. This has the effect of increasing the estimated gain in life years, and hence in QALYs, from adding necitumumab to GCis. The company shows that alternative assumptions result in higher ICER estimates: for example, £79,412 per QALY gained if Weibull distributions are used to extrapolate beyond the OS KM endpoints.

The ERG's preferred set of modelling assumptions included use of the ITT (EGFR-expressing) patient group and Weibull extrapolations from points on the KM with more than 20 patients remaining per arm. This resulted in an estimated ICER of £169,612 per QALY gained. Sensitivity and scenario analysis showed considerable variation around this ICER. However, it remained high, and the estimated probability that GCis + N presents a cost-effective treatment option for first-line treatment for squamous NSCLC for patients with EGFR-expressing tumours is very low unless the maximum acceptable ICER is more than £100,000 per QALY gained.

7 REFERENCES

(1) Thatcher N, Hirsch FR, Luft AV, et al. Necitumumab plus gemcitabine and cisplatin versus gemcitabine and cisplatin alone as first-line therapy in patients with stage IV squamous non-small-cell lung cancer (SQUIRE): an open-label, randomised, controll phase 3 trial. *Lancet Oncology* 2015;16:763-74.

(2) European Medicines Agency. Necitumumab summary of product characteritics. Secondary Necitumumab summary of product characteritics 2016. http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/003886/WC500202694.pdf.

(3) National Institute for Health and Care Excellence. Lung cancer: diagnosis and management (CG121). London: NICE, 2011.

(4) Electronic Medicines Compendium (eMC). Docetaxel Hospira 10 mg/ml concentrate for solution for infusion (SPC). Secondary Docetaxel Hospira 10 mg/ml concentrate for solution for infusion (SPC) 2015. https://www.medicines.org.uk/emc/medicine/23925.

(5) Electronic Medicines Compendium (eMC). Gemcitabine 200 mg powder for solution for infusion, gemcitabine 1 g powder for solution for infusion, gemcitabine 2 g powder for solution for infusion (SPC). Secondary Gemcitabine 200 mg powder for solution for infusion, gemcitabine 1 g powder for solution for infusion, gemcitabine 2 g powder for solution for infusion (SPC) 2014. https://www.medicines.org.uk/emc/medicine/23656

(6) Electronic Medicines Compendium (eMC). Paclitaxel medac 6 mg/ml concentratic solution for infusion. Secondary Paclitaxel medac 6 mg/ml concentration for solution infusion 09.03.16 2015. https://www.medicines.org.uk/emc/medicine/21173.

(7) Electronic Medicines Compendium (eMC). Vinorelbine 10 mg/ml concentrate for s for infusion. Secondary Vinorelbine 10 mg/ml concentrate for solution for infusion 20 https://www.medicines.org.uk/emc/medicine/19453.

(8) Thatcher N, Hirsch FR, Luft AV, et al. Necitumumab plus gemcitabine and cisplati versus gemcitabine and cisplatin alone as first-line therapy in patients with stage IV squamous non-small-cell lung cancer (SQUIRE): an open-label, randomised, controlled phase 3 trial. *The Lancet Oncology* 2015;16(7):763-74.

(9) National Institute for Health and Care Excellence. Erlotinib for the treatment of no small-cell lung cancer (TA162). London, 2008.

(10) National Institute for Health and Care Excellence. Nintedanib for previously treated locally advanced, metastatic, or locally recurrent non-small-cell lung cancer (TA347). London, 2015.

(11) Eli and Lilly and Company. PORTRAZZA (necitumumab) injection, for intravenous use: highlights of prescribing information. Secondary PORTRAZZA (necitumumab) injection, for intravenous use: highlights of prescribing information 2015. http://www.accessdata.fda.gov/drugsatfda_docs/label/2015/125547s000lbl.pdf.

(12) Electronic Medicines Compendium (eMC). Cisplatin 1 mg/ml concentrate for solu for infusion. Secondary Cisplatin 1 mg/ml concentrate for solution for infusion 2015. https://www.medicines.org.uk/emc/medicine/25944.

(13) Pirker R, Pereira JR, Szczesna A, et al. Cetuximab plus chemotherapy in patients with advanced non-small-cell lung cancer (FLEX): an open-label randomised phase III trial. *The Lancet* 2009;373(9674):1525-31.

(14) Food and Drug Administration (FDA). FDA briefing document. Oncologic Drugs Advisory Committee Meeting July 9, 2015: BLA 125547/0 Necitumumab, 2015.

(15) Socinski MA, Luft A, Szczesna A, et al. Subgroup analyses by performance status (PS) in the phase III SQUIRE study: first-line necitumumab (N) plus gemcitabine-cisplatin (GC) vs. GC in squamous non-small cell lung cancer (NSCLC). *Journal of Clinical Oncology* 2015;33(15 Suppl 1).

(16) Socinski MA, Thatcher N, Luft AV, et al. Safety and efficacy of necitumumab (N) during single-agent treatment to progression following gemcitabine-cisplatin (GC) chemotherapy plus necitumumab (IMC-11F8/LY3012211) in the first-line treatment of patients with stage iv squamous non-small cell lung cancer (SQ-NSCLC) (SQUIRE). *International Journal of Radiation Oncology Biology Physics* 2014;90(5 Suppl 1).

