STA REPORT

LIVERPOOL REVIEWS AND IMPLEMENTATION GROUP (LRiG)

Nintedanib for previously treated locally advanced or metastatic non-small cell lung cancer

This report was commissioned by the NIHR HTA Programme as project number 13/106/01

Completed 15 October 2014

Contains AIC and CIC

UNIVERSITY OF LIVERPOOL

LIVERPOOL REVIEWS AND IMPLEMENTATION GROUP

A MEMBER OF THE RUSSELL GROUP

Title: Nintedanib for previously treated locally advanced or metastatic non-small cell lung cancer

Produced by: Liverpool Reviews & Implementation Group (LR*i*G)

Authors: Nigel Fleeman, Research Fellow (Clinical Effectiveness), LR*i*G, University of Liverpool

Adrian Bagust, Professor (Health Economic Modelling), LR*i*G, University of Liverpool

Angela Boland, Assistant Director, LR*i*G, University of Liverpool

Kerry Dwan, Research Fellow (Medical Statistician), LR*i*G, University of Liverpool

Marty Richardson, Research Fellow (Medical Statistician), LR*i*G, University of Liverpool

Sophie Beale, Research Associate (Decision Analysis), LR*i*G, University of Liverpool

Eleanor Kotas, Information Specialist, LR*i*G, University of Liverpool

Lindsay Banks, North West Medicines Information Centre, Pharmacy Practice Unit, Liverpool

Alastair Greystoke, Honorary Consultant in Medical Oncology Newcastle Upon Tyne NHS Foundation Trust

Correspondence to: Nigel Fleeman, Research Fellow, Liverpool Reviews and Implementation Group, University of Liverpool, Room 2.10, Whelan Building, The Quadrangle, Brownlow Hill, Liverpool L69 3GB

Date completed: 15 October 2014

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 13/106/01

Declared competing interests of the authors: None

Acknowledgements: The authors would like to thank Paula Scullin (Consultant Oncologist, Belfast HSC Trust) and Julian Nam (Health Economist, University of Glasgow) who provided feedback on draft versions of the report.

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

This report should be referenced as follows: Fleeman N, Bagust A, Boland A, Dwan K, Richardson M, Beale S, Kotas E, Banks L, Greystoke A. Nintedanib for previously treated locally advanced or metastatic non-small cell lung cancer: A Single Technology Appraisal. LR*i*G, University of Liverpool, 2014

Contributions of authors:

Fleeman N	Project lead, drafted clinical results section and supervised the final report
Bagust A	Checking and validation of the economic model and critique
Boland A	Critical appraisal of clinical and economic evidence
Dwan K	Critical appraisal of the statistical evidence
Richardson M	Critical appraisal of the statistical evidence
Beale S	Critical appraisal of clinical and economic evidence
Kotas E	Cross checking of the submission search strategy
Banks L	Critical appraisal of the submission
Greystoke A	Clinical advice and critical appraisal of the clinical sections of the manufacturer's submission

All authors read and commented on draft versions of the ERG report.

Table of contents

List of tables

List of figures

LIST OF ABBREVIATIONS

AE	adverse event
AESI	adverse events of special interest
AIC	Akaike information criteria
ALT	alanine aminotransferase
AST	aspartate transaminase
AUC	area under the curve
BSA	body surface area
BSC	best supportive care
BD	twice daily
CEAC	cost-effectiveness acceptability curve
CI	confidence interval
CS	company's submission
CTCAE	Common Terminology Criteria for Adverse Events
CTR	clinical trial report
D	Death
ECOG	Eastern Cooperative Oncology Group
EGFR	epidermal growth factor receptor
EMA	European Medicines Agency
ERG	Evidence Review Group
HR	hazard ratio
HRQoL	health related quality of life
ICER	incremental cost-effectiveness ratio
ITT	intention-to-treat
K-M	Kaplan-Meier
LUCADA	National Lung Cancer Audit database
MTC	mixed treatment comparison
NSCLC	non-small-cell lung cancer
ORR	overall response rate
OS	overall survival
PD	progressive disease
PF	progression-free
PFS	progression-free survival
PPS	post-progression survival
PS	performance status
PSA	probability sensitivity analysis
QALY	quality adjusted life year
QoL	quality of life
RCT	randomised controlled trial
RDI	Relative Dose Intensity
RECIST	Response Evaluation in Solid Tumours
SAEs	serious adverse events
SEER	Surveillance, Epidemiology and End Results
SmPC	summary of product characteristics
TKI	tyrosine-kinase inhibitor
ToT	time on treatment
TSAP	trial statistical analysis plan
TTD	time to deterioration
VEGF	vascular endothelial growth factor
VEGFR	vascular endothelial growth factor receptors

1 SUMMARY

1.1 Scope of the submission

The remit of the Evidence Review Group (ERG) is to comment on the clinical and cost-effectiveness evidence submitted to the National Institute for Health and Care Excellence (NICE) as part of the single technology appraisal (STA) process. Clinical and economic evidence have been submitted to NICE from Boehringer Ingelheim in support of the use of nintedanib (Vargatef) for previously treated locally advanced or metastatic non-small cell lung cancer (NSCLC) of adult patients with adenocarcinoma tumour histology.

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

The population specified in the scope is adults with locally advanced or metastatic NSCLC that has progressed following prior chemotherapy. The decision problem addressed by the company is patients with locally advanced, metastatic or locally recurrent NSCLC of adenocarcinoma tumour histology who had previously received first-line chemotherapy. This is in line with the anticipated full marketing authorisation for nintedanib which differs to that of the scope by including the term "locally recurrent" and restricting NSCLC to adenocarcinoma. The ERG notes that to be classified as locally recurrent, a patient would initially present with early stage disease (stage I, II or IIIA) and be treated with surgery or radical radiotherapy and then relapse in the same area without metastases. Since the anticipated license also stipulates patients must have previously received first-line chemotherapy, then all patients would have locally advanced or metastatic disease at the time of second-line treatment. Treatment for locally advanced (be it recurrent or present since diagnosis) or metastatic disease at this point in the disease course is identical.

The anticipated license also specifies that nintedanib should be administered in combination with docetaxel. Both docetaxel monotherapy and erlotinib monotherapy are considered as comparators in the company's submission (CS). However the company states that erlotinib is not a relevant comparator to nintedanib plus docetaxel and this is only considered a comparator by the company for secondary analyses. The ERG agrees with the company that erlotinib is not a relevant comparator. A preliminary recommendation by NICE in February and August 2014 is that erlotinib should not be recommended for treating locally advanced or metastatic NSCLC that has progressed after non-targeted chemotherapy in people with tumours that are EGFR-negative. Furthermore, in clinical practice the ERG notes that the majority of patients with EGFR-positive disease will already have received erlotinib (or another tyrosine-kinase inhibitor [TKI]) as first-line treatment so would not receive erlotinib as a second-line treatment. Finally, patients who would be considered fit enough (i.e. Eastern Cooperative Oncology Group [ECOG] performance status [PS] 0 to 1) to receive nintedanib

would also need to be considered fit enough to receive docetaxel (since docetaxel is administered in combination with nintedanib). Hence only docetaxel is considered to be a relevant comparator by the ERG.

Clinical evidence is presented for all outcomes specified in the scope and cost-effectiveness results are expressed in terms of incremental cost per quality adjusted life year (QALY) gained. No subgroups were specified in the decision problem and no equality issues were identified.

1.3 Summary of clinical effectiveness evidence submitted by the company

Direct evidence is presented for nintedanib plus docetaxel vs placebo plus docetaxel from one phase III double-blind randomised controlled trial (RCT) (LUME-Lung 1). The company states that as not all patients in LUME-Lung 1 had histology of adenocarcinoma but that as patients who did not have adenocarcinoma are expected to be outside the licensed population only data for patients with adenocarcinoma are presented. While some of these patients had locally recurrent, as opposed to locally advanced or metastatic disease at diagnosis, the vast majority (94.2%) had metastatic disease at screening.

The findings from LUME-Lung 1 suggested that nintedanib plus docetaxel significantly improve progression-free survival (PFS) and overall survival (OS) in comparison to placebo plus docetaxel. The gain in median PFS is 1.2 months (4.0 months vs 2.8 months; hazard ratio [HR] 0.77, 95% confidence interval [CI]: 0.62 to 0.96) based on the primary analysis with a median follow-up of 7.1 months. Based on the final analysis, after a median follow-up of 31.7 months, the gain in PFS is 1.4 months (4.2 months vs 2.8 months; HR 0.84, 95% CI: 0.71 to 1.00). The gain in median OS is 2.3 months (12.6 months vs 10.3 months; HR 0.83, 95% CI: 0.70 to 0.99). Pre-specified and post-hoc subgroup analyses for both PFS and OS support the findings for the population of patients with adenocarcinoma as a whole.

Specific adverse events (AEs) occurring more often in the nintedanib plus docetaxel arm than in the placebo plus docetaxel arm and considered to be AEs of special interest (AESIs) were diarrhoea (43.4% vs 24.6%), nausea (28.4% vs 17.7%) and vomiting (19.4% vs 12.3%). These AEs were successfully managed by dose reduction, dose interruption and/or symptomatic treatment and led to permanent nintedanib discontinuation in <1% of patients. Other reported AESIs associated with nintedanib treatment included increases in alanine aminotransferase (ALT) (37.8% vs 9.3%) and aspartate transaminase (AST) (30.3% vs 7.2%). These were reported to be generally reversible and led to permanent nintedanib discontinuation in <2% of patients. The incidence of Common Terminology Criteria for

Adverse Events (CTCAE) grade ≥3 AEs and CTCAE grade ≥3 SAEs were greater in the nintedanib plus docetaxel arm (75.9% and 31.3%) than the placebo plus docetaxel arm (68.5% and 26.6%). The AEs of greatest concern were fatal AEs and some imbalances were reported between treatment arms; fatal AEs being more common in the nintedanib plus docetaxel arm (6.3%) compared to the placebo plus docetaxel arm (2.4%). However, the company considers that these figures may be partially confounded by a longer median duration of treatment with nintedanib/placebo (4.2 months vs 3.0 months respectively) and docetaxel (median 5 and 4 cycles in the intervention and comparator arms respectively).

There was no significant difference over time, or between arms, in global health status/quality of life (QOL) or self-reported health related quality of life (HRQoL) assessments for cough, dyspnea or pain in LUME-Lung 1. Statistically significant improvements were observed for three individual pain items ('have pain', 'pain in chest' and 'pain in arm and shoulder') in favour of nintedanib plus docetaxel, while time to deterioration (TTD) for diarrhoea was significantly worsened in this arm.

Additional evidence is presented for nintedanib plus docetaxel compared to docetaxel and erlotinib by means of mixed treatment comparisons (MTCs) and, where possible, Bucher indirect comparisons. Compared to docetaxel, the base-case MTC analyses (which include four trials) report significant improvements in OS (HR 0.83, 95% CI: 0.70 to 0.99) and PFS (HR 0.77, 95% CI: 0.62 to 0.96) with the addition of nintedanib. The base-case MTC analyses also report significant improvements in OS (HR 0.64, 95% CI: 0.46 to 0.90) and PFS (HR 0.70, 95% CI: 0.50 to 0.998) for nintedanib plus docetaxel compared to erlotinib. The Bucher indirect comparisons (which includes two trials) support these findings (OS HR 0.56, 95% CI: 0.38 to 0.82; PFS 0.58, 95% CI: 0.39 to 0.87). Scenario analyses (including three of the trials from the base-case plus an additional trial) and sensitivity analyses of the base-case (including eight trials) and scenario analyses (including eight trials) were also conducted. These analyses all broadly support the base-case findings. For overall response rate (ORR), the base-case results suggest that there was no significant difference between nintedanib plus docetaxel in comparison with docetaxel or erlotinib.

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

The ERG is satisfied with the search strategy employed by the company to identify clinical effectiveness studies. It is not aware of any additional relevant ongoing or completed studies relevant to the decision problem.

The ERG is of the opinion that the LUME-Lung 1 study is well-designed and conducted, with low risk of bias. However, eligibility criteria mean that the patient population may not be representative of patients generally seen in clinical practice in England. Specifically, the trial excludes patients with any major pleural effusion or evidence of cavitary or necrotic tumours and therapeutic anticoagulation (except low dose heparin) or antiplatelet therapy (except for chronic low-dose therapy with acetylsalicylic acid ≤325mg/day). In addition, the proportion of patients aged ≥65 years is relatively small (28.3%) and such patients may have a poorer prognosis than younger patients. Given the focus of the decision problem on patients with adenocarcinoma, the ERG agrees it was appropriate for the company to only present data from LUME-Lung 1 for this patient population. Notwithstanding the exclusions of certain types of patients referred to above, the patient population is similar to the adenocarcinoma population likely to be treated for locally recurrent, locally advanced or metastatic disease in clinical practice in England. However, perhaps as a result of the eligibility criteria, it is noted that the rate of post-study therapy is relatively high (55.8%) which suggests this is an atypically fitter patient population than would be found in clinical practice in England. This is, however, not uncommon in clinical trials.

The ERG does not consider a comparison of nintedanib plus docetaxel to erlotinib is appropriate to decision problem. However, this was specified in the NICE scope and the company has therefore undertaken such a comparison via MTCs. The ERG has identified a number of methodological limitations related to the conduct of the MTCs (explored below in section 1.9.2) and advises that results from the MTCs should be treated with caution.

1.5 Summary of cost effectiveness evidence submitted by the company

The company developed a de novo partitioned survival Markov model that comprises three health states: progression-free (on or off treatment), progressive disease (PD) and death. All patients enter the model in the progression-free state. The model, when projecting PFS and OS data from LUME-Lung 1, fits a variety of standard parametric functions to the available trial data. Variants of this model structure have been used in the modelling of metastatic oncology for a number of previous NICE STAs. The model has been developed in Microsoft Excel using a 3-weekly cycle length. It includes a half-cycle correction and the time horizon is set at 15 years. As recommended by NICE, a discount rate of 3.5% has been used for both costs and outcomes; outcomes are measured in QALYs. The model perspective is that of the UK NHS. Resource use, costs and utilities were estimated based on information from LUME-Lung 1, published sources and clinical experts.

For the comparison of nintedanib plus docetaxel vs docetaxel, the company's incremental cost-effectiveness ratio (ICER) per QALY gained is £50,776. For the comparison of nintedanib plus docetaxel vs erlotinib, the company's ICER per QALY gained is £27,008. The company carried out a wide range of deterministic sensitivity analyses for these two comparisons. The results from the ten parameters that had the most influence on the ICER per QALY gained ranged from £44,034 to £59,711 for nintedanib plus docetaxel vs docetaxel and from £17,721 to £238,678 for nintedanib plus docetaxel vs erlotinib (in the latter comparison, the HR for OS was the single most influential variable). The results of the company's probabilistic sensitivity analysis (PSA) suggest that for nintedanib plus docetaxel vs docetaxel, there is a 2% and a 50% chance of nintedanib plus docetaxel being cost-effective at willingness to pay thresholds of £30,000 and £50,000 per QALY gained respectively; and a 65% and 94% chance of nintedanib plus docetaxel being cost-effective compared to erlotinib using the same thresholds.

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

The ERG is satisfied with the search strategy employed by the company to identify cost-effectiveness studies and is reasonably confident that no other relevant published articles exist.

Overall, the ERG found the company's model to be well structured. For most functions the assumptions and options are labelled and annotated where necessary; however, in some cases, the ERG has found it difficult to confirm details of the data sources employed (e.g. analyses related to Surveillance, Epidemiology and End Results program [SEER] and the National Lung Cancer Audit database [LUCADA]). The ERG identified eleven factors that limit confidence in the reliability of the company's model and/or results. These relate to: inappropriate methods used to project time-to-event outcomes (OS, PFS and time-on-treatment); mid-cycle adjustment error; inappropriate methods used to estimate cost of treatment doses; underestimate of true cost of febrile neutropenia; monitoring costs; non-UK standard approach to discounting; overall average disutility estimate for fatigue used for both regimens; error in stable disease costs and erroneous restriction of docetaxel to four cycles. The ERG is concerned by the number of implementation errors that have been identified, some of which have important consequences for the size of the estimated ICER per QALY gained for the comparison of nintedanib plus docetaxel vs docetaxel.

The ERG does not consider a comparison of nintedanib plus docetaxel to erlotinib is appropriate to decision problem. However, this was specified in the NICE scope and the company has therefore undertaken such a comparison. The ERG considers that this is

seriously flawed due to inconsistencies apparent in the available time-to-event data leading to conflicting results from the MTC. The ERG has applied other relevant amendments to the submitted model for this comparison, but the uncertainty in OS, PFS and time on treatment (ToT) probably far outweighs all other effects but cannot be quantified.

1.7 Summary of company's case for end of life criteria being met

The company makes a case that nintedanib plus docetaxel meets the criteria set by NICE for end of life treatment. Namely:

- The life expectancy of the patient population was short (< 24 months). Patients with advanced NSCLC have a short life expectancy of less than 24 months on average. Using the extrapolated results from the LUME-Lung 1 trial data implemented in the cost effectiveness model, the median OS of patients on docetaxel monotherapy (current standard of care) is 10.23 months and the mean OS is 15.96 months.
- The number of patients who would be eligible for the treatment is small. The total eligible population in England for nintedanib plus docetaxel based on the anticipated marketing authorisation is estimated to be 703.
- The increase in OS is >3 months. Extension to life due to nintedanib plus docetaxel vs docetaxel monotherapy in the target population with the base-case assumptions within the model is a mean of 3.96 months. The extension in OS over erlotinib is a mean of 5.16 months.

1.8 ERG commentary on end of life criteria

The ERG agrees that patients with advanced NSCLC have a life expectancy of less than 24 months. It also agrees that only a small number of patients would be eligible for treatment with nintedanib plus docetaxel. By applying the Kaplan-Meier (K-M) trial results using the area under the curve (AUC) method until the long-term OS trends were established and then projecting remaining estimated survival using exponential trends, the ERG calculated the extension in mean OS to be 3.05 months for nintedanib plus docetaxel compared with docetaxel. It was not possible for the ERG to derive a mean estimate for OS gain for nintedanib plus docetaxel vs erlotinib.

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

1.9.1 Strengths

Clinical effectiveness

The ERG considers LUME-Lung 1 presents good quality evidence of clinical effectiveness which is directly relevant to the decision problem.

Cost-effectiveness

The company presented comprehensive and very detailed economic sections both within the CS and in the supplementary evidence. The company attempted to fully address the NICE scope. The ERG's requests to the company for additional economic analyses and further information were completed on time and to a high standard.

1.9.2 Weaknesses and areas of uncertainty

Clinical effectiveness

The ERG considers the MTCs are unnecessary because erlotinib is a comparator of no relevance to the vast majority of the patient population that would be considered for treatment with nintedanib plus docetaxel. The ERG further observes that LUME-Lung 1 is the only trial in which any patients (18.8%) received pemetrexed as a first-line treatment, as is now typically the case in clinical practice in England and so, arguably, all of the other trials included in the MTCs are of limited relevance to the decision problem. There are also other major methodological weaknesses and areas of uncertainty with the conduct of the MTCs, namely:

1. the proportional hazards assumption is not supported by the LUME-Lung 1 trial data for PFS or OS. Thus any estimation of the relative effectiveness of nintedanib plus docetaxel vs erlotinib (i.e. a calculated HR) will lack credibility and be effectively meaningless

2. differences in trial and patient characteristics mean that there is heterogeneity across trials which suggests that comparing data from these trials is inappropriate

Methodological issues also exist, namely: the use of both unadjusted and adjusted PFS and OS data, the use of PFS assessed by central independent review and local investigators and the use of primary PFS as opposed to updated PFS from LUME-Lung 1. However, these are not considered by the ERG to have major importance, particularly given the weaknesses and areas of uncertainty identified previously.

A greater number of fatal AEs have been observed in the nintedanib plus docetaxel arm than in the placebo plus docetaxel arm of the LUME-Lung 1 trial. However, the numbers are small and the company is using ongoing surveillance to monitor this issue.

Whilst LUME-Lung 1 is directly relevant to the decision problem, specific exclusion criteria employed in this trial may have excluded some patients who would ideally be considered for treatment in clinical practice in England. These are patients with major pleural effusion,

evidence of cavitary or necrotic tumours, or receiving therapeutic anticoagulation (except low dose heparin) or antiplatelet therapy (except for chronic low-dose therapy with acetylsalicylic acid ≤325mg/day). This may also partially explain why a higher proportion of patients in the trial than would be expected in clinical practice in England received third-line treatment.

Cost-effectiveness

The ERG identified a number of weaknesses and areas of uncertainty in the company's model for the comparison of nintedanib plus docetaxel vs docetaxel. The ERG considers that the high number of implementation errors is a major weakness of the model. These errors are present in estimates of both costs and benefits and therefore influence the size of the base-case ICER per QALY gained in a number of ways (mostly resulting in increasing the size of the ICER).

The most important area of uncertainty identified by the ERG is related to OS estimation. The company used a Log-Logistic survival model, whereas the ERG used the unadjusted trial data directly for the majority of patients, followed by projecting long-term survivors using trends evident in the data set. The company used data from the SEER and LUCADA to support the parametric survival modelling applied in the model. However, it was not possible for the ERG to assess whether this approach was valid; the analyses reported by the company did not provide references for the specific data sets used, nor did the company present sufficient explanation of the data employed. When the ERG replaced the company's preferred OS model with the ERG's preferred OS model, there was a major impact on the size of the ICER per QALY gained; it increased substantially as the size of the ERG's estimated OS incremental gain was reduced.

The ERG does not consider the company's comparison of nintedanib plus docetaxel vs erlotinib to be relevant to the decision problem. Furthermore, even if the comparison was considered to be relevant, the ERG has noted a number of flaws in the company's MTCs that render the clinical effectiveness results unreliable. The ERG considers that these problems are so fundamental that it is not possible to rectify them and modify the company's model to provide improved estimates of OS, PFS and the relative cost-effectiveness of nintedanib plus docetaxel and erlotinib.

1.10 Summary of exploratory and sensitivity analyses undertaken by the ERG

For the comparison of nintedanib plus docetaxel vs docetaxel, the company's base-case ICER (£50,776 per QALY gained) would increase to £85,292 per QALY gained if all 11 ERG

recommended revisions were applied and would increase to £82,995 per QALY gained if all but the limit on the number of cycles of docetaxel treatment were applied.

The ERG has been unable to estimate an ICER for the comparison of nintedanib plus docetaxel vs erlotinib for the reasons stated in the ERG's critique of the clinical effectiveness and cost-effectiveness evidence.

2 BACKGROUND

2.1 Critique of company's description of underlying health problems

section 2.1 of the CS[1] provides a brief overview of NSCLC.sections 2.1 and 2.2 of the CS[1] provide data on the number of patients with NSCLC and section 2.3 provides details about the life expectancy of people with NSCLC in England. These sections appropriately present the key issues relating to the underlying health problems of patients with NSCLC and are summarised as presented in the CS[1] in Box 1.

Box 1 Lung cancer disease course and epidemiology

Types of lung cancer

- [Non-small cell lung cancer (NSCLC) is] the most common type, accounting for 85% to 90% of cases
- Adenocarcinoma is the most common [40%] histological sub-type of NSCLC[2]
- Patients with NSCLC have a poor prognosis that has not changed significantly in the past decades

The disease course

- Lung cancer does not usually cause noticeable symptoms until it is locally advanced or has spread through much of the lungs or into other parts of the body (i.e. metastatic lung cancer)
- This means that the outlook for lung cancer is poor compared with other types of cancer[3]

Epidemiology

- Lung cancer is the second most common cancer in the UK; there are around 41,500 new cases diagnosed each year, with 35,406 new cases in England and Wales in 2010, and more than one in five cancer deaths (22%) in the UK are from lung cancer[4]
- Smoking causes more than 8 in 10 lung cancers in the UK[5]
- At diagnosis, 10 to 15% of patients have locally advanced cancer, i.e. stage IIIB and 40% of patients have metastatic cancer i.e. stage IV[6,7]
- Moreover, patients with stage IIIB and stage IV NSCLC have the lowest 5-year survival rate, at 5% and 1%, respectively[2,8-10]

In relation to epidemiology, the ERG adds that the LUCADA database audit published in 2012 reported approximately 57% of patients with NSCLC were stage IIIB or stage IV.[11] This figure is consistent with the estimates presented by the company in Box 1 (50% to 55%). A recent National Institute for Health Research Horizon Scanning Centre document[12] states that the incidence of stage III/IV NSCLC is 78%. This implies an incidence of stage IIIA disease of around 20% to 30% if the estimates for stage IIIB and IV cited by the company and ERG are subtracted.

2.2 Critique of company's overview of current service provision

As stated in section 2.1 of the CS,[1] the type of treatment that patients with locally advanced or metastatic NSCLC receive depends on several factors, including, but not limited to, tumour histology and EGFR mutation status. Patients with mutation free (i.e. EGFR-negative) locally advanced or metastatic lung cancer usually receive platinum doublet chemotherapy in the first-line setting, typically pemetrexed plus cisplatin for patients with adenocarcinoma.[3] As stated in section 2.5 of the CS,[1] TKIs - erlotinib, gefitinib or afatinib - are all NICE recommended options[13-15] for patients with EGFR-mutations. At present all three of these drugs have been made available to NHS patients at discount prices, as set out in patient access schemes.

According to the company, approximately 30%[16] to 50%[17] of patients with locally advanced or metastatic NSCLC receive second-line treatment. The current options for second and subsequent lines of treatment, as stated in sections 2.1 and 2.5 of the CS,[1] are summarised in Box 2. The company's advisory board, which comprised five clinical experts, estimated that [redacted] of all patients who had received second-line treatment would go on to receive third-line treatment, with approximately one third of these patients receiving this treatment as part of an ongoing clinical trial.[1] The company's own data on file[18] that reports on data from the final quarter of 2012 appears to support this view. These data show that 13.33% of patients who received second-line treatment also received third-line cytotoxic treatment.

Box 2 Current service provision for patients with NSCLC following first-line treatment

Second-line treatment

- The major goal of second-line treatment is to prolong life without worsening HRQ[o]L
- There are a number of new therapies that target patients with relatively rare mutations (e.g. EGFR), but patients with adenocarcinomas and without actionable mutations [e.g. EGFR] who progress following first-line chemotherapy have limited therapy options
- Following failure of first-line chemotherapy, treatment options are limited to docetaxel monotherapy or erlotinib [19,20]
- Docetaxel monotherapy can be considered for second-line treatment of locally advanced or metastatic NSCLC when cancer has relapsed after previous chemotherapy
- Erlotinib is recommended, within its licensed indication, as an alternative to docetaxel as a second-line treatment option for patients with NSCLC only on the basis that it is provided by the manufacturer at an overall treatment cost (including administration, AEs and monitoring costs) equal to that of docetaxel

Third-line treatment and subsequent lines of therapy

- Currently, there are no NICE-recommended technologies

In section 2.6 of their submission,[1] the company notes that the use of erlotinib as a second-line treatment is being reviewed by NICE and presents recommendations issued by NICE in February 2014. The ERG notes that this guidance is in the process of being replaced, with draft guidance published on the NICE website on 7 August 2014. One of the Appraisal Committee's preliminary recommendations[21] from both February and August 2014 is that

erlotinib should not be recommended for treating locally advanced or metastatic NSCLC that has progressed after non-targeted chemotherapy in people with tumours that are EGFR-negative. Furthermore, erlotinib is only recommended in second-line treatment for patients with tumours that are EGFR-positive, or of unknown status, in limited circumstances (Box 3).

In addition, the ERG notes that as the recommended first-line treatment for patients with tumours that are EGFR-positive is a TKI,[13-15] there are unlikely to be many patients with EGFR-positive tumours for whom erlotinib is considered an appropriate second-line treatment. Furthermore, as noted on page 35 of the CS,[1] the opinion of clinical experts is that patients who are sufficiently fit to allow them to tolerate treatment with docetaxel receive docetaxel rather than erlotinib. It is, therefore, unlikely that the same group of patients who would be eligible to receive erlotinib is the same as that who would be considered for docetaxel.

Box 3 Draft NICE guidance on the use of erlotinib as second-line treatment, 7th August 2014

- Erlotinib should not be recommended for treating locally advanced or metastatic NSCLC that has progressed after non-targeted chemotherapy in people with tumours that are EGFR-negative
- Erlotinib is recommended as an option for treating locally advanced or metastatic non-small-cell lung cancer that has progressed in people who have had non-targeted chemotherapy because of delayed confirmation that their tumour is epidermal growth factor receptor tyrosine kinase (EGFR-TK) mutation-positive, only if the manufacturer provides erlotinib with the discount agreed in the patient access scheme
- Erlotinib is recommended as an option for treating locally advanced or metastatic non-small-cell lung cancer that has progressed after chemotherapy in people with tumours of unknown epidermal growth factor receptor tyrosine kinase (EGFR-TK) mutation status, only if:
    - the result of an EGFR-TK mutation diagnostic test is unobtainable because of an inadequate tissue sample or poor quality DNA and
    - the treating clinician considers that the tumour is very likely to be EGFR-TK mutation-positive and
    - the person's disease responds to the first 2 cycles of treatment with erlotinib and
    - the manufacturer provides erlotinib with the discount agreed in the patient access scheme

According to the company: "Nintedanib fits well in the existing clinical pathway and can complement docetaxel treatment as an effective second-line option for patients with locally advanced/metastatic or recurrent NSCLC of adenocarcinoma tumour histology, previously treated with one line of chemotherapy." (page 34 of the CS[1]) As highlighted in the CS,[1] nintedanib is a potent, orally-administered small molecule triple angiokinase inhibitor targeting three receptor classes: vascular endothelial growth factor receptors (VEGFR), fibroblast growth factor receptors and platelet-derived growth factor receptors α and β.[22-24] These receptors have a key role in the formation and maintenance of new blood vessels (angiogenesis) and tumour growth.[25-27] Suppression of neo-angiogenesis via inhibition of VEGFR is considered a promising strategy for the treatment of human solid tumours, impacting tumour growth and spread.[25-27] The simultaneous targeting of all three pathways may be more effective than inhibition of angiogenesis via the VEGF pathway alone.

Largely based on the findings from the pivotal trial comparing nintedanib plus docetaxel to placebo plus docetaxel (LUME-Lung 1[24]), nintedanib is expected to be licensed in combination with docetaxel. Indeed, a positive opinion was received by the European Medicines Agency (EMA) on 25 September 2014 as follows: "Vargatef [nintedanib] is indicated in combination with docetaxel for the treatment of adult patients with locally advanced, metastatic or locally recurrent non-small cell lung cancer (NSCLC) of adenocarcinoma tumour histology after first-line chemotherapy."[28] As noted above, the group of patients who would be eligible to receive second-line docetaxel - and therefore nintedanib - is not likely to be the same as those who would be eligible to receive second-line erlotinib. Therefore the ERG considers with only very few exceptions, nintedanib plus docetaxel would fit into the existing treatment pathway as a comparator to docetaxel rather than erlotinib.

The ERG notes that the aforementioned positive opinion includes patients with locally recurrent NSCLC. In order to be classified as locally recurrent, a patient would initially present with early stage disease (stage I, II or IIIa). The company does not provide information on the service provision for these patients, presumably because the NICE scope is focussed on patients with locally advanced or metastatic disease. However, since the scope also focussed on second-line treatment following chemotherapy, the ERG considers these patients will have locally advanced or metastatic cancer by this stage. The ERG notes that patients with stage I, II or IIIa will initially be treated with surgery or radical radiotherapy and subsequently receive first-line chemotherapy when their disease has relapsed and/or spread.[29] The choice of chemotherapy will again depend on several factors, including, but not limited to, tumour histology and EGFR mutation status.

The estimated number of patients with locally advanced or metastatic adenocarcinoma potentially eligible for second-line treatment with nintedanib plus docetaxel in England is reported by the company to be 703. The ERG agrees with the company that a similar number of patients are likely to be eligible for treatment with nintedanib plus docetaxel. Based on data from the pivotal LUME-Lung 1[24] in which the median number of cycles with docetaxel was five (see also section 4.5) and given the norm in clinical practice in England is to provide a maximum of four cycles of docetaxel, the ERG considers the majority of patients would receive nintedanib in combination with four cycles of docetaxel.

3 CRITIQUE OF COMPANY'S DEFINITION OF DECISION PROBLEM

Table 1 displays the decision problem presented in the CS[1] and that addressed by the company. Each parameter is discussed in detail in the text following the table.

Table 1 Decision problem specified by NICE and addressed in the company's submission

Parameter	Final scope issued by NICE	Decision problem addressed in the company's submission
Population	Adults with locally advanced or metastatic non-small cell lung cancer (NSCLC) that has progressed following prior chemotherapy	Patients with locally advanced, metastatic or recurrent NSCLC of adenocarcinoma tumour histology after first-line chemotherapy
Intervention	Nintedanib in combination with docetaxel	As per final scope
Comparator(s)	Docetaxel monotherapy Erlotinib	Primary analysis: docetaxel monotherapy Secondary analysis: erlotinib monotherapy
Outcomes	The outcome measures to be considered include: • overall survival • progression-free survival • response rates • adverse effects of treatment • health-related quality of life	As per final scope
Economic analysis	The reference case stipulates that the cost effectiveness of treatments should be expressed in terms of incremental cost per quality-adjusted life year. The reference case stipulates that the time horizon for estimating clinical and cost effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared. The availability of any patient access schemes for the intervention or comparator technologies should be taken into account	As per final scope
Subgroups to be considered	None	Not applicable
Special considerations including equity or equality issues	None	Not applicable

Source: adapted from Table 5 of the CS[1]

3.1 Population

The population addressed in the CS[1] differs to the population specified in the scope. The scope states the population is adults with locally advanced or metastatic NSCLC that has progressed following prior chemotherapy. The decision problem addressed by the company is patients with locally advanced, metastatic or locally recurrent NSCLC of adenocarcinoma tumour histology who had previously received first-line chemotherapy. This is in line with the

anticipated full marketing authorisation for nintedanib which also specified nintedanib should be administered in combination with docetaxel (expected in December 2014). The ERG notes that to be classified as locally recurrent, a patient would initially present with early stage disease (stage I, II or IIIA) and be treated with surgery or radical radiotherapy and then relapse in the same area without metastases. Since the anticipated license also stipulates patients must have previously received first-line chemotherapy, then all patients would have locally advanced or metastatic disease at the time of second-line treatment regardless of their initial diagnosis. The ERG notes that while the scope makes no specification about the EGFR mutation status of tumours, in the UK the majority (85% to 90%) of patients have EGFR wild-type tumours (EGFR-negative).[30-32] The ERG further notes that as patients who receive nintedanib also receive docetaxel, the vast majority of eligible patients will be required to have ECOG PS 0 to 1.

3.2 Intervention

The intervention described in the CS[1] is nintedanib. Nintedanib does not currently have a full UK Marketing Authorisation. It does however have a positive opinion from the EMA and it is anticipated that it will be licensed in December 2014 in combination with docetaxel (the specified intervention in the final NICE scope). Nintedanib is provided orally at a dose of 200mg twice daily (BD) and dose adjustments are permitted in patients who experience AEs. The first dose reduction is to 150mg BD and, if required, the dose may be further reduced to 100mg BD. Docetaxel is administered intravenously alongside nintedanib on day 1 of a 21 day cycle at a dose of 75mg/m^2. If necessary, docetaxel doses may be reduced to 60mg/m^2 as per the docetaxel summary of product characteristics (SmPC)[33] and standard clinical practice. Nintedanib may be provided as monotherapy after discontinuation of docetaxel. In the pivotal LUME-Lung 1[24] trial, this was only permitted after four cycles of treatment with docetaxel. The ERG notes that in England, clinicians rarely administer more than four cycles of docetaxel due to the toxicity associated with this drug.

3.3 Comparators

Both docetaxel monotherapy and erlotinib monotherapy are considered as comparators for locally advanced or metastatic disease in the CS.[1] These are the same comparators that are specified in the scope. The company considers docetaxel monotherapy to be the comparator for the primary analysis and considers erlotinib to be the comparator for secondary analyses. This is because as stated on page 184 of the CS,[1] based on feedback from clinical experts, it does not believe that erlotinib is a relevant comparator. The ERG agrees with the company. As noted in section 2.2, the ERG notes that one of the NICE Appraisal Committee's preliminary recommendations[21] is that erlotinib should not be recommended for

treating locally advanced or metastatic NSCLC that has progressed after non-targeted chemotherapy in people with tumours that are EGFR-negative. Furthermore, the characteristics of patients who are considered suitable for second-line erlotinib treatment are different from those who are considered suitable for docetaxel treatment. Given that erlotinib is likely to be preferred when patients have a poorer ECOG PS and/or have EGFR-positive tumours, docetaxel is the most appropriate comparator to nintedanib plus docetaxel in the second-line setting. The company notes that no other agents are licenced or routinely used for this indication (pemetrexed is licensed but not NICE approved). Therefore, no other comparisons are presented (although as reported in section 4.4, there were other comparators employed in the MTCs). The ERG agrees that this is appropriate.

3.4 Outcomes

Clinical evidence is reported in the CS[1] for all outcomes specified in the scope: OS, PFS, response rate (reported as ORR]and disease control rate), AEs of treatment and HRQoL.

3.5 Economic analysis

Results are expressed in terms of incremental cost per QALY gained. Various time horizons are presented with lifetime (15 years) being that of the primary analysis (appropriate for a condition such as lung cancer, with low survival rates). Costs are considered from the perspective of the NHS. No patient access scheme has been submitted.

3.6 Subgroups

No subgroups were specified by NICE or identified by the company.

3.7 Other relevant factors

The company states on page 37 of the CS[1] that it does not consider there will be any equality issues if nintedanib is recommended by NICE.

4 CLINICAL EFFECTIVENESS

4.1 Introduction

This section provides a structured critique of the methods and clinical evidence submitted by Boehringer Ingelheim Ltd in support of the use of nintedanib in combination with docetaxel for the treatment of adult patients with locally advanced, metastatic or recurrent NSCLC of adenocarcinoma tumour histology after first-line chemotherapy. The key components of the clinical evidence presented in the CS[1] are (i) a report of the pivotal trial (LUME-Lung 1[24]) which compared nintedanib plus docetaxel to placebo plus docetaxel (ii) a report of the company's MTC which was conducted in order to compare nintedanib plus docetaxel to erlotinib.

4.2 Critique of the methods of review(s)

The company conducted a systematic literature review to identify RCTs of patients with previously treated second-line NSCLC. The review was designed to identify evidence for any drug, not limited to nintedanib plus docetaxel, erlotinib or docetaxel.

4.2.1 Searches

Sections 6.1.1 and Appendices 1 and 4 of the CS[1] describe the search strategies employed for the systematic review (direct evidence) and the multiple treatment comparison (MTC) (indirect evidence), respectively. While the ERG notes some potential minor limitations with the search strategy employed by the company (as outlined in Appendix 1), the ERG considers that the search strategies employed by the company were appropriate and sufficiently comprehensive to identify relevant studies.

In order to ascertain whether the company had missed any relevant studies or not, the ERG also conducted its own searches, as summarised in Appendix 1. However, the ERG did identify four additional conference presentations[34-37] for the pivotal LUME-Lung 1[24] trial not cited in the CS.[1]

4.2.2 Eligibility criteria

Although the same search strategy was employed to identify studies for inclusion in the systematic review (direct evidence) and MTC (indirect evidence), different eligibility criteria were appropriately employed for each. These are described in detail in Table 6 (pages 44 to 45) and Table 25 (pages 106 to 107) respectively of the CS[1] and summarised in Appendix 2. In general the ERG considers the criteria for both reviews were appropriate although notes that the eligibility of studies for inclusion into the MTC was limited to include only results with

abstracts, an unusual exclusion criterion which could potentially have removed relevant results. However, as noted in section 4.2.1, the ERG conducted its own searches and did not identify any additional eligible studies.

Although the same search was conducted to identify studies for both the systematic review and the MTC, it is unclear if the eligibility criteria for both reviews were simultaneously employed. The ERG notes from an examination of Figures 1 (page 46) and 19 (page 109) in the CS[1] that the number of records screened in the systematic review differed from the number screened in the MTC, suggesting this was not the case.

4.2.3 Quality assessment

The company conducted an assessment of the risk of bias of LUME-Lung 1,[24] the only study to meet the inclusion criteria for the systematic review, and all studies included in the MTC. This assessment included elements of the tool for assessing risk of bias, as recommended by the Cochrane Collaboration.[38] The ERG agrees this is an appropriate tool for assessing the quality of RCTs.

4.2.4 Evidence synthesis

One trial (LUME-Lung 1[24]) was identified by the searches for inclusion into the systematic review and hence the findings were appropriately presented narratively. This trial[24] compared nintedanib plus docetaxel to placebo plus docetaxel. In order to compare nintedanib plus docetaxel to erlotinib, the other comparator specified in the final NICE scope, the company conducted a MTC. The ERG's critique of the company's MTCs is presented in section 4.3.

4.3 Critique of the direct evidence

4.3.1 Identified studies

Only one RCT (LUME-Lung 1[24]) that presented direct evidence relevant to the decision problem was identified by the systematic review. The ERG is not aware of any additional relevant ongoing or completed studies. The company also referred to LUME-Lung 2[39] which compared nintedanib plus pemetrexed to placebo plus pemetrexed. However, data from LUME-Lung 2[39] were solely used to inform the pre-specified statistical analysis of LUME-Lung 1.[24]

As well as being published in a peer reviewed journal,[24] data from LUME-lung 1 were also provided by the company in two clinical trial reports (CTRs): primary PFS[40] and final OS[41] since analyses were conducted at both these time points (see section 4.3.4). Selected appendices to the CTR for final OS were also provided.[42] The company also provided the trial statistical analysis plan (TSAP),[43] the TSAP addendum[44] and the summary of clinical efficacy.[45] Three conference presentations were also cited, two poster presentations,[46,47] and an oral presentation, the slides of which were provided;[48] one of the poster presentations[46] also included data from LUME-Lung 2,[39] the focus of the presentation being to identify potential clinical biomarkers for second-line treatment. These findings are not presented by the company it the CS.[1] The ERG's search also identified four conference presentations not referred to by the company;[34-37] these do not appear to contain any additional data to that included in the CS.[1]

4.3.2 Trial characteristics

The key characteristics of LUME-Lung 1[24] are summarised in Table 2. The study was conducted internationally and randomised 1,314 patients in a 1:1 ratio to nintedanib plus docetaxel or placebo plus docetaxel. Randomisation was stratified by ECOG PS (0 vs 1), previous bevacizumab treatment (yes vs no), histology (squamous vs non-squamous) and presence of brain metastases (yes vs no). The ERG is of the opinion that the LUME-Lung 1[24] study is well-designed and conducted. A large number of patients were recruited to the study and the length of trial follow-up means that the data collected are mature and allow reasonable conclusions to be drawn from the data.

The ERG notes that some of the participating treatment centres were located in the UK although it is not known how many centres or numbers of patients were recruited (this was reported in Appendix 16.1.4 of the CTRs,[40,41] an appendix not included with the CS[1]). However, the ERG notes that the eligibility criteria for entry into this trial (see Appendix 3 for the full eligibility criteria as provided in the CS,[1] pages 58 to 59) do mean the patient

population was likely to be different to that of standard clinical practice in England in a number of different ways. Specifically the trial excludes, patients with clinically significant pleural effusion or evidence of cavitary or necrotic tumours and therapeutic anticoagulation (except low dose heparin) or antiplatelet therapy (except for chronic low-dose therapy with acetylsalicylic acid ≤325mg/day). In clinical practice these patients are likely to have a poorer prognosis than patients included in the trial[49-52] although it is recognised that cavitation may be less of a strong prognostic factor[49] than pleural effusions[50,51] or venous thromboembilsm.[52]

The ERG further notes that previous treatment with docetaxel is a specific exclusion criterion to entry in LUME-Lung 1.[24] Docetaxel is licensed for first-line treatment of NSCLC. However, this is rarely used in clinical practice in England, pemetrexed being the preferred choice for adenocarcinoma patients (see also section 2.2).

Table 2 Trial characteristics of LUME-Lung 1

Characteristics of LUME-Lung 1[24]	
Location	211 locations in 27 countries (Austria, Belarus, Belgium, Bulgaria, China, Croatia, Czech Republic, Denmark, France, Georgian Republic, Germany, Greece, India, Israel, Italy, Lithuania, Poland, Portugal, Romania, Russia, Slovakia, South Korea, South Africa, Spain, Switzerland, Ukraine, United Kingdom)
Design	Phase III multi-centre, randomised, parallel-group, double-blind, placebo-controlled RCT
Population	Patients with locally advanced, metastatic (stage IIIB/IV) or recurrent NSCLC after failure of first-line chemotherapy
Duration of study	23 December 2008 to 15 February 2013 (data cut-off date)
Intervention and comparator	Nintedanib + docetaxel (n=655) Nintedanib 200mg twice daily, orally, on days 2 to 21 of a 21-day cycle in combination with docetaxel 75mg/m^2 IV on day 1 of a 21-day cycle Matched placebo + docetaxel (n=659) Matched placebo twice daily on days 2 to 21 of a 21-day cycle in combination with docetaxel 75mg/m^2 IV on day 1 of a 21-day cycle
Primary outcomes	PFS by central independent review
Secondary outcomes	OS (key secondary endpoint) PFS by local investigator review Tumour response by central independent review and local investigator assessment, including: confirmed objective response; disease control; time to confirmed objective response; duration of confirmed objective response; duration of disease control; change in tumour size; clinical improvement HRQoL Pharmacokinetics safety and tolerability
Duration of follow-up	Median follow-up at the primary PFS analysis (2 November 2010) was 7.1 months (interquartile range: 3.8 to 11.0) and 31.7 months (interquartile range: 27.8 to 36.1 months) at the time of the final OS analysis (15 February 2013)

AE=adverse event; HRQoL=health related quality of life; NSCLC=non-small-cell lung cancer; OS=overall survival; PFS=progression-free survival; RCT=randomised controlled trial
Source: adapted from Table 8 of the CS[1]

4.3.3 Participant characteristics

Not all patients in LUME-Lung 1[24] had histology of adenocarcinoma. As the expected marketing authorisation for nintedanib plus docetaxel is specifically for patients with adenocarcinoma, the company only presented data for the overall population of patients with NSCLC where the results were of relevance to statistical testing (see section 4.3.4). The ERG agrees that this is appropriate. The participant characteristics of 658 (50.1%) patients with adenocarcinoma in LUME-Lung 1[24] are summarised in Table 3. While some patients (15.8%) had early stage disease at diagnosis, at the time of treatment the ERG considers all would have locally advanced or metastatic disease since patients had all received first-line treatment and were now being treated second-line. Indeed, 94.2% of all patients had metastatic disease at screening. The mean time from diagnosis to randomisation into the trial reported in Table 15.1.8: 3 of the CTR[41] was 12.84 months (median 8.74 months).

The company comments that demographic and baseline disease characteristics are well balanced between the two arms of the trial, and that the population is largely representative of patients typically diagnosed with adenocarcinoma although it is noted by the ERG that the proportion of patients aged ≥65 years is relatively small (28.3%). The ERG agrees that the patient characteristics are well balanced.

Data on EGFR mutation status was not routinely collected in LUME-Lung 1[24] although in response to a query from the ERG during the clarification process, the company stated these data has been retrospectively collected for a sample of patients: [redacted]

Notwithstanding the exclusions of certain types of patients identified in section 4.3.2, the patient population is similar to the population who would be treated in clinical practice in England with the exception that a smaller proportion (18.8%) of patients than would be expected today had received prior pemetrexed. Additionally, perhaps as a result of the eligibility criteria, it is noted that post-study therapy is relatively high (55.8%) which suggests this is a fitter patient population than in clinical practice England in Wales. As noted in section 2.2, in England, in clinical practice around [redacted] of all patients who receive second-line treatment subsequently receive third-line treatment (including a third of patients who are enrolled into trials).