(17) Ciuleanu TE, Socinski MA, Obasaju CK, et al. Safety and efficacy of necitumumab continuation therapy: subgroup analysis of phase 3 SQUIRE study. Journal of Clinical Oncology 2015;33(15 Suppl. 1).

(18) Reck M, Gralla RJ, Bonomi P, et al. Maximum severity score (MSS) of baseline patient-reported lung cancer symptom scale (LCSS) as a prognostic and predictive factor for overall survival (OS) in the Phase III SQUIRE study. *Journal of Clinical Oncology* 2015;33(15 Suppl 1).

(19) Brown T, Pilkington G, Bagust A, et al. Clinical effectiveness and cost-effectiveness of first-line chemotherapy for adult patients with locally advanced or metastatic non-small cell lung cancer: a systematic review and economic evaluation. *Health Technol Assess* 2013;17(31):1-278.

(20) Goldstein DA, Chen Q, Ayer T, et al. Necitumumab in Metastatic Squamous Cell Lung Cancer: Establishing a Value-Based Cost. JAMA Oncol 2015;1(9):1293-300.

(21) National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. NICE, London 2013.

(22) Beca J, Hoch JS. Exploring the impact of structural uncertainty in partitioned survival models for oncology. *Value in Health* 2014;17(3):A205-A06.

(23) Coyle D, Coyle K. The inherent bias from using partitioned survival models in economic evaluation. *Value in Health* 2014;17(3):A194.

(24) Goeree R, Villeneuve J, Goeree J, et al. Economic evaluation of nivolumab for the treatment of second-line advanced squamous NSCLC in Canada: a comparison of

modeling approaches to estimate and extrapolate survival outcomes. *Journal of Medical Economics* 2016:1-15.

(25) Sacco JJ, Botten J, Macbeth F, et al. The average body surface area of adult ca patients in the UK: a multicentre retrospective study. *PLoS One* 2010;5(1):e8933.

(26) Latimer NR. Survival analysis for economic evaluations alongside clinical trials-- extrapolation with patient-level data: inconsistencies, limitations, and a practical guid *Decis Making* 2013;33(6):743-54.

(27) Cancer Research UK. Lung cancer survival by stage at diagnosis. Secondary Lu cancer survival by stage at diagnosis 2013. http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/lung-cancer/survival#heading-

(28) cancer.org. Non-small cell lung cancer survival rates, by stage. Secondary Non-cell lung cancer survival rates, by stage 2014. http://www.cancer.org/cancer/lungcan non-smallcell/detailedguide/non-small-cell-lung-cancer-survival-rates.

(29) Hoang T, Dahlberg SE, Schiller JH, et al. Does histology predict survival of adva non-small cell lung cancer patients treated with platin-based chemotherapy? An anal the Eastern Cooperative Oncology Group Study E1594. *Lung Cancer* 2013;81(1):47-

(30) *AJCC Cancer Staging Manual*: Springer-Verlag New York, 2010.

(31) Chouaid C, Agulnik J, Goker E, et al. Health-Related Quality of Life and Utility in Patients with Advanced Non–Small-Cell Lung Cancer. *J Thorac Oncol* 2013;8:997-1003.

(32) Iyer S, Taylor-Stokes G, Roughley A. Symptom burden and quality of life in advanced non-small cell lung cancer patients in France and Germany. *Lung Cancer* 2013;81(2):288-93.

(33) Khan I, Morris S, Hackshaw A, et al. Cost-effectiveness of first-line erlotinib in patients with advanced non-small-cell lung cancer unsuitable for chemotherapy. *BMJ Open* 2015;5(7):e006733.

(34) Papaioannou D, Brazier J, Paisley S. NICE DSU Technical Support Document 9: The identification, review and synthesis of health state utility values from the literature, 2010.

(35) Nafees B, Stafford M, Gavriel S, et al. Health state utilities for non small cell lung cancer. *Health Qual Life Outcomes* 2008;6:84.

(36) Doyle S, Lloyd A, Walker M. Health state utility scores in advanced non-small cell lung cancer. *Lung Cancer* 2008;62(3):374-80.

(37) Locadia M, Bossuyt PMM, Stalmeier PFM, et al. Treatment of venous thromboembolism with vitamin K antagonists: patients' health state valuations and treatment preferences. *Thromb Haemost* 2004;92:1336-41.

(38) Department of Health. Drugs and pharmaceutical electronic market information (eMit). Secondary Drugs and pharmaceutical electronic market information (eMit) 2014.

(39) Joint Formulary Committee. *British National Formulary (BNF68)*. London: BMJ Group and Pharmaceutical Press, 2014.

(40) Curtis L. Unit Costs of Health and Social Care 2014: Personal and Social Servic Research Unit, 2015.

(41) Department of Health. NHS reference costs 2013-2014. http://wwwdhgovuk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndG e/DH_122803 2015.

(42) Department of Health. Drugs and pharmaceutical electronic market information (Secondary Drugs and pharmaceutical electronic market information (eMit) 2015. https://www.gov.uk/government/publications/drugs-and-pharmaceutical-electronic-ma information-emit.

(43) Eddy DM, Hollingworth W, Caro JJ, et al. Model transparency and validation: a r of the ISPOR-SMDM Modeling Good Research Practices Task Force--7. *Value Healt* 2012;15(6):843-50.

(44) Schiller J, Harrington D, Belani CP, et al. <2002 Schiller Four way comparison nejmoa011954.pdf>. N Eng J Med 2002;346(2):92-98.