Table 3 Participant characteristics of patients with adenocarcinoma in LUME-Lung 1

Characteristic		Nintedanib + docetaxel (N=322)	Placebo + docetaxel (N=336)
Sex, n (%)	Male Female	203 (63.0) 119 (37.0)	208 (61.9) 128 (38.1)
Age, years	Mean (SD) Median (range)	58.5 (10.1) 60.0 (29 to 80)	58.6 (9.5) 59.0 (30 to 80)
Age ≥65 years, n(%)		90 (28.0)	96 (28.6)
Race, n (%)	Asian White Other	65 (20.2) 253 (78.6) 4 (1.2)	78 (23.2) 253 (75.3) 5 (1.5)
ECOG performance status, n (%)	0 1†	96 (29.8) 226 (70.2)	99 (29.5) 237 (70.5)
Stage of disease at diagnosis, n(%)	<IIIB IIIB IV	50 (15.6) 55 (17.2) 215 (67.2)	54 (16.1) 45 (13.4) 237 (70.5)
Local recurrence without metastases at screening		22 (6.8)	16 (4.8)
Smoking status, n (%)	Never smoked Ex-smoker Current smoker	115 (35.7) 151 (46.9) 56 (17.4)	115 (34.2) 162 (48.2) 59 (17.6)
Prior first-line therapy	Platinum-based therapy	308 (95.7)	323 (96.1)
	Non-platinum-based therapy	10 (3.1)	10 (3.0)
Prior pemetrexed, n (%)	As platinum therapy As non-platinum therapy	58 (18.0) 3 (0.9)	61 (18.2) 2 (0.6)
Prior bevacizumab, n (%)		24 (7.5)	21 (6.3)
Brain metastases at study entry, n (%)	Present Absent	26 (8.1) 296 (91.9)	23 (6.8) 313 (93.2)
Post study therapy	Any systemic therapy Any chemotherapy Pemetrexed Docetaxel Other chemotherapy EGFR-TK inhibitor Anti-angiogenesis agent Investigational agent	179 (55.6) 123 (38.2) 52 (16.1) 15 (4.7) 90 (28.0) 98 (30.4) 6 (1.9) 18 (5.6)	188 (56.0) 136 (40.5) 62 (18.5) 13 (3.9) 101 (30.1) 105 (31.3) 2 (0.6) 5 (1.5)

ECOG=Eastern Cooperative Oncology Group; EGFR-TK=epidermal growth factor receptor tyrosine kinase; SD=standard deviation
† Including one patient in the nintedanib arm who had an ECOG PS of 2 at screening and at randomisation (i.e. at baseline)
Source: adapted from Table 10 of the CS[1] with additional information taken from Table 15.1.8:2 of the CTR[41]

4.3.4 Description and critique of the statistical approach

Information relevant to the statistical approach taken by the company to analyse data from the pivotal study LUME-Lung 1[24] are taken from the TSAP,[43] trial protocol,[53] CTRs[40,41] and the CS.[1]

Sample size calculation

Details of the sample size calculation performed by the company are reported in the CS[1] (page 71). The study was powered (at the 90% level) to detect a HR for centrally independently assessed PFS for the comparison of nintedanib plus docetaxel vs placebo

plus docetaxel of 0.7843. This would require 713 PFS events. The ERG is satisfied that the company's pre-specified sample size calculation is correct. However as noted in section 3.1 above, only patients with adenocarcinoma were considered relevant to this STA. The company therefore only presents data for the adenocarcinoma population. The ERG notes that although around half of the patients in LUME-Lung 1[24] had adenocarcinoma (see section 4.3.3) this was not a stratification factor (see section 4.3.2) and so patients with adenocarcinoma were not strictly a randomised subgroup although they do constitute the majority of non-squamous patients which was a stratification factor. However, as noted in section 4.3.3, baseline characteristics were well balanced between the two groups suggesting the analyses were valid.

Protocol amendments

A list of changes implemented after a protocol amendment (dated 15 May 2009) is included in the CTR[40] (pages 120 to 121). The changes included slight adjustments to the exclusion criteria, clarification of ongoing safety evaluations, and timings of the screening period. All changes were made before analyses began, and so were not driven by the results of the trial. The ERG considers that it is very unlikely that any of the changes would influence the outcomes or analyses of LUME-Lung 1,[24] or would be a cause for concern.

Clinical endpoints and statistical analyses

The company provides a list of outcome measures used in LUME-Lung 1[24] in Table 13 (page 66) of the CS[1] (also summarised in Appendix 4 of the ERG report). The ERG is satisfied that all outcomes were pre-specified in the TSAP[43] and reported in full in the CTRs.[40,41]

The intention-to-treat (ITT) population was used in all efficacy analyses. The primary outcome of PFS by central review was analysed using the K-M method, and a stratified log-rank test. Cox regression analyses were also carried out to estimate treatment effect, including adjustment for stratification factors.

Secondary outcomes relevant to the decision problem included OS, PFS by local investigator review, best tumour response, HRQoL and AEs. OS, the key secondary outcome of the trial, was also analysed using a stratified log-rank test. Tumour response was reported for both central independent review and local investigator review according to modified Response Evaluation in Solid Tumours (RECIST) criteria and analysed using logistic regression.

Within clinical trials, time-to-event data like PFS and OS are commonly reported as HRs, derived from the Cox proportional hazards model. Such a model does not appear to be appropriate for the PFS and OS results of this trial since hazards are not independent of time (see Appendix 7) and the HR (and 95% CIs) presented for PFS and OS offer inaccurate estimates of relative efficacy. Instead of assuming proportional hazards, alternative approaches may be more appropriate to better reflect relative efficacy in the data..

The CS[1] (page 71) describes the stages of analyses in Table 14. These are summarised in Appendix 4 of the ERG report. The ERG is satisfied that each of these stages was pre-specified in the trial protocol.[53]

Following the hypothesis-generating trial LUME-Lung 2,[39] an amendment to the statistical plan of LUME-Lung 1[24] was implemented such that statistical testing of OS would only be conducted if a significant difference had been observed for the primary analysis of PFS and had been confirmed by the updated analysis of PFS. If this condition was satisfied, OS analyses would then be conducted in a sequential fashion, i.e. the null hypothesis was to be tested in each population only if a significant treatment effect had been shown in the previous population. This hierarchical method was utilised to control the type 1 error rate (detecting an effect when one is not present), which can be high when performing a large number of statistical tests. The sequence of populations was:

1. Adenocarcinoma patients who had progressed within 9 months of starting first-line therapy (i.e. the T<9m adenocarcinoma population)

2. Adenocarcinoma population

3. Overall trial population

The CS[1] clarifies that the amendment to the TSAP[43] was made before database lock and unblinding of data used in the final OS analysis; the ERG considers that this amendment is unlikely to bias the results from LUME-Lung 1.[24]

Subgroup analyses

A number of pre-specified analyses for the primary endpoint of PFS assessed by central review and for the secondary outcome OS were pre-specified in the protocol. The company also conducted post-hoc subgroup analyses and a number of baseline characteristics (CS,[1] page 77) were also investigated for subgroup effects. The subgroup types analysed are summarised in Appendix 5 of the ERG report. The ERG notes that there is a large number of

subgroup analyses but is satisfied that the results of all of the pre-specified and post-hoc subgroup analyses are provided in the CS.[1]

Sensitivity analyses

A number of sensitivity analyses for the primary endpoint of PFS assessed by central review and for the secondary outcome OS were pre-specified in the protocol.[53] These are summarised in Appendix 5 of the ERG report. However, no sensitivity analysis of PFS in the adenocarcinoma population was performed. The ERG is satisfied that the results of all of the pre-specified and post-hoc sensitivity analyses are provided in the CTR.[41]

4.3.5 Risk of bias

The company conducted an assessment of the risk of bias using the criteria recommended by NICE in the Guide to the Methods of Technology Appraisal.[54] The risk of bias assessment is presented in Table 4. The ERG is satisfied with the risk of bias assessment presented in the CS[1] and agrees that the study has an overall low risk of bias.

Table 4 Assessment of risk of bias conducted by company for LUME-Lung 1trial

Criteria	Response
Was randomisation carried out appropriately?	Yes
Was the concealment of treatment allocation adequate?	Yes
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes
Were the care providers, participants and outcome assessors blind to treatment allocation?	Yes
Were there any unexpected imbalances in drop-outs between groups?	No
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Yes

Source: Table 17 of the CS[1]

4.3.6 Results

The focus of the results section in both the CS[1] and this ERG report is the adenocarcinoma population from LUME-Lung 1,[24] as this is the population relevant to the decision problem.

However, results of the primary PFS analysis for the overall trial population and OS for the T<9m adenocarcinoma population have been presented wherever necessary (and are clearly labelled), due to the fact that these populations were part of the previously described hierarchical OS statistical analysis (section 4.3.4). By presenting these results, the justification for conducting the analysis of OS for the patients with adenocarcinoma has been demonstrated.

Progression-free survival

The CS[1] reports that median follow-up was 7.1 months at the time at the primary PFS analysis (2 November 2010), the results of which are presented in Table 5.

Table 5 Primary analysis of centrally independently assessed PFS in LUME-Lung 1 trial (November 2010)

Outcome	Nintedanib + docetaxel (median)*	Placebo + docetaxel (median)*	HR vs placebo arm (95% CI)†	p-value	Risk reduction
PFS in overall ITT population	3.4 months	2.7 months	0.79 (0.68 to 0.92)	0.0019	21%
PFS in adenocarcinoma population§	4.0 months	2.8 months	0.77 (0.62 to 0.96)	0.0193	23%

CI=confidence interval; HR=hazard ratio; ITT=intention-to-treat; PFS=progression-free survival
* Based on unadjusted Kaplan-Meier estimates for each treatment arm
† A proportional hazards model stratified by three factors (ECOG PS at baseline, presence of brain metastases at baseline, prior bevacizumab therapy) was used to derive the HR, 95% CI, and p-value (corresponding to the stratified log-rank test p-value)
§ Analysis conducted retrospectively
Source: Table 18 of the CS[1]

The results suggest that the use of nintedanib plus docetaxel significantly improved PFS in comparison to placebo plus docetaxel in both the overall trial population and in the subgroup of patients with adenocarcinoma. However, the ERG suggests that these results should be interpreted with caution, due to the violation of the proportional hazards assumption (see section 4.3.4). In particular, it is evident that the trial survival arms converge within the duration of the trial indicating that PFS gain from use of nintedanib is restricted to the first 12 months of treatment (see Figure 12 in section 5.5.3) and that the hazard ratio is not time-invariant (Figure 19 in Appendix 7).

The CS[1] reports that median follow-up was 31.7 months at the time of the updated PFS analysis (15 February 2013), the results of which are summarised in Table 6.

Table 6 Updated analysis of centrally independently assessed PFS in in LUME-Lung 1 trial (February 2013)

Outcome	Nintedanib + docetaxel (median)*	Placebo + docetaxel (median)*	HR vs placebo arm (95% CI)†	p-value	Risk reduction
PFS in the overall trial population	3.5 months	2.7 months	0.85 (0.75 to 0.96)	0.0070	15%
PFS in adenocarcinoma population§	4.2 months	2.8 months	0.84 (0.71 to 1.00)	0.0485	16%

CI=confidence interval; HR=hazard ratio; PFS=progression-free survival
* Based on unadjusted Kaplan-Meier estimates for each treatment arm
† A proportional hazards model stratified by three factors (ECOG PS at baseline, presence of brain metastases at baseline, prior bevacizumab therapy) was used to derive the HR, 95% CI, and p-value (corresponding to the stratified log-rank test p-value)
§ Analysis conducted retrospectively
Source: Table 19 (page 89) of the CS[1]

The CS[1] states that the results obtained in the updated analysis support the findings from the primary PFS analysis. The ERG agrees that the results are consistent across both analyses as nintedanib plus docetaxel is shown to significantly improve PFS in comparison to placebo plus docetaxel in both the overall trial population and the adenocarcinoma population at the updated analysis.

Progression-free survival by local investigator review

The ERG notes that the PFS results as assessed by local investigator review were very similar to those obtained by central review. The treatment effect for the adenocarcinoma population significantly favoured nintedanib plus docetaxel over placebo plus docetaxel (HR 0.78, 95% CI: 0.62 to 0.97, p=0.0246).

Progression-free survival subgroup analyses

Subgroup analyses were performed at the time of the final OS analysis (15 February 2013). Results from the PFS (central review) subgroup analyses of baseline characteristics for adenocarcinoma patients are provided by the company in Figure 17 of the CS[1] (page101). The majority of pre-specified and post-hoc subgroup analyses show the effect of nintedanib plus docetaxel to be consistent with the treatment benefit observed in the primary analysis. The only exceptions to this are two subgroups (i) more than 9 months since start of first-line treatment and (ii) Asian region where there was a trend in favour of placebo plus docetaxel.

The results of tests for interaction were also provided to identify whether any subgroup of patients experienced a significantly greater treatment benefit than the remaining population. Significant interactions were observed for 'time since start of first-line therapy' (p=0.0032) and metastases in 'adrenal glands' (p=0.0336); these results suggest that patients who progressed within 9 months of first-line therapy, and those with metastases in the adrenal glands, experience a greater treatment effect than the remaining population.

Progression-free survival sensitivity analyses

Sensitivity analyses were only performed for PFS in the whole trial population, not only for those with adenocarcinoma.

Overall survival

Nintedanib plus docetaxel significantly improved median OS in comparison to placebo plus docetaxel in the population of adenocarcinoma patients who progressed within 9 months of first-line therapy (10.9 months vs 7.9 months respectively; HR 0.75, 95% CI: 0.60 to 0.92, p=0.0073). Therefore, analysis of OS in the population of interest, all adenocarcinoma patients, was permitted and the results are summarised in Table 7. Median OS was significantly longer with nintedanib plus docetaxel than with placebo plus docetaxel in the

adenocarcinoma population. However, the ERG is concerned that survival hazards appear not to be time invariant (see Figure 20, Appendix 7) and therefore may be misleading.

Table 7 OS in the adenocarcinoma population in LUME-Lung 1 trial (February 2013)

Outcome	Nintedanib + docetaxel (median)*	Placebo + docetaxel (median)*	HR vs placebo arm (95% CI)†	p-value
Overall survival	12.6 months	10.3 months	0.83 (0.70 to 0.99)	0.0359

CI=confidence interval; HR=hazard ratio; OS=overall survival
* Based on unadjusted Kaplan-Meier estimates for each treatment arm
† A proportional hazards model stratified by three factors (ECOG PS at baseline, presence of brain metastases at baseline, prior bevacizumab therapy) was used to derive the HR, 95% CI, and p-value (corresponding to the stratified log-rank test p-value)
Source: Table 20 of the CS[1]

Overall survival subgroup analyses

Subgroup analyses were performed at the time of the final OS analysis (15 February 2013) in the adenocarcinoma population.

Results from the pre-specified and post-hoc OS subgroup analyses of baseline characteristics for adenocarcinoma patients are provided by the company in Figure 18 of the CS[1] (page102). The subgroup analyses also show treatment effects in favour of nintedanib plus docetaxel, supporting the findings of the primary analysis. The only exceptions to this are two baseline characteristics: (i) presence of brain metastases and (ii) below stage IIIB disease at diagnosis. The company notes that a significant interaction was observed for 'best response to first-line treatment' (p=0.0766), indicating that patients whose best response to first-line therapy was PD would benefit more in terms of OS than the rest of the population. The ERG agrees with the company that this subgroup has a relatively small sample size (n=117) and the results should therefore be interpreted with caution.

The ERG is satisfied that all pre-specified subgroups were reported and show a consistent effect for OS across the majority of baseline characteristics.

Overall survival sensitivity analyses

The results of the two sensitivity analyses performed for OS in the adenocarcinoma population are presented in the text of the CS[1] and summarised here in Table 8.

Table 8 Sensitivity analyses of OS in the adenocarcinoma population in LUME-Lung 1 (February 2013)

Analysis	HR (95% CI)	p-value
Main OS analysis	0.83 (0.70 to 0.99)	0.0359
Sensitivity analysis 1 - Cox proportional hazards model with three of the stratification factors used at randomisation as covariates (ECOG PS at baseline, prior bevacizumab treatment, presence of brain metastases at baseline)	0.83 (0.70 to 0.98)	0.0295
Sensitivity analysis 2 - Model included the stratification factors and the baseline sum of the longest diameters (SLD) of the target lesions (mm) as covariates	0.81 (0.69 to 0.97)	0.0186

CI=confidence interval; HR=hazard ratio; OS=overall survival
Source: Text (page 99) of the CS[1] and Table 11.4.1.2.1.7: 2 of CTR[41]

The sensitivity analyses show that the results of the OS analysis remain very similar when including three of the stratification factors (ECOG PS at baseline, prior bevacizumab treatment and presence of brain metastases at baseline), or the stratification factors and baseline sum of the longest diameters of the target lesions as covariates in the model..

Tumour response based on central independent review

The results from the tumour response assessment (central independent review) are summarised in Table 9. No significant difference in ORR between nintedanib plus docetaxel patients and placebo plus docetaxel patients (4.7% vs 3.6%, odds ratio 1.32 [95% CI 0.61 to 2.93], p=0.4770) was observed. The ERG considers the ORRs in both arms to be lower than would be anticipated in typical clinical trials (see also section 4.4.5).

Table 9 Tumour response and disease control in the adenocarcinoma population in LUME-Lung 1 (February 2013)

Type of response (according to modified RECIST version 1.0 by central independent review)	Nintedanib + docetaxel (n=322)	Placebo + docetaxel (n=336)	Odds ratio* (95% CI)
Patients with objective tumour response, ORR [n (%)]	15 (4.7)	12 (3.6)	1.32 (0.61 to 2.93) p=0.4770
Complete response, n (%)	0	0	-
Partial response, n (%)	15 (4.7)	12 (3.6)	-
Unconfirmed complete/partial response n (%)	10 (3.1)	7 (2.1)	-
Median duration of confirmed objective response (months)	4.9	4.3	-
Median time to confirmed objective response (months)	1.6	5.1	-
Stable disease† n (%)	179 (55.6)	136 (40.5)	-
Patients with disease control§ n (%)	194 (60.2)	148 (44.0)	1.93 (1.42 to 2.64) p<0.0001
Median duration of disease control (months)	5.7	6.3	-
Progressive disease‡ n (%)	87 (27.0)	147 (43.8)	-
Other¥ n (%)	41 (12.7)	41 (12.2)	-

CI=confidence interval; ORR=overall response rate
* Odds ratios were obtained from logistic regression model adjusted for baseline ECOG PS
† stable disease was assumed if a follow-up imaging indicated stable disease at least once and at least 6 weeks after randomisation (i.e. at or after Day 43).
§ A patient was considered to have disease control if he/she had a best objective response of stable disease or better.
‡ Including patients with stable disease from a radiological imaging earlier than Day 43 followed by progressive disease
¥ Including patients with stable disease from a radiological imaging earlier than Day 43 followed by a non-evaluable response
Source: Table 21 of the CS[1]

4.4 Critique of the indirect evidence

4.4.1 Included studies in the MTC and statistical approach

Nine trials[24,55-62] were included in the review of the indirect evidence. The ERG did not identify any additional studies that met the company's eligibility criteria. However, not all nine studies were incorporated in any single MTC analysis. Four studies[48,56,59,62] were included in the base-case analyses, three[24,59,62] of which were also included in scenario analyses alongside a fifth study.[60] The remaining four studies[55,57,58,61] were only included in sensitivity analyses alongside those included in the base-case (sensitivity analyses i) or scenario analyses (sensitivity analyses ii); hence there were only ever a maximum of eight studies included in any given analysis.

The features of the types of analyses are as follows:

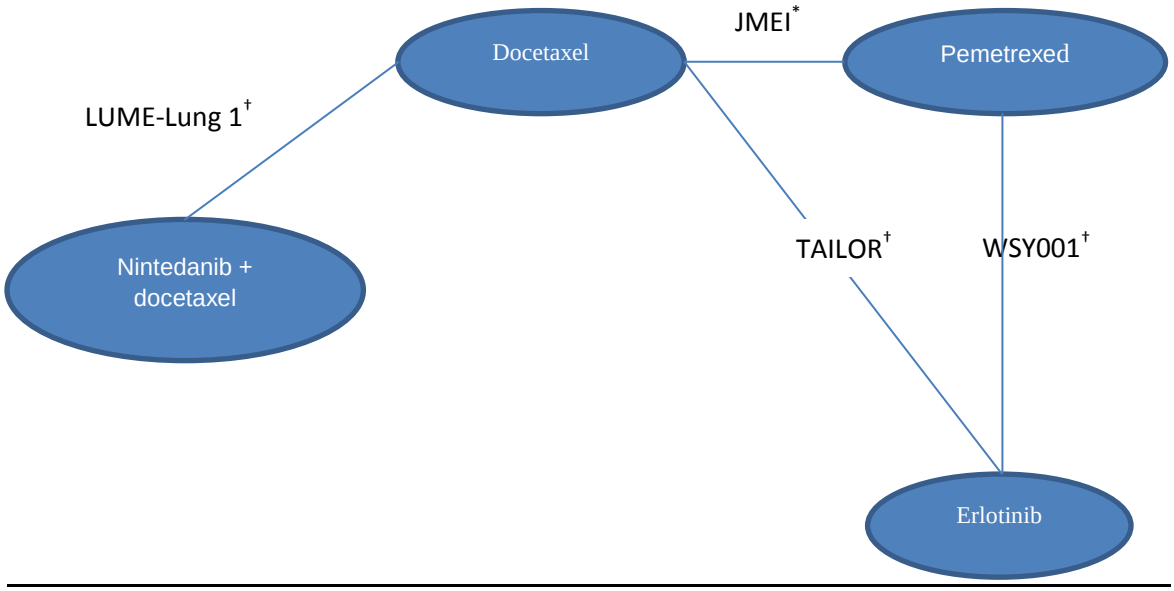
1. Base-case: includes all trials that meet eligibility criteria but excludes studies in which a high proportion (>20%) of patients have EGFR-positive adenocarcinoma and studies which include 'chemotherapy' as a single comparator where chemotherapy could be one or more possible regimens i.e. it must be possible to compare the intervention to all included comparators separately. The base-case analysis network diagram is reported in Figure 1.

2. Scenario analysis: assumes docetaxel and pemetrexed are of equal efficacy. The CS[1] states that this assumption was used to allow as many treatments to be compared with nintedanib plus docetaxel as possible. Hence the TITAN[60] study, excluded from the base-case, could be included in the scenario analysis because chemotherapy (docetaxel or pemetrexed) was the comparator. However the JMEI[56] study could not be included since this trial compared docetaxel to pemetrexed.

3. Sensitivity analyses: studies in which >20% of patients had EGFR-positive adenocarcinoma were also included in a MTC alongside

    i. the trials included in the base-case or

    ii. the trials included in the scenario analyses.



* Trial included only patients with adenocarcinoma
† Subgroup of patients with adenocarcinoma

Figure 1. Network diagram for MTC base-case analyses
Source: adapted from Figure 20 of the CS[1]

The company explains that the rationale for excluding patients with EGFR-positive adenocarcinoma from all but the sensitivity analyses was to enable a comparison between nintedanib plus docetaxel and other TKIs in a population similar to the patient population in LUME-Lung 1.[24] The majority of patients in LUME-Lung 1[24] would be expected to have EGFR-mutation negative adenocarcinoma [redacted] (see also section 4.3.3 and Table 14, section 4.4.3).

For each analysis, the company attempted to compare efficacy and safety. Efficacy outcomes were OS, PFS and ORR and safety outcomes were AEs for the following: fatigue, nausea and diarrhoea. However, for AEs, it was not possible to conduct a MTC for the base-case because none of the AE outcomes were reported in a sufficient number of trials in the base-case in order to be able to conduct a MTC.

The studies, comparators and analyses are summarised in Table 10 and Table 11.

Table 10 Studies included in the review of indirect evidence identified by the company

Trial name	Intervention	Comparator	Analyses included in
LUME-Lung 1[24]	Nintedanib + docetaxel	Placebo + docetaxel	Base-case, scenario and sensitivity
TAILOR[59]	Erlotinib	Docetaxel	Base-case, scenario and sensitivity
WSY001[62]	Erlotinib	Pemetrexed	Base-case, scenario and sensitivity
JMEI[56]	Pemetrexed	Docetaxel	Base-case and sensitivity
TITAN[60]	Erlotinib	Chemotherapy (docetaxel or pemetrexed)	Scenario and sensitivity
GEF-ERL[55]	Gefitinib	Erlotinib	Sensitivity
KCSG-LU08-01[57]	Gefitinib	Pemetrexed	Sensitivity
V-15-32[61]	Gefitinib	Docetaxel	Sensitivity
S103[58]	Pemetrexed + erlotinib	Pemetrexed or erlotinib	Sensitivity

Source: adapted from Figure 20 and Figure 21 of CS[1]

Table 11 Comparisons with nintedanib plus docetaxel in the MTCs undertaken by the company

Analyses	Comparators	Outcomes
Base-case	Docetaxel Erlotinib Pemetrexed	Overall survival Progression-free survival Overall response rate
Scenario	Chemotherapy (docetaxel and/or pemetrexed) Erlotinib	Overall survival Progression-free survival Overall response rate Safety
Sensitivity i	Docetaxel Pemetrexed Erlotinib Gefitinib Pemetrexed + erlotinib	Overall survival Progression-free survival Overall response rate
Sensitivity ii	Chemotherapy (docetaxel and/or pemetrexed) Erlotinib Gefitinib Pemetrexed + erlotinib	Overall survival Progression-free survival Overall response rate Safety

Sensitivity i: sensitivity analyses for base-case; sensitivity II; sensitivity analyses for scenario analyses
Source: adapted from Table 36 and Table 37 of CS[1]

For efficacy and safety outcomes, the company conducted MTCs and, where possible, Bucher indirect comparison results using the methods described in Appendix 6. The ERG is satisfied that the modelling approach was suitable. The ERG considers that conducting Bucher indirect comparisons is an effective method of assessing consistency within the network and therefore the reliability of the MTC results. If results from the MTC for any given comparison are considerably different to those obtained by the Bucher indirect comparison, it is likely that the MTC is not measuring the treatment effect accurately.

However, the ERG does not consider conducting any MTC was appropriate. There are multiple reasons for this, the primary reasons relating to the appropriateness of the MTC to the decision problem:

1. Erlotinib is not an appropriate comparator for the population of patients who would potentially be eligible to receive nintedanib plus docetaxel. As noted earlier in sections 0 and 3.3, the NICE Appraisal Committee's preliminary recommendations[21] are that erlotinib should not be recommended for treating locally advanced or metastatic NSCLC that has progressed after non-targeted chemotherapy in people with tumours that are EGFR-negative. Furthermore the characteristics of the vast majority of patients who are considered suitable for second-line erlotinib treatment are different from those who are considered suitable for second-line docetaxel treatment, most notably in terms of ECOG PS, EGFR mutation status and previous treatment received. Therefore the comparison with docetaxel is most appropriate and direct evidence for this is available from LUME-Lung 1.[24]

2. The ERG further observes that LUME-Lung 1[24] is the only trial in which any patients received pemetrexed as a first-line treatment as is now typically the case in clinical practice in England. In this trial, 19.1% of patients were previously treated with pemetrexed, more than in any other included trial.

In addition, there are methodological issues:

3. Although this is still an issue of some academic debate, the ERG considers that the proportional hazards assumption is not supported by LUME-Lung 1[24] trial data for PFS or OS. As LUME-Lung 1[24] is the only trial providing evidence for nintedanib plus docetaxel, any comparison also including evidence from this trial will incorporate this HR, which affects the robustness of these other comparisons. Thus any estimation of the relative effectiveness of nintedanib plus docetaxel vs erlotinib (i.e. a calculated HR) will lack credibility and be effectively meaningless. A full assessment of this issue is provided by the ERG in Appendix 7.

4. Differences in trial and patient characteristics (as described in detail in sections 4.3.2 and 4.3.3) suggest there is heterogeneity across trials which may mean MTCs are inappropriate:

    a. In the base-case analyses, while LUME-Lung 1[24] and TAILOR[59] both report similar median follow-up times, JMEI[56] and WSY001[62] report much shorter follow-up times. This heterogeneity may mean that the trials are too dissimilar

to allow a valid comparison of outcomes in an MTC. Additional sources of heterogeneity have also been identified in terms of differences in eligibility criteria across trials (see section 4.4.2) and participant characteristics (see section 4.3.3).

b. There also appears to be heterogeneity across the trials included in the scenario analyses.[24,59,60,62] TITAN[60] includes many more patients with ECOG PS 2 (20%) than would be expected in a patient population considered for treatment with nintedanib plus docetaxel. Furthermore, unlike any of the other trials in the base-case, TITAN[60] also permitted treatment crossover following disease progression. Hence the median OS for the chemotherapy arm may be inflated. It also compares erlotinib to chemotherapy in which chemotherapy consists of docetaxel or pemetrexed, thereby assuming the two treatments to be of equal efficacy. The ERG is not aware of any evidence that supports this assumption specifically in an adenocarcinoma population. Finally, by including this trial, the MTC is no longer making comparisons to docetaxel but to chemotherapy. However the chemotherapy arm includes pemetrexed which is not a second-line treatment option in England. Taking these factors into account, the ERG considers that the efficacy and safety results generated by the scenario analyses are neither relevant nor robust.

c. Trials included only in the sensitivity analyses[55,57,58,61] appear to be different to those in the base-case and scenario analyses, in particular these trials have high proportions of patients with EGFR-positive mutations and are based in Asia. Combining data from these trials with data from trials in the base-case and scenario analyses appears to be inappropriate as patients from Asia may have different tumour biology and comorbidities to those in the UK and EGFR mutation status is known to be related to the efficacy of some drugs. The ERG considers that the efficacy and safety results generated by the sensitivity analyses are not robust.

d. For the MTCs of safety outcomes, the company explains that due to low event rates, and the fact that only a small number of trials reported these outcomes, a network could only be formed when assuming equal tolerability of docetaxel and pemetrexed (scenario analysis). However, the findings from JMEI,[56] which compared these two drugs, albeit in a broader NSCLC population (52.9% had adenocarcinoma), reported differences between the two drugs, with a more favourable safety profile for pemetrexed. Therefore

this assumption does not hold. Furthermore, as identified above, the ERG considers there are differences in trial and patient characteristics between the trials included in the base-case, scenario and sensitivity analyses. The ERG considers that none of safety results generated by the MTC analyses are robust.

For all of the reasons outlined above, the ERG does not consider the comparison of nintedanib plus docetaxel with erlotinib is relevant to this STA.

4.4.2 Trial characteristics of included studies

The characteristics of trials included in the base-case and scenario analyses are summarised in Table 12 as well as the characteristics of those trials included in the sensitivity analyses only. The ERG notes that only TAILOR[59] was conducted solely in Europe whereas four trials[55,57,61,62] were conducted solely in Asia; three[55,57,61] of the Asian studies were included only in the sensitivity analyses. The location of trials is likely to be important because patients may have different tumour biology and comorbidities depending on their ethnic origin and where they live.

The company argues that all of the included trials had similar eligibility criteria. However, the ERG notes that there were some differences.

The ERG considers that two eligibility criteria (ECOG PS and complications such as brain metastases and pleural effusions) may be the main drivers of outcome in patients with adenocarcinoma. In the base-case and scenario analyses, only LUME-Lung 1[24] restricted trial entry to ECOG PS ≤1. Six trials explicitly stated they excluded patients with brain metastases: three in the base-case: LUME-Lung 1,[24] WSY001[62] and JMEI;[56] TITAN[60] in the scenario analyses and KCSG-LU08-01[57] and S103[58] in the sensitivity analyses. It is however noted by the ERG that LUME-Lung 1[24] excluded patients with active brain metastases and so this exclusion criterion may have enabled patients with a poorer prognosis to have been included than in the other trials in this respect. Two trials (LUME-Lung 1[24] and JMEI[56]), both in the base-case, excluded clinically significant or uncontrolled pleural effusions. Therefore patients in LUME-Lung 1[24] and JMEI[56] in particular may be expected to have slightly better prognoses than patients in the other trials although similar exclusion criteria may have been employed in the other trials but were not reported; for example TAILOR,[59] GEF-ERL[55] and V-15-32[61] reported only limited eligibility criteria. The ERG acknowledges that existence of brain metastases is a relatively common exclusion criteria for entry into trials. Nevertheless, such exclusions do result in a patient population different to those who would be treated in clinical practice. In this instance, because patients who receive nintedanib do so in

combination with docetaxel, the exclusion of patients with ECOG PS≥2 is however appropriate.

Additional eligibility criteria which could also impact on patient outcomes include EGFR mutation status, previous treatment and smoking status. WSY001[62] included only patients with EGFR wild type disease (EGFR-negative) whereas GEF-ERL[55] included only patients with EGFR activating mutations (EGFR-positive). Patients in the latter study would be expected to perform better when treated with a TKI or chemotherapy than patients in the former study. Furthermore, it should be noted that the majority of patients treated in clinical practice in England would be EGFR-negative. KCSG-LU08-01[57] and V-15-32[61] only permitted entry to never-smokers whereas the majority of patients with NSCLC treated in England are current or ex-smokers. Prior pemetrexed (or drugs directed at pemetrexed molecular targets) or TKIs were explicitly not permitted in three trials in the base-case (WSY001[62] and JMEI[56] and S103[58]), TITAN[60] in the scenario analyses and three trials (TAILOR,[59] GEF-ERL[55] and KCSG-LU08-01[57]) in the sensitivity analyses. These are potentially important exclusion criteria as not only may these affect outcomes but in clinical practice in England today, as noted in section 2.2, these are the first-line treatments of choice: pemetrexed for EGFR-negative disease and a TKI for EGFR-positive disease.

Alongside differences in eligibility criteria, the ERG also observes that in V-15-32,[61] docetaxel was administered every 3 weeks as a one-hour intravenous infusion of 60 mg/m^2 (the approved dosage in Japan). This trial,[61] alongside KCSG-LU08-01[57] and TITAN,[60] also permitted treatment crossover, unlike any of the other trials. This is an important consideration because treatment crossover could confound OS in these trials. Finally, it should also be noted that the median follow-up times varied considerably in the trials (range 7.5 to 33 months). This is important because if follow-up is not similar across trials, bias may be introduced into studies with shorter follow-up and less mature data as a result of increased censoring.

Table 12 Trial characteristics of trials included in only the MTC base-case and scenario analyses

Trial	Location	Inclusion criteria	Exclusion criteria	Median follow-up
LUME-Lung 1[24]	Europe, Asia, South Africa	• Histologically or cytologically confirmed stage IIIB-IV or recurrent NSCLC of any histology, following relapse or failure of one previous first-line chemotherapy (in the case of recurrent disease one additional previous regimen was allowed for adjuvant, neoadjuvant, or neoadjuvant + adjuvant therapy) • Life expectancy of ≥3 months • At least one target lesion measurable according to RECIST criteria • ECOG PS 0 to 1	• Prior docetaxel or VEGF/VEGFR inhibitor (other than bevacizumab) usage • Radiographic evidence of cavitary or necrotic tumours, centrally located tumours with radiographic evidence (CT or MRI) of local invasion of major blood vessels, or a recent history (<3 months) of clinically significant haemoptysis or a major thrombotic or clinically relevant major bleeding event in the past 6 months • Active brain metastases or leptomeningeal disease • Pre-existing ascites and/or clinically significant pleural effusion	31.7 months
TAILOR[59]	Italy	• Patients with wild-type EGFR advanced NSCLC, who had recurrence or progression after failing platinum-based chemotherapy • Adequate vital functions • ECOG PS ≤2	• Previous treatment with taxanes or anti-EGFR drugs or drugs directed at pemetrexed molecular targets (i.e., thymidylate synthase and dihydrofolate reductase inhibitors)	33 months
WSY001[62]	China	• Aged 18 to 75 years • Pathologically or cytologically confirmed stage IIIB or IV lung adenocarcinoma or postoperative recurrent lung adenocarcinoma incurable by surgery or radiotherapy within 6 months of neoadjuvant or adjuvant chemotherapy • EGFR wild-type and EGFR FISH-positive disease • Received 1 prior platinum-based chemotherapy (including neoadjuvant or adjuvant chemotherapy) • Adequate bone marrow function • Adequate liver function • Adequate renal function • Presence of 2-dimensional measurable disease • Life expectancy of ≥3 months • ECOG PS 0 to 2	• Prior treatment with TKI or pemetrexed • Symptomatic brain metastases • Prior malignant disease (except for basal cell carcinomas) • Pregnancy	14.7 months

Trial	Location	Inclusion criteria	Exclusion criteria	Median follow-up
JMEI[56]	Not reported	• Histologically or cytologically confirmed stage III or IV NSCLC not amendable to curative therapy • Received treatment with only one prior chemotherapy for advanced disease (one prior additional therapy allowed for neoadjuvant, adjuvant, or neoadjuvant + adjuvant therapy) • Adequate bone marrow function • Adequate hepatic function • Adequate renal function • ECOG PS 0 to 2	• Patients with prior docetaxel or pemetrexed treatment • CTCAE ≥grade 3 peripheral neuropathy • An inability to interrupt nonsteroidal anti-inflammatory drugs • Uncontrolled pleural effusions, symptomatic or uncontrolled brain metastases, or significant weight loss (≥ 10% body weight in the preceding 6 weeks) were ineligible.	7.5 months
TITAN[60]	International	• Histologically documented locally advanced, recurrent, or metastatic NSCLC • Disease progression while receiving four cycles of a standard first-line platinum-based chemotherapy doublet (representing a population with poor prognosis); patients who had disease progression during the four cycles of a standard platinum-based chemotherapy doublet could enrol once they had recovered from any toxic effects of the chemotherapy treatment • Adequate haematologica function • Adequate hepatic function • Adequate renal function • Ability to comply with study and follow-up procedures • ECOG PS 0 to 2	• Previous exposure to anti-human-EGFR-directed drugs or drugs directed at pemetrexed molecular targets (i.e., thymidylate synthase and dihydrofolate reductase inhibitors) • Prior chemotherapy or systemic anti-neoplastic therapy other than the permitted platinum-based regimens • Uncontrolled or untreated brain metastasis • Spinal cord compression or other malignancies within the past 5 years (except carcinoma in situ)	24.8 months (chemotherapy arm) 27.9 months (erlotinib arm)

CTCAE= Common Terminology Criteria for Adverse Events; ECOG=Eastern Cooperative Oncology Group; EGFR=epidermal growth factor receptor; FISH=fluorescence in situ hybridisation**;** NSCLC=non-small-cell lung cancer; PS=performance status; RECIST=Response Evaluation Criteria in Solid Tumors; TKI=tyrosine-kinase inhibitor; VEGFR=vascular endothelial growth factor receptor

Source: adapted from Tables 9, 26 and 36 of the CS[1] with additional criteria added from cited source publications (For JMEI[56] eligibility criteria were reported in Hanna *et al*[63])

Table 13 Trial characteristics of trials included in only the MTC sensitivity analyses

Trial	Location	Inclusion criteria	Exclusion criteria	Median follow-up
GEF-ERL[55]	South Korea	• Histologically confirmed stage IIIB or IV NSCLC including recurrent or metastatic disease following failure of first-line chemotherapy • WHO performance status of 0 to 2 • Presence of either an activating EGFR mutation, or two of three clinical factors associated with higher incidence of EGFR mutations. • Brain metastasis permitted if treated at least 4 weeks before entry and clinically stable without steroid treatment for 1 week	• Previous treatment with EGFR signalling inhibitors and radiation therapy within the preceding 4 weeks	16.3 months
KCSG-LU08-01[57]	Korea	• Histologically or cytologically confirmed pulmonary adenocarcinoma that progressed after just 1 previous platinum-based chemotherapy regimen for advanced disease (stage not reported) • Never-smoked (a total of ≤100 cigarettes in their lifetime) • ECOG PS 0 to 2	• Patients with prior TKI or pemetrexed treatment • Symptomatic or uncontrolled brain metastases were ineligible.	15.9 months
V-15-32[61]	Japan	• Histologically or cytologically confirmed stage IIIB or IV NSCLC not amenable to curative surgery or radiotherapy, or postoperative recurrent NSCLC • Failure of prior treatment with one or two chemotherapy regimens (≥1 platinum-based regimen) • WHO PS 0 to 2 • Protocol amendment allowed recruitment of patients without measurable lesions	• Not reported	21 months
S103[58]	Not reported	• Histologically or cytologically confirmed, locally advanced or metastatic non-squamous NSCLC following failure of first-line chemotherapy regimen • ECOG PS 0 to 2 • Only never-smoking patients (<100 lifetime cigarettes) were eligible.	• Prior exposure to agents directed at the human EGFR axis or at pemetrexed molecular targets (e.g. TS or DHFR inhibitors) • Brain metastasis (unless treated and stable after radiotherapy ≥2 weeks) • Concurrent administration of any other antitumour therapy.	14.7 months

DHFR= dihydrofolate reductase; ECOG=Eastern Cooperative Oncology Group; EGFR=epidermal growth factor receptor;; NSCLC=non-small-cell lung cancer; PS=performance status; TKI=tyrosine-kinase inhibitor; TS= thymidylate synthase; VEGFR=vascular endothelial growth factor receptor; WHO=World Health Organisation
Source: adapted from Tables 27 and 36 of the CS[1]

4.4.3 Participant characteristics of included studies

Baseline characteristics of patients summarised in the CS[1] are reported in Table 14 (for trials included in the base-case and scenario analyses) and Table 15 (for trials included in the sensitivity analyses). The ERG considers that the baseline characteristics that are the main drivers of outcomes in patients with adenocarcinoma are ECOG PS, response to prior therapy and EGFR mutation status.

Narratively, the company only focuses on ECOG PS, noting that in TITAN[60] and GEF-ERL,[55] which were included only in the scenario and/or sensitivity analyses, there were a higher proportion of patients (20.0% and 14.6% respectively) with ECOG PS 2 than any of the trials in the base-case. Because patients receive docetaxel with nintedanib, then the ERG considers that the vast majority of patients included in studies should all have ECOG PS≤1. While it is difficult to quantify the proportion, a minimum of 85% would seem reasonable. TITAN[60] (included only in the scenario analyses) does not meet this criterion.

The ERG notes that the proportion of patients with adenocarcinoma ranged from 50% to 100% in the base-case. The three studies (LUME-Lung 1,[24] JMEI[56] and TAILOR[59]) with <75% are appropriately included because they do report subgroup analyses for patients with adenocarcinoma. The ERG further notes that while LUME-Lung 1[24] included some patients with early stage disease at diagnosis, the majority (91.2%) of patients with adenocarcinoma had stage III/IV disease at diagnosis and even more (94.2%) had metastatic disease at screening. In WSY001[62] 71.5% were reported to have stage III/IV disease, the remainder (28.5%) described as having recurrent disease. In JMEI[56] all adenocarcinoma patients were reported to have stage III (18%) or IV (82%) disease at baseline. No information about staging is provided in TAILOR,[59] it being stated patients with metastatic disease were enrolled who “had recurrence or progression after failing platinum-based chemotherapy.”

Response to prior therapy differed across the trials. The proportion of patients with a complete or partial response or stable disease to previous treatment varied from 56.1% in WSY001[62] to 70.7% in JMEI;[56] in LUME-Lung 1[24] it was 70.7% and in TAILOR[59] was 63.9%. As noted in section 4.3.6, in LUME-Lung 1[24] a significant interaction was observed for ‘best response to first-line treatment’ indicating that patients whose best response to first-line therapy was PD would benefit more in terms of OS than the rest of the population. However, as noted by both the company and ERG, this subgroup has a relatively small sample size (n=117) and the results should be interpreted with caution.

With regard to other baseline characteristics, in studies included in the base-case, the ERG observes that median age varied from 54.3 to 60 years, proportion of females from 27.3% to 39.2%, patients with wild-type mutations (EGFR-negative) ranged from [redacted] to 100% and the proportion of never smokers from 17.4% to 35.7%. However, data were not presented for mutation status or never smokers for JMEI[56] and data were incomplete for mutation status for LUME-Lung 1;[24] it is assumed the majority of patients in both trials would be EGFR-negative, an assumption apparently supported by the limited data available from LUME-Lung 1.[24]

In some respects, the characteristics of TITAN,[60] which is included only in the scenario analyses, is like those included in the base-case. There were again a high proportion of patients with unknown EGFR status but it appears from the data available, if it is assumed the ratio of EGFR-positive to EGFR-negative patients in the patients with unknown mutation status is the same as that in the known mutation status, that the majority were EGFR-negative. Arguably what makes this trial most unlike those in the base-case, however, is the aforementioned higher proportion of patients with ECOG PS 2 suggesting a greater proportion of patients with more severe disease in this trial. The ERG further notes that all patients in this trial had stage IIIB (21.7%) or stage IV (88.3%) disease at baseline.

With regard to the trials included only in the sensitivity analyses, it is apparent from Table 15 that three trials appear similar to each other in most respects (GEF-ERL,[55] KCSG- LU08-01[57] and S103[58]) whereas the fourth (V-15-32[61]) appears to be different as it has fewer numbers of female patients, never smokers and patients with adenocarcinoma. The ERG notes that in all trials, EGFR-mutation status is only available from a minority of patients. If it is assumed the ratio of EGFR-positive to EGFR-negative patients in the patients with unknown mutation status is the same as that in the known mutation status patients, then the data appear to support the company's assertion that the proportion of patients with EGFR-positive disease ≥20%; indeed, in each trial there would be a majority of patients with EGFR-positive disease. All patients had stage IIIB and IV disease at baseline in S103[58] and KCSG-LU08-01.[57] In GEF-ERL[55] the proportion was 84.4% with the majority of other patients described as having recurrent disease (13.5%). In V-15-32[61] all patents had stage III/IV disease or were described as being recurrent (83.0% and 17.0% respectively).

Table 14 Patient characteristics of trials included in only the MTC base-case and scenario analyses

Trial and arm	Number at baseline	Adenocarcinoma		Age (years)	Wild-type mutations (EGFR-negative) (%)	ECOG PS 0 to 1 (%)	Female (%)	Never smokers (%)
		%	N					
LUME-Lung 1[24]	1314	50.1	658			100.0		
Nintedanib + docetaxel	655	49.2	322	Median: 60 Range: 53 to 67	[redacted]	100.0	27.3	35.7*
Placebo + docetaxel	659	51.0	336	Median: 60 Range: 54 to 66	[redacted]	100.0	27.3	34.2*
TAILOR[59]	219	69.4	152			92.7		
Erlotinib	109	63.3	69	Median: 66 Range: 40 to 81	100	93.6	29.4	17.4
Docetaxel	110	75.5	83	Median: 67 Range: 35 to 83	100	91.7	33.6	27.2
WSY001[62]	123	100	123			94.3		
Erlotinib	61	100	61	Median: 54.3 Range: 30 to 74	100	93.4	34.4	24.6
Pemetrexed	62	100	62	Median: 55.1 Range: 33 to 75	100	95.2	37.1	27.4
JMEI[56]	571	52.9	302			86.8*		
Pemetrexed	283	55.8	158	Median: 57.4* Range not reported	Not reported	84.8*	39.2*	Not reported
Docetaxel	288	50.0	144	Median: 56.7* Range not reported	Not reported	88.9*	34.0*	Not reported
TITAN[60]	424	49.5	201			80.0		
Erlotinib	203	47.3	96	Median: 59 years Range: 36 to 80 years	36.9 Indeterminate: 15.8 Missing: 43.3	80.8	20.7	14.8
Chemotherapy	221	51.6	114	Median: 59 years Range: 22 to 79 years	33.5 Indeterminate: 16.3 Missing: 45.7	79.2	27.6	19.9

* Subgroup of patients with adenocarcinoma only. [redacted]
Source: adapted from Table 26 with additional data on EGFR mutations and ECOG PS taken from the cited source publications

Table 15 Patient characteristics of trials included in only the MTC sensitivity analyses

Trial and arm	Number at baseline	Adenocarcinoma		Age (years)	Wild-type mutations (EGFR-negative) (%)	ECOG PS 0 to 1 (%)	Female (%)	Never smokers (%)
		%	N					
GEF-ERL[55]	96	90.6	87			85.4		
Gefitinib	48	91.7	44	Median: 60 Range: 37 to 83	25.0 Missing: 56.3	85.4	85.4	91.7
Erlotinib	48	89.6	43	Median: 56 Range: 32 to 81	41.7 Missing: 41.7	85.4	85.4	95.8
KCSG-LU08-01[57]	135†	100.0	135			91.1		
Gefitinib	68†	100.0	68	Median: 58 Range: 40 to 77	22.1 Missing: 50.0	91.2	85.3	100.0
Pemetrexed	67†	100.0	67	Median: 64 Range: 30 to 78	23.9 Missing: 44.8	91.0	85.1	100.0
V-15-32[61]	489¥	77.7	380		5.3 Missing: 88.3	95.7		
Gefitinib	244¥	78.4	191	≤64 years: 56.3		95.5	38.4	29.0
Docetaxel	239¥	77.0	184	≤64 years: 55.3		95.9	38.1	35.7
S103[58]	240	93.8	225		7.9 Missing: 82.1	92.9		
Erlotinib + pemetrexed	78	92.3	72	Median: 55.8 Range not reported		91.0	74.4	100.0
Erlotinib	82	92.7	76	Median: 53.9 Range not reported		92.7	65.9	100.0
Pemetrexed	80	96.3	77	Median: 55.9 Range not reported		95.0	56.3	100.0

Source: adapted from Table 27 with additional data on EGFR mutations and ECOG PS taken from the cited source publications
† Population analysed for safety and efficacy analyses
¥ Population evaluated for safety (described as intention-to-treat population in source paper)

4.4.4 Risk of bias

The company conducted an assessment of the risk of bias of the studies included in the base-case MTC, the results are presented in the CS[1] and shown in Table 16. The ERG considers that the conclusions drawn by the company are valid and that the included studies have an overall low risk of bias.

Table 16 Company's assessment of risk of bias for trials included only in the MTC base-case analyses

Criteria	LUME-Lung 1[24]	TAILOR[59]	WSY001[62]	JMEI[56]
1. Was randomisation carried out appropriately?	Yes	Yes	Yes	Yes
2. Was the concealment of treatment allocation adequate?	Yes	Yes	Not clear	Not clear
3. Were the groups similar at the outset of the study in terms of prognostic factors?	Yes	Yes	Yes	Yes
4. Were the care providers, participants and outcome assessors blind to treatment allocation?	Yes	Yes	No	Not clear
5. Were there any unexpected imbalances in drop-outs between groups?	Not clear	No	No	No
6. Is there any evidence to suggest that the authors measured more outcomes than they reported?	No	No	No	No
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?	Yes	Yes	Yes	Yes
Overall quality (" + + ", " + ", "-")	**+ +**	**+ +**	**+ +**	**+ +**

Source: Appendix 5 (Table 155) of the CS[1]
Information for LUME-Lung 1[24] taken from trial protocol[53]

Trials in the scenario and sensitivity analyses were also assessed for risk of bias (Appendix 5 of the CS[1]). These were considered to be of similarly low risk of bias with the exception of KCSG-LU08-01[57] which was deemed to be at higher risk of bias because of unexpected imbalances in drop-outs between treatment arms.

4.4.5 Individual study findings

Efficacy results from the studies included in the base-case analyses are provided in Table 17. The findings from the studies included in the scenario and sensitivity analyses are not presented here because the patient characteristics of these trials are considered by the ERG to be too different to those of the patient population relevant to the decision problem.

Significant improvements in OS were reported for nintedanib plus docetaxel compared to placebo plus docetaxel in LUME-Lung 1[24] and erlotinib vs pemetrexed in TAILOR.[59] Significant improvements in PFS were only reported in LUME-Lung 1[24] for nintedanib plus docetaxel vs placebo plus docetaxel. In patients treated with adenocarcinoma, median OS varied from 9.2 months (adjusted OS in JMEI[56]) to 13.4 months (pemetrexed arm of WSY001[62]); the OS for nintedanib plus docetaxel therefore appears to compare favourably in LUME-Lung 1[24] (12.6 months). Median PFS ranged from 2.8 months (placebo plus docetaxel arm in LUME-Lung 1[24] to 4.2 months (nintedanib plus docetaxel arm in LUME-Lung 1[24]).

Although median OS was not presented for the adenocarcinoma population in TAILOR,[59] the ERG notes that for the overall population median OS was 8.2 months in the erlotinib arm as compared to 11.7 months for erlotinib in WSY001[62] in which all patients had adenocarcinoma. The median OS for docetaxel in the overall population of TAILOR,[59] was 5.4 months and was lower than the adjusted median OS reported for the adenocarcinoma subgroup of patients treated with docetaxel in JMEI[56] (9.2 months) and OS for the placebo plus docetaxel arm in the adenocarcinoma subgroup of LUME-Lung 1[24] (10.3 months). The median PFS for the erlotinib arm in the overall population in TAILOR[59] (2.9 months) was also slightly lower than for the erlotinib arm in WSY001[62] (4.1 months). Median PFS for the docetaxel arm in the overall population in TAILOR,[59] (2.4 months) was however similar to that of the placebo plus docetaxel arm of adenocarcinoma patients in LUME-Lung 1[24] (2.8 months) and slightly less than the adjusted PFS in the docetaxel arm of JMEI[56] (3.5 months). These findings may be indicative that the trials included different patient populations, as suggested by the ERG in 4.4.1.

Response rates were only reported for three of the trials.[24,56,62] The ERG notes that the ORR for patients treated with docetaxel in JMEI[56] (9.9%) was much greater than reported for placebo plus docetaxel in LUME-Lung 1[24] (3.6%). For patients treated with pemetrexed it was also higher (12.8%) in JMEI[56] than in WSY001[62] (8.1%). The highest ORR was reported for erlotinib (19.7%) in WSY001.[62] The findings for ORR were lowest for either arm in LUME-Lung 1.[24]

Table 17 Individual study findings (inputted into the MTC base-case analyses by the company)

Outcomes		LUME-Lung 1[24]‡		TAILOR[59]		WSY001[62]		JMEI[56]	
		Nintedanib + docetaxel	Placebo + docetaxel	Erlotinib	Docetaxel	Erlotinib	Pemetrexed	Pemetrexed	Docetaxel
N efficacy		322	336	69	83	61	62	158	144
Unadjusted OS	Months	12.6	10.3	NR¥	NR¥	11.7	13.4	NR	NR
	HR (95% CI) p-value	0.83 (0.7 to 0.99) p=0.0359		0.67 (0.48 to 0.95); reported as significant		1.01 (0.66 to 1.54) p= 0.97		NR	
Adjusted OS†	Months	NR	NR	NR	NR	NR	NR	9.0	9.2
	HR (95% CI) p-value	0.81 (0.69 to 0.97) p= 0.0186 (two-sided)		NR		NR		0.92 (0.69 to 1.22) p=0.551	
Unadjusted PFS	Months	4.0	2.8	NR¥	NR¥	4.1	3.9	NR	NR
	HR (95% CI) p-value	0.77(0.62 to 0.96) p= 0.0193		0.76 (0.54 to 1.05)		0.92 (0.62 to 1.37) p= 0.683		NR	
Adjusted PFS†	Months	4.2	2.8	NR	NR	NR	NR	3.5	3.5
	HR (95% CI) p-value	0.84 (0.71 to 1) p= 0.0485 (two-sided)		NR		NR		0.83 (0.65 to 1.06) p= 0.135	
Response	Criteria	RECIST		NR		RECIST		Southwest Oncology Group Criteria	
Objective response	Definition	Objective tumour response (CR + PR)		NR		PR + CR		CR, PR*	
ORR	N evaluated	322	336	NR	NR	61	62	158	144
	N	15	12	NR	NR	12	5		
	%	4.7	3.6	NR	NR	19.7	8.1	12.8	9.9

CI=confidence interval; CR=complete response; HR=hazard ratio; NR=not reported; ORR=overall response rate; OS=overall survival; PFS=progression-free survival; PR=partial response
† No study reported what variables were adjusted for except LUME-Lung 1[24] for OS: [redacted]
‡ For the LUME-Lung 1[24] trial adjusted OS, PFS, and ORR data for the adenocarcinoma subgroup are available from CTR[41]
¥ For TAILOR, median OS and median PFS are only reported for overall population, not adenocarcinoma subgroup
* Complete response: complete disappearance of all measurable and evaluable disease; Partial response: ≥50% decrease in the sum of products of perpendicular diameters of all measurable lesions
Source: adapted from Table 32 of the CS[1] with additional data taken from the source papers

4.4.6 Results from mixed treatment comparisons

As noted in sections 2.2, 3.3 and 4.4.1 above, the ERG does not consider a comparison of nintedanib plus docetaxel to erlotinib is appropriate to decision problem. Furthermore, the ERG also considers there are a number of methodological issues with the MTC and taken together, the ERG does not therefore consider a comparison of nintedanib plus docetaxel with erlotinib is relevant to this STA. Nevertheless, for completeness, a brief description of the results and critique follows.

The following analyses were conducted by the company:

- Base-case analyses
- Sensitivity analyses for base-case (sensitivity analysis i)
- Scenario analyses
- Sensitivity analyses for scenario analyses (sensitivity analysis ii)

While only comparisons of nintedanib plus docetaxel to docetaxel and erlotinib are considered relevant to the NICE scope, some results are presented relative to other comparators included in the MTCs for completeness.

Summary of company's results: overall survival

The results from the base-case analysis for OS are presented in Table 18 and the probabilities of each treatment being the best at improving OS are presented in Table 19. The results from the base-case analysis suggest that nintedanib plus docetaxel is significantly more effective than either docetaxel alone or erlotinib alone. Results from the Bucher indirect comparisons support the findings from the MTC. Nintedanib plus docetaxel is most likely to be the best treatment, suggesting superiority over docetaxel and erlotinib.

Table 18 Summary of OS findings from MTC base-case analysis

Treatment	HR (95% CI) to fixed-effects
Nintedanib + docetaxel vs docetaxel	
Result from MTC	0.83 (0.70 to 0.99)
Result from Bucher indirect comparison	Not applicable
Nintedanib + docetaxel vs pemetrexed	
Result from MTC	0.82 (0.60 to 1.11)
Result from Bucher indirect comparison	0.90 (0.65 to 1.26)
Nintedanib + docetaxel vs erlotinib	
Result from MTC	0.64 (0.46 to 0.90)
Result from Bucher indirect comparison	0.56 (0.38 to 0.82)
Deviance information criterion	0.4095

OS=overall survival; HR=hazard ratio; CI=confidence interval
Notes: The results from the base-case analysis do not feature the random-effect model as there were no instances of two trials with the same comparison
Source: adapted from Table 38 of the CS[1]

Table 19 Probabilities of each treatment being the best at improving OS in base-case analysis

Treatment	Probability of being best
Nintedanib + docetaxel	70.44%
Docetaxel	9.81%
Pemetrexed	16.42%
Erlotinib	3.33%

Source: adapted from Table 39 of the CS[1]

The findings from the scenario and sensitivity analyses broadly support those of the base-case analyses. Nintedanib plus docetaxel also had the highest probability of being the best treatment in the sensitivity analysis (i) of the base-case (49.2%), followed by erlotinib plus pemetrexed (37.17%), a comparator that was not included in the original base-case. In the scenario analysis that assumes equivalent efficacy of docetaxel and pemetrexed, nintedanib plus docetaxel also had the highest probability of being the most effective treatment (78.95%). In the sensitivity analysis (ii) for the scenario analysis, erlotinib plus pemetrexed (54.39%), a comparator that was not included in the base-case analysis, had the highest probability of being the most effective, followed by nintedanib plus docetaxel (34.21%).

Summary of company's results: progression-free analyses

The results from the base-case analysis for PFS are presented in Table 20. The probabilities of each treatment being the best at improving PFS are presented in Table 21. The results suggest that nintedanib plus docetaxel significantly improves PFS in comparison to docetaxel and erlotinib. Results from the Bucher indirect comparisons support the findings from the MTC. Nintedanib plus docetaxel was most likely to be the best treatment, suggesting superiority over docetaxel and erlotinib.

Table 20 Summary of PFS findings from MTC base-case analysis

Treatment	HR (95% CI) to fixed-effects
Nintedanib + docetaxel vs docetaxel	
Result from MTC	0.77 (0.62 to 0.96)
Result from Bucher indirect comparison	Not applicable
Nintedanib + docetaxel vs pemetrexed	
Result from MTC	0.84 (0.61 to 1.15)
Result from Bucher indirect comparison	0.93 (0.67 to 1.29)
Nintedanib + docetaxel vs erlotinib	
Result from MTC	0.70 (0.50 to 1.00)¥
Result from Bucher indirect comparison	0.58 (0.39 to 0.87)
Deviance information criterion	1.568

PFS=progression-free survival; HR=hazard ratio; CI=confidence interval
The results from the base-case analysis do not feature the random-effect model as there were no instances of two trials with the same comparison
¥ The estimate for the upper bound of the 95% credible interval was 0.9958, making the result statistically significant
Source: adapted from Table 40 of the CS[1]

Table 21 Probabilities of each treatment being the best at improving PFS in base-case analysis

Treatment	Probability of being best
Nintedanib + docetaxel	69.69%
Docetaxel	5.01%
Pemetrexed	18.53%
Erlotinib	6.77%

Source: adapted from Table 41 of the CS[1]

The findings from the scenario and sensitivity analyses broadly support those of the base-case analyses although not to the same extent as for the OS analyses. Erlotinib plus pemetrexed, a comparator not in the original base-case, had the highest probability of being the best treatment in the sensitivity analysis (i) of the base-case (61.99%), followed by nintedanib plus docetaxel (25.01%). In the scenario analysis that assumes equivalent efficacy of docetaxel and pemetrexed, nintedanib plus docetaxel had the highest probability of being the most effective treatment (83.57%). In the sensitivity analysis (ii) for the scenario analysis, erlotinib plus pemetrexed (72.23%), a comparator that was not included in the base-case analysis, had the highest probability of being the most effective, followed by nintedanib plus docetaxel (16.42%).

Summary of company's results: overall response rate

Table 22 shows the results of the base-case analysis for ORR. The results suggest that there was no significant difference in ORR between nintedanib plus docetaxel in comparison with docetaxel or erlotinib.

Table 22 Summary of ORR findings from MTC base-case analysis

Treatment	HR (95% CI) to fixed-effects
Nintedanib + docetaxel vs docetaxel	
Result from MTC	1.33 (0.61 to 2.95)
Result from Bucher indirect comparison	Not applicable
Nintedanib + docetaxel vs pemetrexed	
Result from MTC	0.98 (0.33 to 2.84)
Result from Bucher indirect comparison	0.98 (0.34 to 2.83)
Nintedanib + docetaxel vs erlotinib	
Result from MTC	0.33 (0.07 to 1.56)
Result from Bucher indirect comparison	Not applicable
Deviance information criterion	37.47

CI=confidence interval; ORR=overall response rate
The results from the base-case analysis do not feature the random-effect model as there were no instances of two trials with the same comparison.
Source: adapted from Table 42 of the CS[1]

In the sensitivity analysis (i) for the base-case, nintedanib plus docetaxel was statistically inferior to erlotinib, gefitinib and erlotinib plus pemetrexed using a fixed-effects model. The findings from the random-effects model also suggest nintedanib plus docetaxel to be inferior although the wider confidence intervals mean that the difference is no longer statistically significant. The scenario analysis found nintedanib plus docetaxel shows no significant difference in ORR compared with chemotherapy (docetaxel or pemetrexed) or erlotinib. The sensitivity analysis (ii) for the scenario analysis found nintedanib plus docetaxel was not significantly different from chemotherapy (docetaxel or pemetrexed) or erlotinib but was significantly inferior to gefitinib and erlotinib plus pemetrexed.

Summary of company's results: adverse events

The safety outcomes of any grade fatigue, nausea and diarrhoea were only able to be analysed as part of the sensitivity analysis where docetaxel and pemetrexed were assumed to be of comparable efficacy. Although LUME-Lung 1[24] reported additional AEs, including CTCAE grade ≥3 fatigue and nausea, these outcomes could not be compared as either no other linked trial reported equivalent data, or the event rates in one or more of the treatment arms were zero.

ERG critique of the company's results from the mixed treatment comparisons

If the problems with the appropriateness and conduct of the MTCs highlighted in section 4.4.1 are ignored, the ERG makes a number of further observations in relation to the findings reported from the MTCs:

1. Results from the Bucher indirect comparisons support the findings from the MTCs suggesting that inconsistency in the network is not a concern, as additional evidence

from the wider treatment networks corroborate the evidence from simple indirect comparisons.

2. It is stated that unadjusted data were used wherever possible, although only adjusted data were available for JMEI.[56] This trial did not specify the variables which were adjusted for and this lack of information makes it difficult to assess the impact that these adjustments may have had on the data, and therefore the results of the MTCs.

3. Data used to derive results for PFS was PFS assessed by central independent review for LUME-Lung 1,[24] whereas for JMEI[56] and TAILOR[59] local investigator assessed PFS data were used; it is unclear whether the results used in the MTC from WSY001[62] were the results from central independent review or local investigator assessment. However, considering the similarities in the findings from central independent review (HR 0.77, 95% CI: 0.62 to 0.96) and local investigator assessment (HR 0.78, 95% CI: 0.62 to 0.97) for LUME-Lung 1,[24] it seems unlikely that this would greatly impact the results of the MTC.

4. The ERG observes that the company inputted data from the primary PFS analysis for LUME-Lung 1[24] into the MTC; it would have been more informative to use the data from the updated analysis (HR 0.84, 95% CI: 0.71 to 1.00).

5. The company states that trials which provided the active treatment arm with placebo versions of the comparator were not distinguished from trials which did not provide a placebo. The ERG does not consider this to be of major concern, as although the one trial (LUME-Lung 1[24]) which provided a placebo is less likely to be at risk of bias (see also section 4.4.4 [risk of bias]), it is unlikely that this difference would introduce a significant amount of heterogeneity between trials. The ERG notes that due to the small number of studies, a comparison between the fixed and random-effects models to test for heterogeneity could only be conducted for the sensitivity analyses for both OS and PFS. The base-case analysis showed some inconsistency for both OS and PFS effect sizes when direct and indirect evidence was compared. The company suggests that this may be due to differences in EGFR mutation status across studies. The ERG agrees with this assessment and believes the inconsistency may also be caused by differences in patient populations as discussed in sections 4.3.2 and 4.3.3.

The ERG's critique of AEs, including consideration of the evidence input into and derived from the MTC, is presented in section 4.5.

4.5 Critique of the adverse events data

Comparison of adverse events from the direct evidence

In LUME-Lung 1[24] AEs were collected for the full trial population and the subgroup of patients with a histology of adenocarcinoma. In the CS[1] AEs are appropriately only presented for the adenocarcinoma subgroup since this is the population that is relevant to the decision problem.

The company reports that treatment with nintedanib plus docetaxel resulted in additional AEs compared with docetaxel treatment alone. Indeed, drug-related AEs reported in Table 59 of the CS[1] were 81.3% in the nintedanib plus docetaxel arm compared to 72.4% in the placebo plus docetaxel arm. However, the company argues that these data must be considered in the context of there being longer median treatment duration in the nintedanib plus docetaxel arm (Table 23). The ERG notes that in clinical practice in England, the maximum number of docetaxel cycles is likely to be four but notes the median number in the nintedanib plus docetaxel arm was five. It further notes that in both arms, the maximum number of cycles exceeded 40.

Table 23 Treatment exposure in the adenocarcinoma population in LUME-Lung 1

Length of treatment and dose intensity	Nintedanib + docetaxel	Placebo + docetaxel
Median duration of nintedanib/placebo treatment (range)	4.2 months (0.10 to 41.53)	3.0 months (0.07 to 31.10)
Mean dose intensity of nintedanib/placebo (%, SD)	91.2 (15.0)	93.8 (13.3)
Number of docetaxel courses (median, range)	5.0 (1 to 45)	4.0 (1 to 41)
Mean overall dose intensity of docetaxel (%, SD)	98.1 (4.5)	98.7 (3.7)

Source: Table 53 of the CS[1]

The most common specific types of AEs reported by adenocarcinoma patients in LUME-Lung 1[24] are summarised in Table 24. Types of AEs reported by patients in the nintedinab plus docetaxel arm included diarrhoea (43.4%), nausea (28.4%) and vomiting (19.4%) which the company states were successfully managed by dose reduction, dose interruption and/or symptomatic treatment and led to permanent nintedanib discontinuation in <1% of patients (Table 25). These were identified as AESIs relating to nintedanib by the company. Other reported AESIs associated with nintedanib treatment included ALT/AST increase (37.8% vs 9.3% and 30.3% vs 7.2% respectively, Table 24) which were reported to be generally reversible and led to permanent nintedanib discontinuation in <2% of patients (Table 25). For the majority of patients with adenocarcinoma requiring a dose reduction to manage AEs, a single dose reduction of nintedanib or placebo was sufficient (Table 25).

Table 24 Proportion of types of AEs in the adenocarcinoma population in LUME-Lung 1

AEs	Nintedanib + docetaxel		Placebo + docetaxel	
	% any CTCAE grade (% CTCAE grade ≥3)		% any CTCAE grade (% CTCAE grade ≥3)	
All AEs	96.3	(75.9)	94.3	(68.5)
Occurring in ≥5% in either arm:				
• Diarrhoea	43.4	(6.3)	24.6	(3.6)
• Neutrophil count decrease	40.9	(36.3)	40.5	(34.8)
• ALT increased	37.8	(11.6)	9.3	(0.9)
• Fatigue	30.9	(4.7)	29.4	(4.2)
• AST increased	30.3	(4.1)	7.2	(0.6)
• Nausea	28.4	(0.9)	17.7	(0.6)
• WBC decreased	27.8	(19.7)	28.2	(18.3)
• Decreased appetite	23.4	(1.3)	15.6	(1.5)
• Vomiting	19.4	(1.3)	12.3	(0.6)
• Alopecia	17.5	(0.3)	20.4	(0)
• Dyspnoea	16.9	(4.7)	15.6	(6.0)
• Neutropenia	13.8	(11.9)	15.3	(13.5)
• Cough	13.1	(0.9)	18.9	(0.6)
• Pyrexia	12.2	(0.6)	14.1	(0.3)
• Stomatitis	11.3	(1.3)	7.8	(0.3)
• Haemoglobin decreased	10.9	(0.9)	13.8	(2.1)
• Constipation	6.9	(0)	11.7	(0.3)
SAEs	34.7	(31.3)	32.1	(26.6)
Occurring in ≥5% in either arm:				
• Febrile neutropenia	5.6	(5.6)	1.8	(1.8)
• Malignant neoplasm progression	3.8	(3.8)	2.4	(2.1)
• Dyspnoea	2.8	(2.5)	5.4	(4.8)
• Pneumonia	2.8	(2.2)	3.6	(1.8)
• Diarrhoea	1.9	(1.6)	2.1	(1.8)
• General physical health deterioration	1.9	(1.9)	1.5	(1.2)
• Neutropenia	1.9	(1.6)	3.3	(3.3)
• Asthenia	1.6	(1.3)	0.6	(0.3)
• Respiratory failure	1.6	(1.6)	0.3	(0.3)
• Vomiting	1.6	(0.6)	1.2	(0.6)
• Atrial fibrillation	1.3	(0.9)	0	(0)
• Chest pain	1.3	(0.9)	1.8	(1.5)
• Pleural effusion	1.3	(1.3)	1.8	(1.2)
• Sepsis	1.3	(1.3)	0.3	(0.3)
• Pyrexia	0.6	(0)	1.2	(0)

AEs=adverse events; ALT=Alanine aminotransferase; AST=Aspartate transaminase; CTCAE= Common Terminology Criteria for Adverse Events; WBC=white blood cell
* As judged by the local investigator
Source: adapted from Tables 62 and 63 of the CS[1]

Table 25 AEs leading to dose interruptions, reductions or discontinuations in LUME-Lung 1

AEs	Nintedanib + docetaxel	Placebo + docetaxel
	%	%
At least 1 temporary interruption of nintedanib/placebo	52.2	41.4
At least 1 temporary interruption of nintedanib/placebo >14 consecutive days	10.0	6.6
1 dose reduction of nintedanib/placebo	17.2	6.6
2 dose reductions of nintedanib/placebo	4.7	0
AEs leading to dose reduction of nintedanib or placebo	21.6	6.6
AEs leading to dose reduction of nintedanib or placebo occurring in ≥1% in either arm:		
• Diarrhoea	8.1	3.3
• ALT increased	7.8	0.6
• AST increased	3.8	0
• Vomiting	2.2	0.6
• Nausea	1.3	0.3
AEs leading to dose reduction of docetaxel	16.6	12.3
AEs leading to permanent discontinuation of last study medication	20.9	17.7
AEs leading to permanent discontinuation of last study treatment occurring in ≥1% in either arm:		
• ALT increased	1.6	0
• Malignant neoplasm progression	1.6	1.5
• AST increased	1.3	0.3
• Dyspnoea	1.3	3.3

AEs=adverse events; ALT=Alanine aminotransferase; AST=Aspartate transaminase
Source: adapted from Tables 54, 55, 56, 58 and 59 of the CS[1]

The ERG notes from Table 24 that the incidence of AEs and SAEs was similar between treatment arms but the incidence of grade ≥3 AEs and SAEs was greater in the nintedanib plus docetaxel arm. The ERG notes grade 3 AEs tend to be particularly significant and can lead to drug discontinuation and hospitalisation but grade 2 AEs may also be clinically relevant by also impacting negatively on HRQoL. It is further noted that dose reduction schemes for nintedanib/placebo specified in Table 11 (pages 63 to 64) of the CS[1] included grade 2 AEs, namely vomiting of CTCAE grade ≥2 within 3 days after docetaxel therapy, diarrhoea of CTCAE grade 2 for >7 consecutive days and AST or ALT elevations of CTCAE grade 2 in conjunction with bilirubin elevations of CTCAE grade ≥1, or AST or ALT elevations of CTCAE grade ≥3. CTCAE grade 2 diarrhoea was included as an AE in its economic model (see Table 35 in section 5.4.7). From the CTR[41] (page 332, Table 12.2.2.4.1.2: 2) the ERG observes 17.8% of patients in the nintedanib plus docetaxel arm reported CTCAE grade 2 diarrhoea compared to 7.2% in the placebo plus docetaxel arm; CTCAE grade 2 + was 24.0% and 10.8% respectively.

Aside from AESIs related to nintedanib, a number of other AESIs were also identified and reported in the CS.[1] These were generally balanced across treatment arms. Exceptions identified by the company were:

- AESIs related to VEGFR inhibitor class effects: a higher frequency of any CTCAE grade hypertension in the nintedanib plus docetaxel arm (3.4% vs 0.6%). However, the incidence of CTCAE grade ≥3 hypertension was balanced across arms (0.9% vs 0.6%)
- AESIs based on potential associations/complications of AEs: any CTCAE grade dehydration only occurred in the nintedanib plus docetaxel arm (1.9% any CTCAE grade and 0.6 % CTCAE grade ≥3)
- AESIs related to potential interaction with concomitant chemotherapy: mucositis was more common in the nintedanib plus docetaxel arm (16.6%) than the placeboplus docetaxel arm (11.4%); however, the incidence of grade ≥3 mucositis was balanced across arms (1.3 % vs 0.6% respectively)
- AESIs selected based on competitor labelling: any CTCAE grade cutaneous skin reactions and any CTCAE grade rash were more common in the nintedanib plus docetaxel arm than placebo plus docetaxel arm (15.6% vs 10.5% and 12.5% vs 8.7% respectively; the incidence of both grade ≥3 cutanous skin reactions and grade ≥3 rash was however balanced across arms (1.3 % vs 0.6% for cutaneous skin reactions and 0.3% and 0% for rash)
- AESIs related to cardiac events: any CTCAE grade cardiac arrhythmias occurred at a slightly higher incidence in the nintedanib plus docetaxel arm (11.6%) compared with the placebo plus docetaxel arm (7.5%); however, the incidence of grade ≥3 cardiac arrhythmias was balanced across arms (2.2 % vs 1.5% respectively).

Other AEs identified as AESIs by the company were interstitial lung disease, photosensitivity conditions and anaphylactic reaction. Frequencies of these AESIs were uncommon (1.3%, 0.3% and 0 respectively for nintedanib plus docetaxel compared to 0.3%, 0.6% and 0.3% respectively in the placebo plus docetaxel arm). All were CTCAE grade <3 except for interstitial lung disease (0.3%) and anaphylactic reaction (0.3%) in the placebo plus docetaxel arm.

The AEs reported in LUME-Lung 1[24] which are of greatest concern, are fatal AEs where some imbalances were reported between treatment arms, fatal AEs being more common in the nintedanib plus docetaxel arm (Table 26). The only exception was fatal AEs occurring within 6 weeks of treatment which the company argues were well-balanced "indicating that the combination therapy with nintedanib and docetaxel had no acute toxicity[(42)]" (page 154 of the CS[1])

Table 26 Summary of fatal AEs in LUME-Lung 1 in the adenocarcinoma population

Fatal adverse events (AEs)	Nintedanib + docetaxel	Placebo + docetaxel
	n (%)	n (%)
All fatal AEs	56 (17.5)	32 (9.6)
• Fatal AEs occurring within 6 weeks	13 (4.0)	12 (3.6)
• Fatal AEs not attributed to progressive disease	20 (6.3)	8 (2.4)
• Fatal AEs attributed to progressive disease*	36 (11.3)	24 (7.2)
Drug-related fatal AEs	6 (1.9)	1 (0.3)
• Sepsis	2 (0.6)	0
• Dehydration	1 (0.3)	0
• Diverticulum intestinale†	1 (0.3)	0
• Ischaemic stroke	1 (0.3)	0
• Large intestine perforation†	1 (0.3)	0
• Neutropenic infection	1 (0.3)	0
• Dyspnoea	0	1 (0.3)

* Attribution to progressive disease by the local investigator, as documented on the Case Report Form
† One patient experienced more than 1 fatal AE considered drug-related (patient with large intestine perforation and diverticulum intestinale)
Source: adapted from Tables 60 and 61 of the CS[1]

The company argues that data on fatal AEs are confounded in two ways. Firstly, the company argues the extent of exposure was longer on nintedanib plus docetaxel compared to docetaxel alone. As noted in Table 23, the median number of cycles of docetaxel that patients received was greater in the nintedanib plus docetaxel arm than in the docetaxel arm (5 vs 4 respectively). Therefore it is argued that the higher exposure to docetaxel may have contributed, at least in part, to the higher incidence of fatal AEs of sepsis caused by neutropenia in the nintedanib arm through the known myelotoxic effect of docetaxel. Consequently neutropenia and sepsis are considered possible side effects of nintedanib therapy in combination with docetaxel and are regarded as important identified risks for future monitoring and ongoing safety surveillance. Secondly, the analysis focusing on the on-treatment fatal AEs resulted in a skewed view of the deaths that occurred during the study. The company states that further review of PD and non-PD deaths occurring during the entire observation period revealed no other safety pattern suggestive of nintedanib associated toxicities.[42]

The ERG considers that the number of deaths related to AEs is relatively small but agrees with the company that AE related deaths need to be monitored in future. The ERG considers that the greater number of PD related deaths in the nintedanib plus docetaxel arm could be related to the fact that PFS was longer in this arm and so patients were on treatment longer; this may also account for differences in non- PD deaths. However, the ERG does not consider that the greater number of cycles of docetaxel received by patients treated with nintedanib is likely to have been a confounder since, as reported by the National Confidential

Enquiry into Patient Outcome and Death, most patients with life threatening toxicity tend to experience fatal AEs during the first cycle of treatment;[64] it is reported by the company that dose intensity was similar between arms (98.1% vs 98.7%) .

Comparison of adverse events from the indirect evidence

As noted in section 4.4.1, it was only possible to conduct MTCs for safety outcomes if it was assumed pemetrexed and docetaxel had equal tolerability, an assumption which the ERG reiterates is not supported by the evidence (e.g. see JMEI[56]). The ERG has however presented the data input into the MTC as this shows AEs across two trials: LUME-Lung 1[24] and WSY001.[62] However, as WSY001[62] it should be noted that WSY001[62] is a trial of Asian patients conducted in China and AEs in a population in England may differ as a result of differences in co-morbidities, smoking history and pharmacokinetics between these populations. The ERG notes that the data from these trials support the generally held view that erlotinib is generally better tolerated than nintedanib plus docetaxel or docetaxel alone.

Table 27 Safety results for adenocarcinoma populations of trials included in MTC base-case analysis

Outcomes		LUME-Lung 1[24]		WSY001[62]	
		Nintedanib + docetaxel (n=320)	Placebo + docetaxel (n=333)	Erlotinib (n=61)	Pemetrexed (n=62)
Any CTCAE grade AE: fatigue	N	99	98	12	16
	%	30.9	29.4	19.7	25.8
Any CTCAE grade AE: nausea	N	91	59	1	15
	%	28.4	17.7	1.6	24.2
Any CTCAE grade AE: diarrhoea	N	139	82	10	2
	%	43.4	24.6	16.4	3.2
CTCAE grade ≥3 fatigue	N	15	14	0	0
	%	4.7	4.2	0	0
CTCAE grade ≥3 nausea	N	3	2	0	2
	%	0.9	0.6	0	3.2

AE=adverse event; CTCAE= Common Terminology Criteria for Adverse Events
Source: adapted from Table 34 of the CS[1]

4.6 Critique of the health related quality of life data

The company reports data on HRQoL data for LUME-Lung 1[24] that appears to have been reported in a poster presentation at the World Conference on Lung Cancer, Sydney, Australia, October 2013.[47] It is stated that data are reported for patients with adenocarcinoma only although baseline data were only available for all patients, regardless of histology. The ERG also notes that the company states that longitudinal analysis reported that nintedanib plus docetaxel did not result in a change in global health status/QOL in patients with adenocarcinoma. Self-reported HRQoL assessments by EORTC

questionnaires also revealed that there were no significant differences in cough, dyspnea or pain in patients over time or between those receiving nintedanib plus docetaxel and those receiving placebo plus docetaxel. Nintedanib-treated patients did however achieve numerically better cough and pain scores than placebo-treated patients, suggesting an improvement in HRQoL for these domains. Furthermore, statistically significant differences were observed between groups for three individual pain items ('have pain', 'pain in chest' and 'pain in arm and shoulder'). On the other hand, the TTD for diarrhoea was significantly worsened in the nintedanib plus docetaxel arm; there was no significant difference between arms for nausea and vomiting, or appetite loss (Table 28).

Table 28 Time to deterioration of nausea and vomiting, appetite loss and diarrhoea in patients with adenocarcinoma in LUME-Lung 1

Symptom	HR (95% CI)
Nausea and vomiting	1.23 (1.00 to 1.51)
Appetite loss	1.13 (0.92 to 1.38)
Diarrhoea	1.86 (1.51 to 2.30)*

*p<0.05
HR=hazard ratio
Source: Table 24 of the CS[1]

The ERG notes that the response to the HRQoL questionnaire appears to be very good; the company states that over 80% of patients completed HRQoL questionnaires over the first 20 cycles of treatment and approximately 70% of patients completed the questionnaire at the end of the treatment visit. It is noted that the main drivers of HRQoL in this population tend to be cancer related symptoms. Taking into account the findings for ORR and PFS (see section 4.3.6) in which it was observed that the addition of nintedanib did not make a major difference to response rates but did lead to increased rates of tumour control and slower progression on average, it is perhaps unsurprising that dramatic differences in HRQoL were not seen on initiation of therapy. It is interesting to observe significant differences in pain symptoms as fatigue, dyspnoea and cough are often reported to be more troublesome to patients and their families.[65] The worsened TTD for diarrhoea for patients treated with nintedanib plus docetaxel is unsurprising given the greater proportion of diarrhoea AEs in this arm (see section 4.5). The increased rates of diarrhoea did not seem have any major impact on global health status/QoL.

No attempt was made by the company to compare HRQoL between nintedanib plus docetaxel and erlotinib. For reasons stated above (sections 2.2, 3.3 and 4.4.1), the ERG does not consider such a comparison is relevant to the decision problem, even if such a comparison were possible.

4.7 Conclusions of the clinical effectiveness section

Clinical evidence has been submitted to NICE from the company in support of the use of nintedanib for previously treated locally advanced, metastatic or recurrent NSCLC of adult patients with adenocarcinoma tumour histology. The NICE scope did not specify adenocarcinoma nor did it refer to locally recurrent disease. This population is however in line with the anticipated marketing authorisation. While none of the scope, decision problem or anticipated marketing authorisation refer to the EGFR mutation status of NSCLC tumours, in England, the majority of patients (85 to 90%) are likely to be EGFR-negative. The ERG further notes that because patients who receive nintedanib also receive docetaxel, then patients who are likely to be eligible for treatment with nintedanib will also be ECOG PS 0 to 1.

Direct evidence is presented for nintedanib plus docetaxel vs placebo plus docetaxel from one RCT (LUME-Lung 1[24]). Indirect evidence for nintedanib + docetaxel vs erlotinib is presented from MTCs. While both docetaxel and erlotinib are specified as comparators in the NICE scope, given that erlotinib is likely to be preferred when patients have a poorer performance status and/or have EGFR-positive tumours, or be treatment naïve for a TKI, the ERG agrees with the company that erlotinib is not a relevant comparator and that docetaxel is the only appropriate comparator for this STA.

LUME-Lung 1[24] is a phase III double-blind RCT which compares nintedanib plus docetaxel vs placebo plus docetaxel. It is considered to have a low risk of bias. As a result of the exclusions of certain types of patients, the patient population appears to be fitter than would be found in clinical practice in England. This could partially explain why the post-study therapy rate is relatively high (55.8%).

The findings from this trial suggest nintedanib plus docetaxel significantly improved PFS and OS in comparison to placebo plus docetaxel in the subgroup of patients with adenocarcinoma. After a median follow-up of 31.7 months the gain in median PFS was 1.4 months (4.2 months vs 2.8 months) and gain in median OS was 2.3 months (12.6 months vs 10.3 months). However, the ERG does not consider that the assumption of proportional hazards is consistent with the trial data, and therefore use of these results in cost-effectiveness modelling should not be based implicitly or explicitly on this assumption. Nintedanib plus docetaxel also resulted in an increase in some types of AEs, particularly diarrhoea, nausea, vomiting and increases in ALT/AST. The majority of these AEs can be managed by dose reductions of nintedanib. The ERG is in agreement with the company that apparent improvements seen in terms of PFS and OS in the adenocarcinoma patients were achieved without substantial alterations in self-reported HRQoL.

5 COST EFFECTIVENESS

5.1 Introduction

This section provides a structured critique of the economic evidence submitted by Boehringer Ingelheim Ltd in support of the use of nintedanib in combination with docetaxel for the treatment of adult patients with locally advanced, metastatic or recurrent NSCLC of adenocarcinoma tumour histology after first-line chemotherapy. The two key components of the economic evidence presented in the CS[1] are (i) a systematic review of the relevant literature and (ii) a report of the company's de novo economic evaluation. The company also provided an electronic copy of their economic model that was developed in Microsoft Excel.

5.2 ERG comment on company's review of cost-effectiveness evidence

5.2.1 Objective of the company's cost-effectiveness review

On page 175 of the CS,[1] the company explains that "The scope of the systematic review is to review all available published data on economic evaluations of second-line therapies for locally advanced or metastatic NSCLC that could inform a HTA submission based on Boehringer Ingelheim's second-line comparative trials of nintedanib". This single systematic review was performed to identify clinical, cost-effectiveness, resource use and cost data as well as studies reporting utility scores for health states within the model.

Details of the cost-effectiveness search strategies employed are included in Appendix 10 of the CS. Medline (via PubMed), Medline R-In Process (via PubMed), EMBASE, and The Cochrane Library (via NHS EED) were searched for data on economic models, costs, resource use associated with NSCLC, HRQoL and utilities. HEED and EconLit were searched for data on HRQoL and utilities. The time horizon for the search for full economic studies was 2000 to February 2014 and for cost analyses was 2012 to 2013.

The search of the literature yielded no relevant studies. The ERG is satisfied with the company's search strategy and is confident that the company did not miss any relevant published articles.

5.2.2 Eligibility criteria used in the study selection

The inclusion/exclusion criteria used in the study selection are presented in Table 29. The ERG is satisfied that these criteria are relevant to the decision problem.

Table 29 Economic evaluation search inclusion/exclusion criteria

Parameter	Inclusion criteria	Exclusion criteria
Population	Relapsed or refractory NSCLC (RR NSCLC) (receiving second-line chemotherapy or relapsed/refractory to first-line chemotherapy)	Any patient population other than RR NSCLC
Interventions	Any second-line chemotherapy for RR NSCLC: • Monotherapy • Combination therapy with other chemotherapy Other interventions that are considered standard care in the patient population that will be relevant to the economic model	Patients who were treatment-naïve or had received more than first-line therapy
Outcomes	Economic models: • Cost-utility analyses • Cost-effectiveness analyses • Cost-benefit analyses • Cost-minimisation analyses	No outcomes of interest included
Study design	Economic models: Economic studies	Not an economic model
Language restrictions	English language	Non-English language
Date	Economic models: 2002 onwards	Prior to the year 2002*
Country	Any	None

*Abstracts published prior to 2011 and systematic reviews published prior to 2009 were excluded
Source: Table 72 of CS[1]

5.2.3 Included and excluded studies

No relevant studies were identified by the company.

5.2.4 Conclusions of the cost effectiveness review

The ERG notes that since nintedanib in combination with docetaxel has not yet received a full marketing authorisation from the EMA for the second-line treatment of adult patients with adenocarcinoma, the lack of economic evaluations of relevance to the decision problem is not unexpected.

5.3 ERG critique of the company's literature review

The ERG is satisfied with the company's search strategy and stated eligibility criteria for inclusion/exclusion. The ERG is confident that the company did not miss any relevant published papers.

The ERG acknowledges that the company reported the methods and results of a series of literature reviews at key points throughout the cost-effectiveness section in the CS;[1] the ERG considered the results of these additional reviews to be very helpful.

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

5.4.1 NICE reference case checklist

Table 30 NICE Reference case checklist completed by ERG

Attribute	Reference case	Does the de novo economic evaluation match the reference case?
Decision problem	The scope developed by NICE	Partial - the population was limited to patients with adenocarcinoma
Comparator(s)	Alternative therapies routinely used in the NHS	Yes
Perspective costs	NHS and Personal Social Services	Partial - only NHS costs were included in the model
Perspective benefits	All health effects on individuals	Yes, health effects to the individual are captured via QALYs
Form of economic evaluation	Cost-effectiveness analysis	Cost-effectiveness analysis
Time horizon	Sufficient to capture differences in costs and outcomes	Lifetime horizon was used (15 years)
Synthesis of evidence on outcomes	Systematic review	Nintedanib + docetaxel vs docetaxel: direct trial data from LUME-Lung 1[24] was used. Nintedanib + docetaxel vs erlotinib: hazard ratios were taken from the results of the company's network meta-analysis (fixed-effects model)
Outcome measure	Quality adjusted life years (QALYs)	QALYs were used which is appropriate
Health states for QALY	Described using a standardised and validated instrument	EQ-5D was used, with data collected mainly from participants in LUME-Lung 1.[24] Data from published sources[66,67] were used in the sensitivity analysis
Benefit valuation	Time-trade off or standard gamble	Time-trade off
Source of preference data for valuation of changes in HRQoL	Representative sample of the public	UK preference tariff based on public sample. Data for assigning valuation health states were collected directly from trial participants
Discount rate	An annual rate of 3.5% on both costs and health effects	Benefits and costs were discounted at the 3.5% rate
Equity	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	All QALYs estimated by the model have the same weight
Sensitivity analysis	Probabilistic sensitivity analysis	Deterministic, scenario and probabilistic sensitivity analyses were undertaken by the company

Table 31 Critical appraisal checklist for the economic analysis completed by ERG

Question	Critical appraisal	ERG comment
Was a well-defined question posed in answerable form?	Yes	
Was a comprehensive description of the competing alternatives given?	Yes	
Was the effectiveness of the programme or services established?	Partial	For the direct comparison of nintedanib + docetaxel vs docetaxel, effectiveness was established using data from LUME-Lung 1[24] For the indirect comparison of nintedanib + docetaxel vs erlotinib, an MTC was undertaken. The ERG does not consider the results of the MTC to be valid or reliable, nor does it consider the comparison to be relevant to the decision problem
Were all the important and relevant costs and consequences for each alternative identified?	Mostly	Specific issues are discussed insection 5.5 and the impact on the ICER is presented insection 6 of the ERG report
Were costs and consequences measured accurately in appropriate physical units?	No	Specific issues are discussed insection 5.5 and the impact on the ICER is presented insection 6 of the ERG report
Were the cost and consequences valued credibly?	No	Specific issues are discussed insection 5.5 and the impact on the ICER is presented insection 6 of the ERG report
Were costs and consequences adjusted for differential timing?	Yes (with errors)	Specific issues are discussed insection 5.5 and the impact on the ICER is presented insection 6 of the ERG report
Was an incremental analysis of costs and consequences of alternatives performed?	Yes	ICERs were calculated correctly
Was allowance made for uncertainty in the estimates of costs and consequences?	Yes	Deterministic, scenario and probabilistic sensitivity analyses were undertaken
Did the presentation and discussion of study results include all issues of concern to users?	Yes	

5.4.2 Description of the company's economic model

A schematic of the company's submitted economic model is provided in the CS[1] and is reproduced in Figure 2.The company's cost-effectiveness model is a partitioned survival Markov model which comprises three health states: progression-free (on or off treatment) (PF); PD and death (D). All patients enter the model in the PF state. At the beginning of each time period patients can either remain in the same health state or progress to a worse health state, i.e. from PF to PD or death; or from PD to death. The model uses the partitioned survival (also known as area under the curve or AUC) method to determine the proportion of patients in each of the three health states during each model cycle. The proportion of patients in the PD state is estimated as the difference between OS and PFS. Estimates of OS and PF are based on PF and OS survival data from LUME-Lung 1[24] and the corresponding parametric survival models. The model assumes that patients receive best supportive care (BSC) following the discontinuation of active second-line treatment. The model also allows some patients in the progressed state to have subsequent treatments. The costs of subsequent treatment are included in the economic evaluation; however, the impact of subsequent therapy is not included in the model. Variants of this model structure have been used in the modelling of metastatic oncology for numerous STAs including two recent NICE STAs that considered locally advanced or metastatic NSCLC (NICE TA192[14] and NICE TA258[13]).

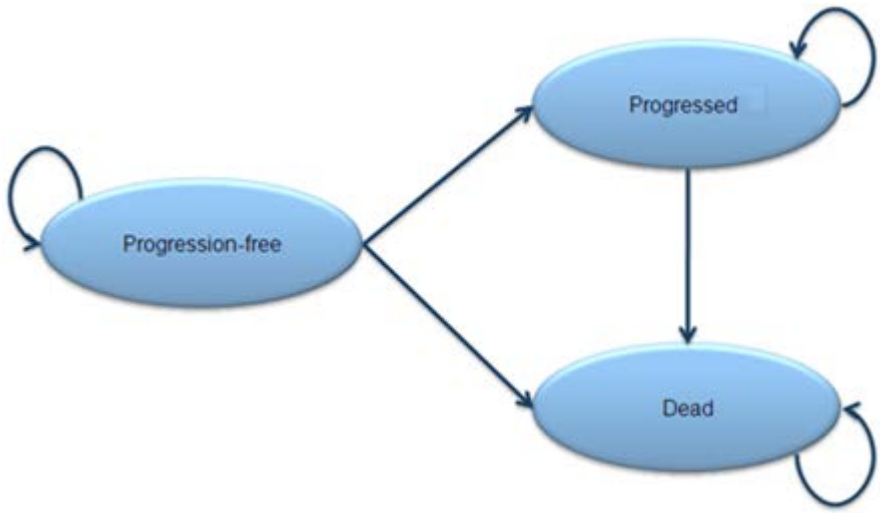


Figure 2 Schematic of company's model
Source: Figure 24 of the CS[1]

The model has been developed in Microsoft Excel using a 3-weekly cycle length. It includes a half-cycle correction and the time horizon is set at 15 years. Health effects are measured in QALYs. A summary of all of the variables applied in the economic model is shown in

Table 79 of the CS;[1] details displayed in the table include the values used, range (distribution) and source.

5.4.3 Population

The company states on page 180 of the CS[1] that the model population was based on the findings of LUME-Lung 1[24] and included patients with locally advanced and/or metastatic, stage IIIB/IV or recurrent NSCLC with adenocarcinoma histology who failed after first-line chemotherapy.

5.4.4 Interventions and comparators

The company's base-case economic evaluation compares nintedanib plus docetaxel with docetaxel. The interventions are implemented in the model in accordance with their current, or anticipated, full marketing authorisations and doses.

Patients receiving nintedanib plus docetaxel are assumed to take two 100mg capsules of nintedanib twice daily; there are 120 capsules in each 100mg pack. The assumed NHS list price per 30-day pack is £2151.10. The ERG notes that there is also a 150mg capsule available; there are 60 capsules in each 150mg pack. In response to a clarification question raised by the ERG, the company indicated that the price of both packs is likely to be the same. Nintedanib plus docetaxel therapy needs to be given for a minimum of four cycles before nintedanib can be administered as monotherapy. There is no administration cost associated with nintedanib. Patients receiving intravenous docetaxel are assumed to receive 75mg/m^2 on day 1 of a 21-day cycle. The monthly cost of docetaxel is estimated to be £49 (using electronic Marketing Information Tool [eMIT] prices[68]) and has a monthly administration cost of £221.43 (NHS Reference Cost 2012/13).[69]

The submitted economic model also permits the comparison of nintedanib plus docetaxel with erlotinib. In the model the dose of erlotinib is assumed to be 150mg per day and the MIMS 2013 price for a pack of 30 tablets is £1631.53.[70] It is noted that erlotinib has an associated patient access scheme, which the company took into account by undertaking a number of sensitivity analyses in which a range of discounts were applied to the list price. However, the company emphasises on page 184 of the CS[1] that erlotinib is not a relevant comparator and considers that patients treated with erlotinib are a different patient population.

Some patients in the model go on to receive subsequent therapy after progression: the company's external expert stipulated that 5% would receive erlotinib, 25% would receive a platinum doublet and 70% would receive BSC. The cost per month of BSC (£406.63 per

cycle [3 weeks]) is as per TA310 (Afatinib NICE submission) [15] as recommended by the company's external expert.

5.4.5 Perspective, time horizon and discounting

The company states that the economic evaluation is undertaken from the perspective of the NHS and Personal Social Services. However, it should be noted that the model does not include all likely Personal Social Services costs. The time horizon is set at 15 years and both costs and outcomes are discounted at 3.5% in line with the NICE Methods Guide to Technology Appraisal.[54]

5.4.6 Treatment effectiveness and extrapolation

Modelling treatment effectiveness (nintedanib plus docetaxel vs docetaxel)

Kaplan-Meier survival curves for OS and PFS for nintedanib plus docetaxel and docetaxel monotherapy were available from LUME-Lung 1[24] and show the proportion of patients in the model's three health states at each time point. These data were incorporated into the cost-effectiveness model by using full parametric approximation of the raw data in the base-case. In the sensitivity analyses, K-M data from LUME-Lung 1[24] were used to model OS (until at least 5% of trial patients are still at risk) and were extrapolated using parametric function as a tail to the Kaplan-Meier data to provide a lifetime time horizon. Survival data from LUME-Lung 1[24] were mature and the proportion of censored patients in both treatment arms were similar. However, in order to facilitate extrapolation of trial data beyond the time horizon, OS and PFS data were analysed using parametric survival models. Parametric survival curves were fitted to PFS and OS K-M curves using two approaches: 1) joint models, statistical models including data for both treatment groups with a term for treatment and 2) separate models, statistical models fitted to each randomised treatment arm separately.

Choice of statistical model

The "goodness of fit" based on Akaike information criteria (AIC) indicated that joint models were appropriate. However, the intercept and scale parameters of the separately fitted curves indicated that the curves should not be forced into the same model, thus separate curves were selected for OS and PFS. The log-logistic model had the lowest AIC among the separately fitted OS models, and the Weibull model had the lowest AIC among the separate proportional hazard models for OS; therefore, these were selected to model the OS data. The log-normal model had the lowest AIC among the separate PFS fits, and the Weibull had the lowest AIC among the separate proportional hazard models for PFS; therefore, these were selected to model PFS.

The company states that the long-term extrapolation of trial data was validated with a group of UK clinicians and against data from SEER using the SEER*Stat software, as well as against data from LUCADA. As noted in section 5.5, no references were provided to identify the specific DEER and LUCADA data sets employed.

Survivals implemented in the model

Survival modelling options programmed into the cost-effectiveness model are displayed in Table 32. In the base-case, the analysis used separate models for PFS and OS, with log-normal distribution for the PFS and log-logistic distribution for the OS.

Table 32 Survival estimation models employed in the company's model

Progression-free survival	Overall survival (OS)
Separate model - Log-normal (base-case)	Separate model - Log-logistic (base-case)
Separate model - Weibull	Separate model - Weibull
Kaplan-Meier curve*	Kaplan-Meier curve*
	Kaplan-Meier curve and SEER lognormal†
	Kaplan-Meier curve and Separate Log-logistic†
	Kaplan-Meier curve and Separate Weibull†
	Kaplan-Meier curve and LUCADA lognormal†

LUCADA= National Lung Cancer Audit database; OS=overall survival; PFS=progression-free survival; SEER=Surveillance, Epidemiology and End Results
* With this option, the model does not extrapolate the PFS/OS with the use of parametric models but it uses the K-M curves for PFS/OS obtained directly from the LUME-Lung 1[24] trial. Note that this option only applies for nintedanib + docetaxel and docetaxel monotherapy
† With this option, the Kaplan-Meier curves from the LUME-Lung 1[24] trial are used for the estimation of OS until patient number at risk drops down to 5% of original patients, afterwards parametric models are used
Source: Table 75 of the CS[1]

Modelling treatment effectiveness of erlotinib

As OS and PFS K-M curves for erlotinib were not available, model OS and PFS inputs for erlotinib were derived by applying HRs (i.e., vs nintedanib plus docetaxel) obtained from the mixed treatment comparisons to the OS and PFS of nintedanib plus docetaxel. The company considers that HRs can only be used if the survival distribution is a proportional hazard model such as exponential, Weibull, or Gompertz. Thus, in the model, erlotinib can be evaluated only if a Weibull distribution is selected for both OS and PFS. The model base-case analysis utilised HRs from the MTC base-case. The HR for OS was 0.64 (95% CI: 0.46 to 0.90) and the HR for PFS was 0.7 (95% CI: 0.5 to 1.0). The company used results from the fixed-effects model because there was one trial per comparison.

5.4.7 Health related quality of life

Utility

Health related quality of life data were collected during LUME-Lung 1[24] using the EQ-5D instrument, in line with the NICE Methods Guide to Technology Appraisal.[54] Data from the LUME-Lung 1[24] were analysed using a longitudinal model adjusted for baseline ECOG score, prior treatment with bevacizumab, presence of brain metastases, controlling for health status and key adverse events. Key model utility values for PF and PD are displayed in Table 33.

Table 33 Utilities for progression-free and post-progression states

Nintedanib + docetaxel and docetaxel arms - Pooled	Progression free without adverse events	
	Mean	Standard error
Week 0	0.710	0.01
Week 3	0.721	0.01
Week 6	0.707	0.01
Week 9	0.699	0.01
Week 12	0.692	0.01
Week 15	0.687	0.01
Week 18	0.682	0.01
Week 21	0.677	0.02
Week 24	0.671	0.02
Week 27	0.666	0.02
Week 30	0.661	0.02
Treatment arm	**Progressive disease**	
	Mean	**Standard error**
Nintedanib + docetaxel	0.64	0.01
Docetaxel	0.64	0.01

Source: Table 80 of the CS[1]

Progression free utility estimates

The analysis estimated utility values over time for PF patients from week 0 to week 30 in 3-week intervals - without a treatment term. An assumption of the linear extrapolation of trend observed until week 30 for the PF health state is employed in the base-case to allow modelling of continuing change in utility in the PF state beyond the trial data.

Progressed disease utility estimates

In contrast to the estimation of PF utilities over time, mean PD utilities were used in the base-case model to accommodate the memory-less feature of the Markov approach.

Utility values used in the model

The company's model uses the utility values derived from LUME-Lung 1[24] in the base-case. Utility values from the literature are also tested within the model. The company used utility values from a recently published paper by Chouaid et al[66] in a sensitivity analysis. This paper reports utilities recorded from relevant patients in Europe, Canada, Australia and Turkey as well as the UK and uses the EQ-5D to obtain utilities for the health states that were used in the company's model.

Table 34 Utilities used in the sensitivity analysis (Chouaid et al[66] 2013)

Health state	Mean (Standard error)
Progression free survival (PFS)	0.74 (0.03)
Post-progression	0.46 (0.08)

Source: Table 83 of the CS[1]

Disutility

The company's model also incorporated the impact of AEs on HRQoL; utility decrements associated with each AE were applied for a period of one model cycle. The company notes that the model may have double counted disutilities as some patients may experience multiple AEs simultaneously. Disutilities due to AEs are presented in Table 35.

Table 35 Disutilities associated with AEs

Adverse event	Disutility	Sources
ALT increased	-0.05	Assumption
Anaemia	-0.07	Nafees et al[67]
Diarrhoea - grade 2	-0.02	Assumption: half of the disutility for grade 3/4 diarrhoea
Diarrhoea - grade 3/4	-0.04	Data on file, Table 18.1[71]
Fatigue	-0.21	Data on file, Table 18.1[71]
Febrile neutropenia	-0.09	NICE TA192,[14] Nafees et al. 2008[67]
Infection	-0.05	Assumption
Liver-related investigations	-0.05	Assumption
Nausea and vomiting	-0.05	Nafees et al[67]
Neutropenia	-0.09	Nafees et al[67]
Neutrophil count decreased	-0.09	Assumption: same as disutility of neutropenia
Rash	-0.033	Nafees et al.[67]
Thrombocytopenia	-0.05	NICE TA181[72]
WBC count decreased	-0.05	Assumption

WBC=white blood cell
Source: Table 84 of the CS[1]

5.4.8 Resources and costs

Drug acquisition and administration costs

Table 36 presents a summary of the drug and IV administration costs per cycle for each comparator for the active second-line treatment phase, the BSC phase and, where relevant, the third-line treatment phase. Adjustments in drug costs due to change in dose intensity and treatment discontinuation as observed in LUME-Lung 1[24] were included in the company's model for second-line treatments. Changes in dose intensity or treatment discontinuation inputs only affected drug costs outcomes; they did not affect clinical outcomes (e.g. OS, PFS and AEs). Wastage was taken into account when calculating the cost of IV treatments.

As nintedanib is taken orally, it is not associated with any additional administration costs.

Table 36 Drug costs used in the company's model

Drug	Units per administration	Price per unit	Route	Administrations per cycle	Administration cost	Costs per cycle*
Nintedanib	400 mg	£0.18	Oral	21	-	£1,354
Docetaxel in combination with nintedanib	75 mg/m^2 400 mg	£5.68 £0.18	IV Oral	1 21	£155 -	£196 + £1,354= £1550
Docetaxel	75 mg/m^2	£5.68	IV	1	£155	£196
Erlotinib	150 mg	£0.36	Oral	21	-	£1,051
Carboplatin†	750 mg	£0.33	IV	1	£155	£250
Vinorelbine†	30 mg/m^2	£2.78	IV	3		£465

IV=intravenous
* Mean dose intensity taken into account: (nintedanib + docetaxel=98.1%, nintedanib=91.2%, docetaxel=98.7% and erlotinib=92%)
† third-line treatment
Source: Table 96 of the CS[1]

Health state costs

The company considered that there was little published literature exploring the detailed resource use commonly associated with NSCLC or other metastatic cancer. To estimate the treatment patterns in NSCLC a resource use questionnaire was constructed. This formed the basis of an interview with an oncologist who specialised in the treatment of patients with lung cancer and who had experience of working on NICE health technology assessment reports. A series of questions was posed separately for each different health state (stable on nintedanib plus docetaxel, stable on docetaxel, stable on erlotinib, stable on BSC; progressed on active treatment, progressed on BSC; and a one-off estimate of resource use at the time of progression) under the umbrella term 'monitoring'. Three main areas of resource use were considered: routine follow up (type and frequency of physician visit,

laboratory tests, radiological scans); treatment at time of progression (hospitalisations, physician visits, laboratory tests, radiological scans, procedures use; and resources used during BSC/palliative care (initial tests, procedures, hospitalisations, physician visits, laboratory tests, radiological scans and procedures). Detailed descriptions of resource use are displayed in Tables 98 to 105 in the CS;[1] in addition a full range of the unit costs employed is also presented in Table 106 of the CS.[1] The unit costs of visit procedures and laboratory tests were mainly derived from the National Schedule of Reference costs (2012/3),[69] whilst some visit costs were taken from the Personal Social Services Research Unit (PSSRU).[73]

Adverse events costs

A single UK consultant provided AE management costs. Estimates were generated via survey and face-to-face discussion. Costs for inpatient hospitalisations were taken from the NHS National Schedule of Reference Costs (2012/13).[69] Outpatient costs were taken from the same source[69] or from the PSSRU.[73] The cost of each AE is summarised in Table 37.

Table 37 Adverse events costs

Type of adverse event	Cost of adverse events
ALT increased	£587
Anaemia	£978
AST increased	£336
Diarrhoea - CTCAE grade 1 and 2	£250
Diarrhoea - CTCAE grade 3 and 4	£1796
Fatigue	£370
Febrile neutropenia	£2012
Infection	£2181
Nausea and vomiting	£1919
Neutropenia	£346
Rash	£639
Thrombocytopenia	£422
WBC count decreased	£423

ALT=alanine aminotransferase; AST=aspartate aminotransferase; CTCAE= Common Terminology Criteria for Adverse Events; WBC=white blood cell
Source: Table 107 of the CS[1]

5.4.9 Model validation

The company reports that a number of steps were taken to ensure that the analysis was validated, including:

- External review by a leading UK clinical expert to ensure that the model adheres to the clinical course of the disease and is reflective of current clinical practice
- Sensitivity analyses
- A senior modeller within the model developers' organisation (with no involvement in the model's development) performed a detailed quality assurance check on the model
- The company performed validation checks (varying parameter values and assumptions). This involved increasing and decreasing various parameters or changing assumptions in the model and then monitoring the impact on outputs. If the outputs were unexpected, further checks were made to determine whether this was the result of an error in the model.

5.4.10 Results included in the company's model

The incremental cost-effectiveness results generated by the company's economic model are presented in Table 38 and Table 39. The ICER for nintedanib plus docetaxel vs docetaxel is estimated by the company to be £50,677 per QALY gained. The ICER for nintedanib plus docetaxel vs erlotinib is estimated by the company to be £27,008.

Table 38 Company's base-case cost-effectiveness results: nintedanib plus docetaxel vs docetaxel

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£) vs baseline (QALYs)	ICER (£) incremental (QALYs)
Nintedanib + docetaxel	[redacted]	[redacted]	[redacted]	-	-	-	-	-
Docetaxel	[redacted]	[redacted]	[redacted]	£11,051	0.33	0.22	£50,776	£50,776

ICER=incremental cost-effectiveness ratio; LYG=life years gained; QALY=quality adjusted life year
N.B. Distributions used - OS: Log-logistic; PFS: Log-normal
Source: Table 129 of the CS[1]

Table 39 Company's secondary cost-effectiveness results: nintedanib plus docetaxel vs erlotinib

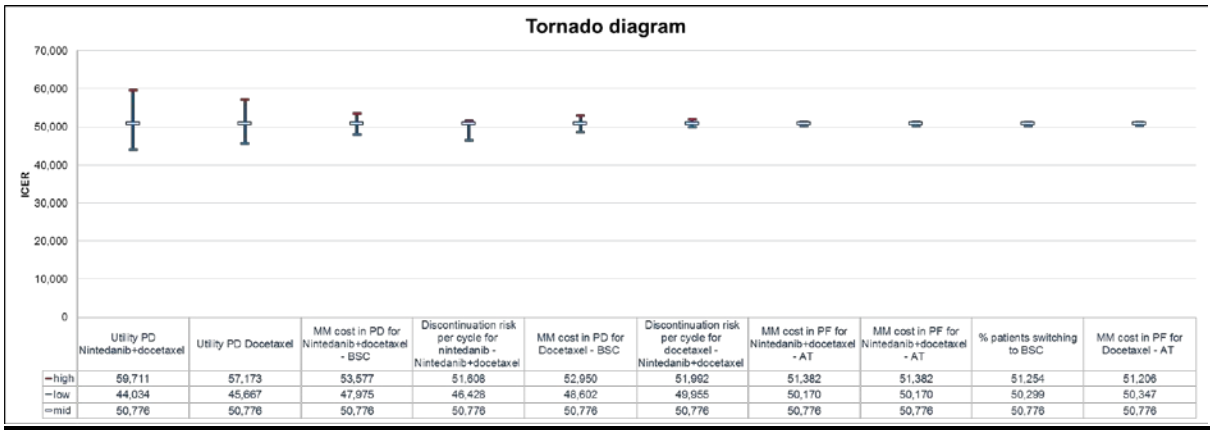
Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£) vs baseline (QALYs)	ICER (£) incremental (QALYs)
Nintedanib + docetaxel	[redacted]	[redacted]	[redacted]	-	-	-	-	-
Erlotinib	[redacted]	[redacted]	[redacted]	£7,571	0.43	0.28	£27,008	£27,008

ICER=incremental cost-effectiveness ratio; LYG=life years gained; QALY=quality adjusted life year
N.B. Distributions used - OS: Weibull distributions; PFS: Weibull survival
Source: Table 130 of the CS[1]

5.4.11 Sensitivity analyses

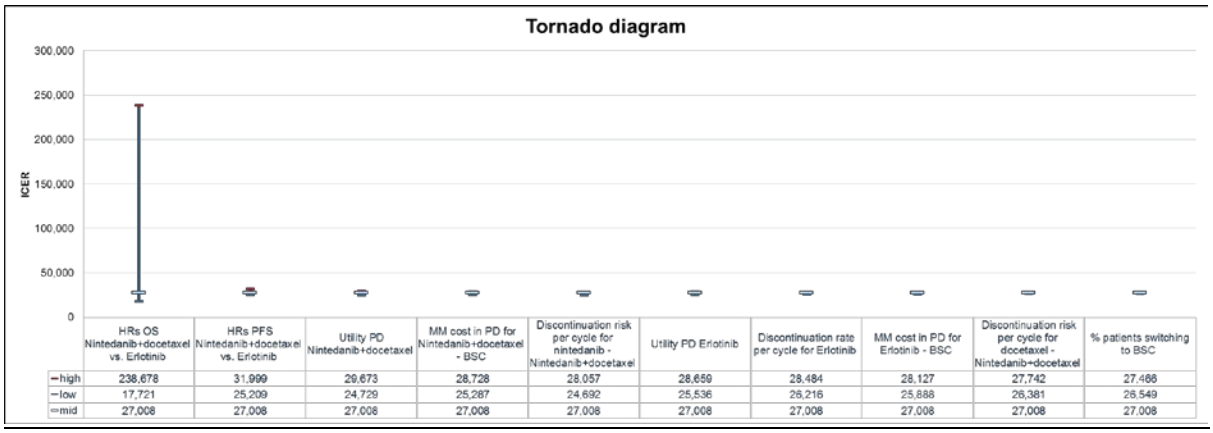
Deterministic sensitivity analyses

The company carried out a wide range of deterministic sensitivity analyses. Results for the ten parameters showing the greatest variability for the comparisons of nintedanib plus docetaxel vs docetaxel and vs erlotinib are shown in Figure 3 and Figure 4 respectively. For the comparison of nintedanib plus docetaxel vs docetaxel, the two most influential variables were univariate changes in utility values after progression for both intervention and comparator. For the comparison of nintedanib plus docetaxel vs erlotinib, the single most influential variable was the HR used for OS.



	Utility PD Nintedanib+docetaxel	Utility PD Docetaxel	MM cost in PD for Nintedanib+docetaxel - BSC	Discontinuation risk per cycle for nintedanib - Nintedanib+docetaxel	MM cost in PD for Docetaxel - BSC	Discontinuation risk per cycle for docetaxel - Nintedanib+docetaxel	MM cost in PF for Nintedanib+docetaxel - AT	MM cost in PF for Nintedanib+docetaxel - AT	% patients switching to BSC	MM cost in PF for Docetaxel - AT
high	59,711	57,173	53,577	51,606	52,950	51,992	51,382	51,382	51,254	51,206
low	44,034	45,667	47,975	46,428	48,602	49,955	50,170	50,170	50,299	50,347
mid	50,776	50,776	50,776	50,776	50,776	50,776	50,776	50,776	50,776	50,776

Figure 3 One-way sensitivity analysis: nintedanib plus docetaxel vs docetaxel
Source: Figure 33 of the CS[1]



	HRs OS Nintedanib+docetaxel vs. Erlotinib	HRs PFS Nintedanib+docetaxel vs. Erlotinib	Utility PD Nintedanib+docetaxel	MM cost in PD for Nintedanib+docetaxel - BSC	Discontinuation risk per cycle for nintedanib - Nintedanib+docetaxel	Utility PD Erlotinib	Discontinuation rate per cycle for Erlotinib	MM cost in PD for Erlotinib - BSC	Discontinuation risk per cycle for docetaxel - Nintedanib+docetaxel	% patients switching to BSC
high	238,678	31,999	29,673	28,728	28,057	28,659	28,484	28,127	27,742	27,466
low	17,721	25,209	24,729	25,287	24,692	25,536	26,216	25,888	26,381	26,549
mid	27,008	27,008	27,008	27,008	27,008	27,008	27,008	27,008	27,008	27,008

Figure 4 One-way sensitivity tornado diagram: nintedanib plus docetaxel vs erlotinib
Source: Figure 34 of the CS[1]

Probabilistic sensitivity analysis

The company undertook probabilistic sensitivity analysis (PSA) to derive the mean ICERs per QALY gained for nintedanib plus docetaxel vs docetaxel and vs erlotinib. PSA was carried out using 5000 iterations of the cost-effectiveness model.

The PSA result for nintedanib plus docetaxel vs docetaxel shows that nintedanib plus docetaxel has a 2% probability of being cost-effective at the £30,000 per QALY gained threshold and a 50% chance of being cost-effective at the £50,000 per QALY gained threshold. The cost-effectiveness plane and cost-effectiveness acceptability curve (CEAC) for this comparison are displayed in Figure 5 and Figure 6.

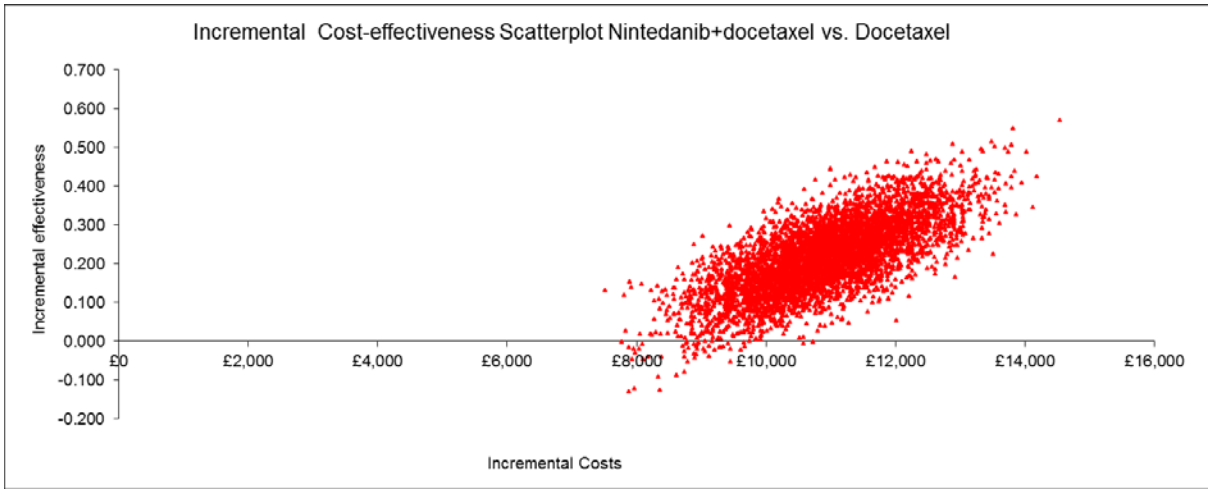


Figure 5 Incremental cost-effectiveness plane for nintedanib plus docetaxel vs docetaxel
Source: Figure 35 of the CS[1]

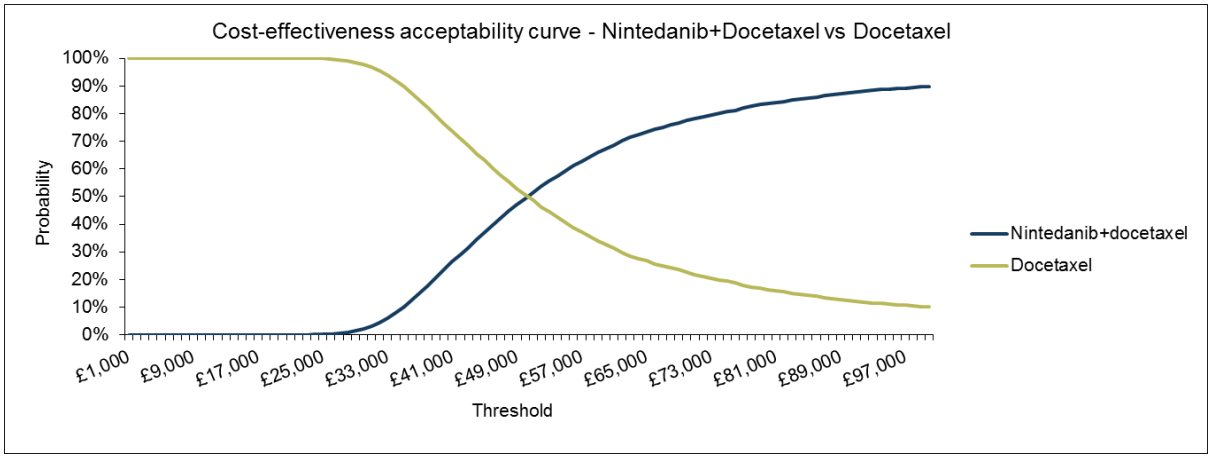


Figure 6 Cost-effectiveness acceptability curve for nintedanib plus docetaxel vs docetaxel
Source: Figure 36 of the CS[1]

The PSA result for nintedanib plus docetaxel vs erlotinib shows that nintedanib plus docetaxel has a 65% probability of being cost-effective at the £30,000 per QALY gained threshold and a 94% chance of being cost-effective at the £50,000 per QALY gained threshold. The cost-effectiveness plane and CEAC for this comparison are displayed in Figure 7 and Figure 8.

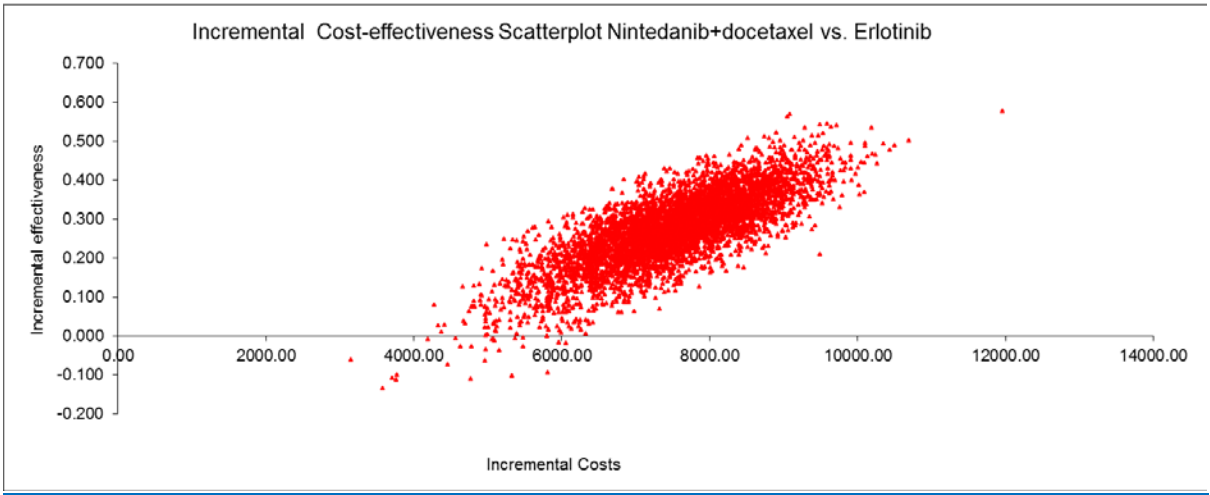


Figure 7 Cost-effectiveness plane for nintedanib plus docetaxel vs erlotinib
Source: Figure 37 of the CS[1]

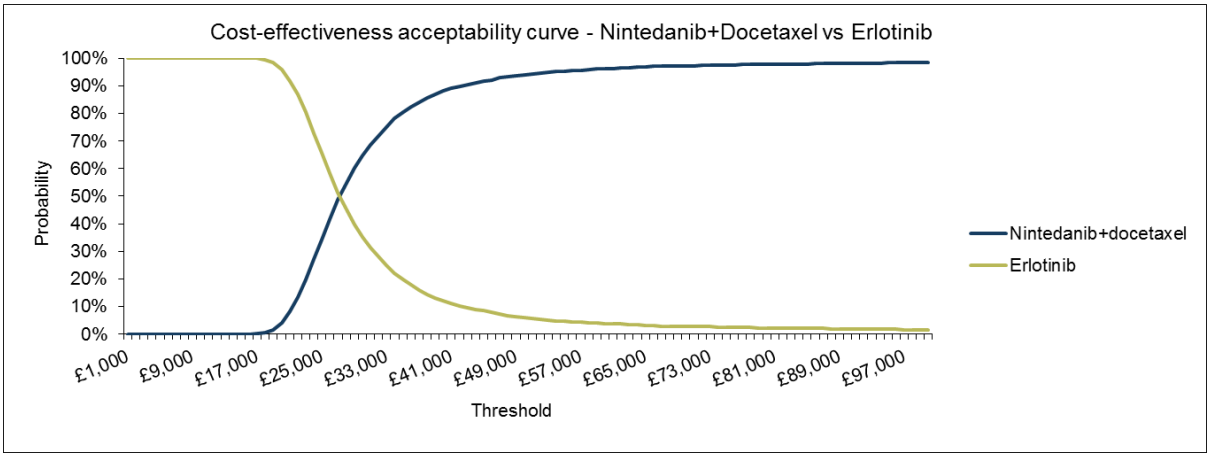


Figure 8 Cost-effectiveness acceptability curve for nintedanib plus docetaxel vs erlotinib
Source: Figure 38 of the CS[1]

Scenario analyses

The company also undertook a series of scenario analyses and explored how varying scenarios relating to survival modelling, indirect comparisons, resource use, utility, time horizon and discount rate might affect the results of the economic evaluation. The results of these scenario analyses are displayed in Tables 135 to 140 in the CS.[1] The company concluded that the base-case ICERs are mainly sensitive to changes in the PFS and OS HRs as well as the costs and utilities associated with the post-progression states.

5.5 Detailed critique of company's economic model

The model submitted by the company for this appraisal is structured as a partitioned survival spreadsheet model following a structure broadly similar to those used in similar appraisals.[13,14] For most functions the assumptions and options are labelled and annotated where necessary. However, in some cases, the ERG has found it difficult to confirm details of the data sources employed.

In line with the issues previously discussed (section 2.2 and 3.3) concerning the relevance of erlotinib as a comparator (largely due to the challenge of identifying a meaningful population for such a comparison), and the unreliability of indirect evidence of relative efficacy (section 4.4.1 and 4.4.6), this critique is primarily focussed on the direct comparison between nintedanib plus docetaxel and docetaxel monotherapy in the adenocarcinoma subgroup of LUME-Lung 1.[24]

A particular concern of the ERG relates to the analyses reported by the company of OS data from the SEER and LUCADA registers (Appendix 13 of CS[1]). No references were provided which identify the specific data sets employed and relevant details (such as date of extraction, selection criteria, duration of follow-up) are missing. The ERG has had to infer from the text that the SEER results appear to relate to all-cause mortality from the date of Stage 4 diagnosis and that the LUCADA data relate to second-line chemotherapy, but without any specific indication of prior treatments, PS and/or other relevant characteristics. The value of these analyses to support the company's chosen parametric survival modelling is therefore difficult to assess, and in particular the relevance of the SEER dataset to the population recruited to LUME-Lung 1[24] must be considered weak.

The following sections detail eleven specific issues identified by the ERG involving errors in data analysis, parameter values or methodology which have been identified in the submitted model, together with estimates made by the ERG of the impact of correcting these problems on the estimated ICER for nintedanib plus docetaxel compared with docetaxel. Within the time available to the ERG, it has not been possible to be certain that other problems do not remain undetected in the company's model.

5.5.1 Methods used to project time-to-event outcomes

In seeking to project OS and PFS data from LUME-Lung 1[24] to represent expected lifetime experience, the company has followed a convention of seeking to fit a variety of standard parametric functions to the available data, and employed the derived functions in place of the trial data throughout their decision model.

The ERG considers that this approach to model calibration to be flawed on several counts:

- The primary purpose of curve-fitting is to anticipate what is likely to happen to the minority of patients who remain at risk (i.e. alive with or without disease progression or remaining on treatment) at the time of data cut. However, the great majority of data events used for this purpose are drawn from patients who are unlike those remaining at risk at the time of data cut, since that majority were at greater propensity to fail (i.e. die, progress or cease treatment) than those still remaining. This is an example of bias against survivors and frequently results in the fitting of inappropriate functions and misleading projection estimates.

- The methods used for fitting parametric functions to a survival data set are essentially descriptive and lack any external validity based on the appropriateness of an underlying disease/treatment process governing them. Therefore, the analyst may be content in having achieved a reasonable correspondence to the available data, but lacks any basis for confidence in any future projection based thereon.

- When a single clinical trial is the primary source for cost-effectiveness assessment, it is important to make the maximum direct use of the available evidence. Replacing a large part of the trial results with a fitted model adds additional uncertainty from imposed modelling assumptions to the unavoidable data sampling uncertainty, so that rather than clarifying the underlying disease dynamic, it only serves to obscure it.

- Most of the 'standard' statistical functions used by the company to model survival lack any logical or empirical basis for representing a biological phenomenon, being only selected for their analytical convenience.

As part of the clarification process, the ERG requested detailed K-M survival analysis results for all of the time-to-event trial data employed in the company's model. The ERG has, for each of the K-M survival analysis results, identified a projective model, using only those data in the period towards the end of the survival curve in which it is apparent that a long-term trend has become established. The early K-M data are used directly in the company's

model, giving way to the projective model only to represent the segment of patient experience which cannot be reliably estimated otherwise.

In projecting ToT the company's model considers only a single parametric function (exponential model with fixed hazard per cycle calibrated over the whole trial period). Here, the same methodology flaws are present, except that no attempt has been made to assess the comparative validity of the exponential hazard function against possible alternatives.

5.5.2 Overall survival estimation

The company's model base-case comparison of nintedanib plus docetaxel vs docetaxel indicates a gain in (undiscounted) overall survival of 4.7 months; only 15% (0.7 months) of this gain is attributed to the pre-progression phase. This is unusual in locally advanced and metastatic cancers where treatment benefit is largely confined to the active treatment period (i.e. PFS). In order to validate this claim, the ERG has carried out its own analysis of the OS and post-progression survival (PPS) trial data, based on K-M results provided by the company in response to a clarification request.

Figure 9 shows a cumulative hazard chart for OS. After about 300 days, a simple linear trend is established in both trial arms and continues indefinitely. This indicates that in both arms OS can be estimated by use of a simple exponential projective model (i.e. there is a constant hazard irrespective of time). Comparing the slopes of the trend lines allows a long-term HR of 0.83 in favour of nintedanib plus docetaxel to be estimated. To verify this finding a similar cumulative hazard chart was prepared for PPS (shown as Figure 10). This confirms that patients in LUME-Lung 1[24] who survived a disease progression event continued to gain survival benefit from treatment with nintedanib plus docetaxel compared with those receiving only docetaxel. Long-term linear trends are apparent in both trial arms beyond 200 days in PPS, and the trends continue to diverge with an estimated long-term HR of 0.79 in favour of nintedanib plus docetaxel.

Estimates of lifetime OS were obtained by the ERG by applying the K-M trial results directly using the area under curve (AUC) method until the long-term OS trends were established and then projecting remaining estimated survival using the exponential trends (as shown in Figure 11). Mean OS in the docetaxel arm is estimated as 453.0 days (14.9 months) compared with 545.7 days (17.9 months) in the nintedanib plus docetaxel arm, a net survival gain of 92.7 days (3.05 months) attributable to the addition of nintedanib to docetaxel. This result is considerably lower than the OS gain obtained from the company's model (4.7 months), and indicates the effect of replacing the company's preferred Log-Logistic survival model to represent the whole trial data set with the ERG's approach (direct use of

unadjusted trial data for the majority of patients, followed by projecting long-term survivors using trends evident in the trial data set).

Replacing the company's preferred OS model with the ERG's approach has a major impact on the cost-effectiveness of nintedanib plus docetaxel compared with docetaxel alone. The incremental discounted cost per patient is reduced by [redacted] while the incremental discounted QALY gain is reduced by [redacted], resulting in the estimated ICER increasing from £50,776 per QALY gained to £68,587 per QALY gained.

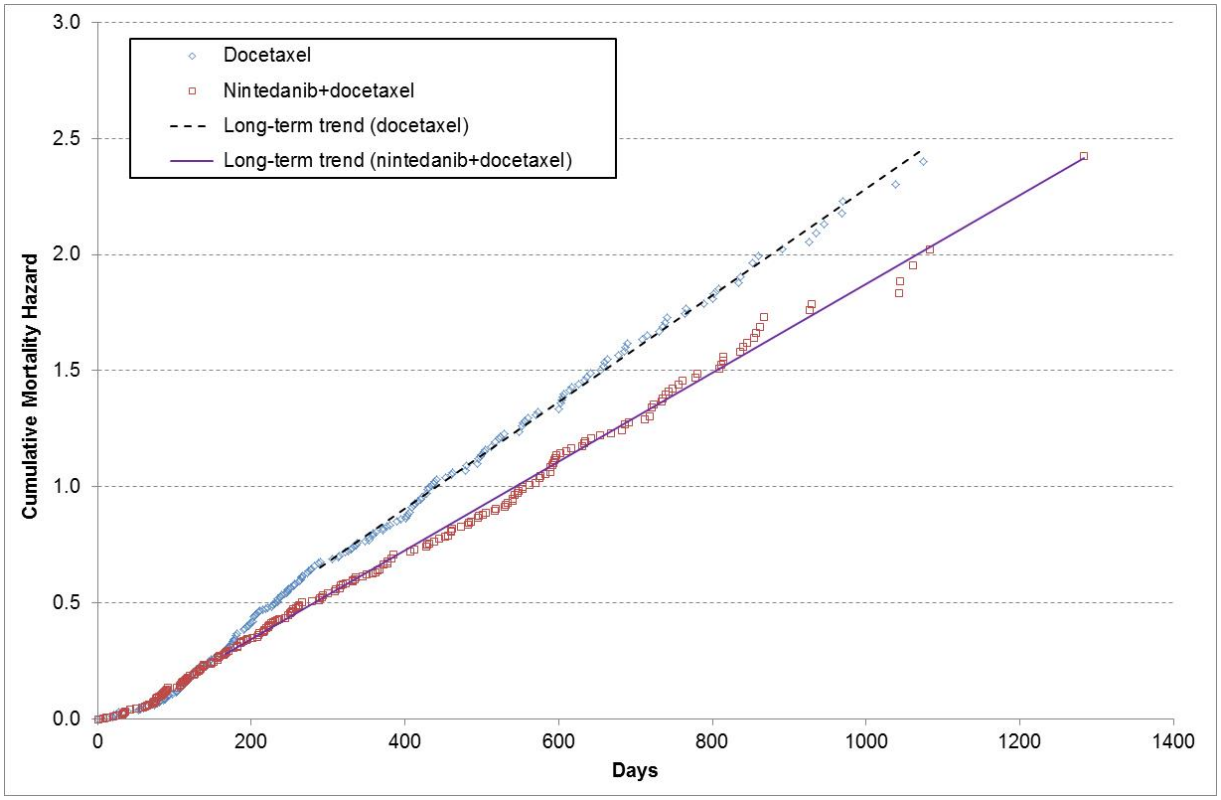


Figure 9 Cumulative OS hazard plot for nintedanib plus docetaxel vs docetaxel

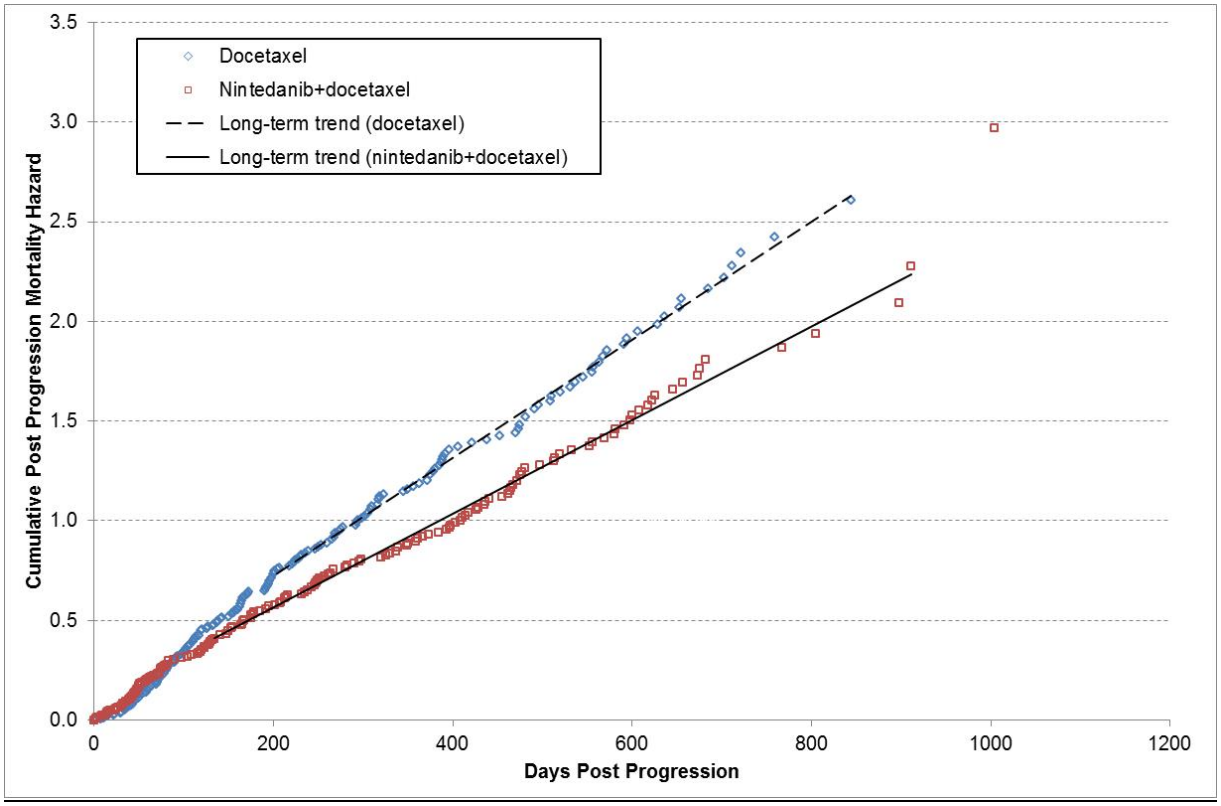


Figure 10 Cumulative PPS hazard plot for nintedanib plus docetaxel vs docetaxel

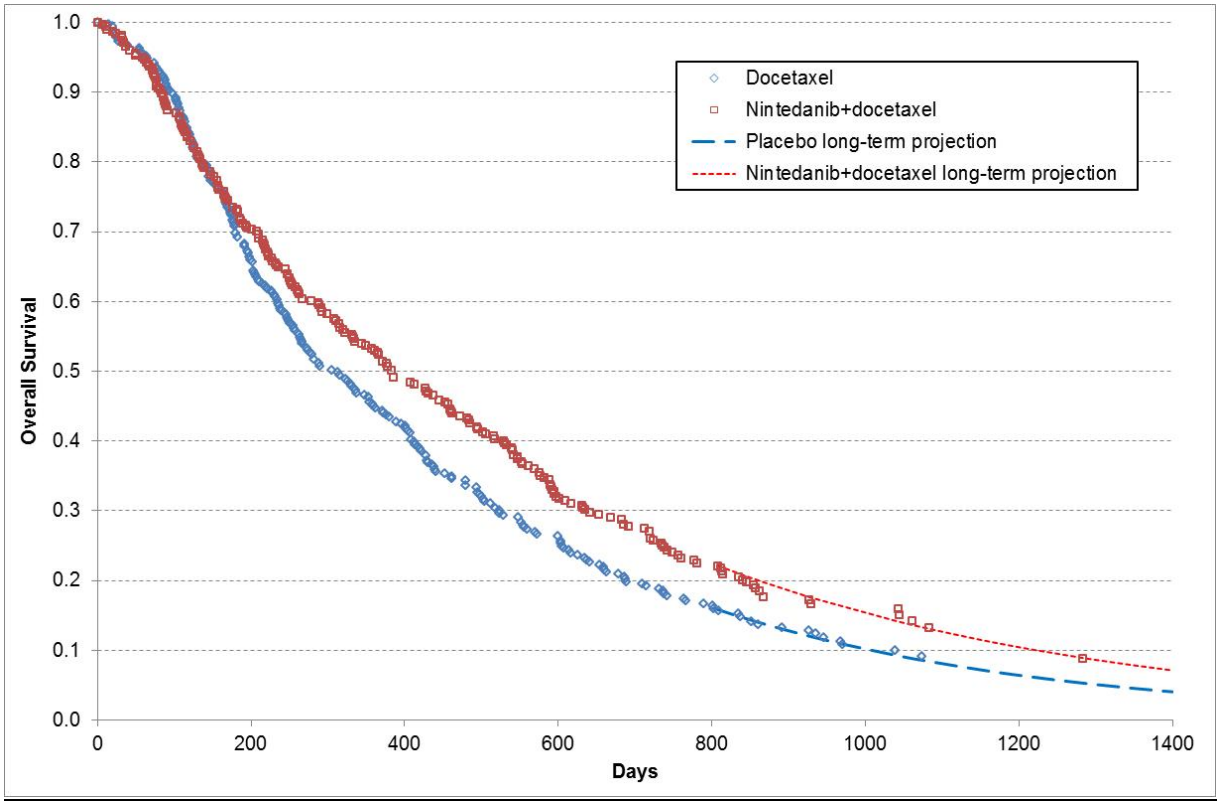


Figure 11 OS plot with ERG long-term projections for nintedanib plus docetaxel vs docetaxel

5.5.3 Progression-free survival estimation

The company's model base-case comparison of nintedanib plus docetaxel vs docetaxel indicates a gain in (undiscounted) PFS of 28.6 days, based on calibrating a Log-Normal hazard distribution to each trial arm and applying these to represent patient experience until all patients have died or suffered disease progression.

Examination of the PFS temporal profile (Figure 12) indicates that although the addition of nintedanib to docetaxel therapy generates a short-term delay in disease progression for some patients (i.e. the PFS curves begin to separate), subsequently this advantage progressively dissipates until the PFS experience of patients in the two trial arms is indistinguishable. Here, the extent of advantage in mean PFS can be readily estimated directly from the K-M analysis results by comparing the AUC estimates up to the point when the curves converge. The ERG identified that convergence occurred at day 375, and the difference in AUC at this time is 36.4 days. This suggests a small additional PFS benefit compared with the gain obtained in the company's model (28.6 days).

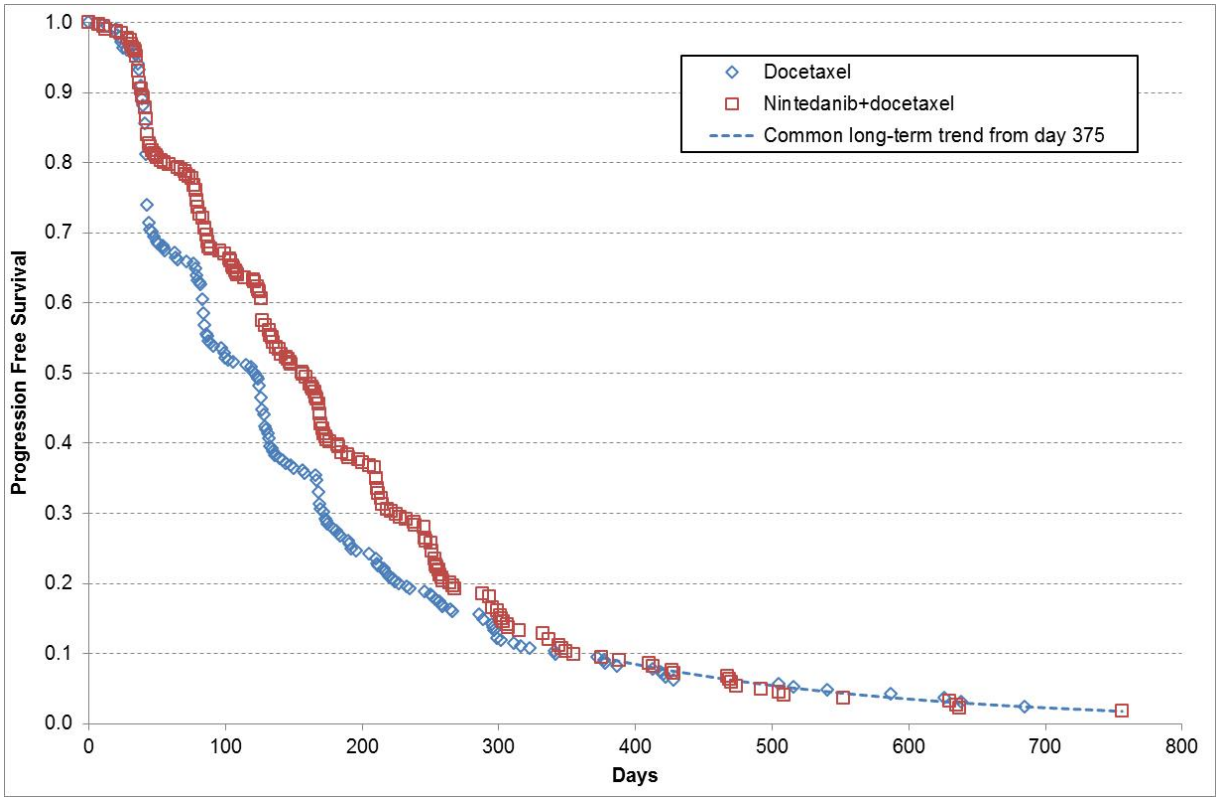


Figure 12 PFS plot with ERG common long-term projection for nintedanib plus docetaxel vs docetaxel

In order to apply the results of this re-analysis to the company's model, the ERG carried out a K-M landmark analysis for patients who were still progression-free at day 375. This indicated that a common long-term exponential model is appropriate for use in both

treatment arms from day 375 onwards, and this is shown in Figure 12. However, it should be noted that any projective model could be employed to both arms of the trial without any effect on the cost-effectiveness analysis as the incremental gain in PFS is unaffected.

Replacing the company's preferred PFS model with the ERG's approach has a modest impact on the cost-effectiveness of nintedanib plus docetaxel compared with docetaxel. The incremental discounted cost per patient is increased by [redacted] while the incremental discounted QALY gain is increased by [redacted], resulting in the estimated ICER increasing from £50,776 per QALY gained to £52,445 per QALY gained.

5.5.4 Time on treatment estimation

The ERG has used the same approach to obtain an accurate representation of the duration of treatments in the arms of LUME-Lung 1.[24] This approach uses the K-M results directly until a long-term exponential trend is established for projection until all patients have died (shown in Figure 13 to Figure 15).

Replacing the company's preferred exponential model with the ERG's approach has a modest impact on the cost-effectiveness of nintedanib plus docetaxel compared with docetaxel. The discounted cost per patient is increased in both treatment arms, so that the incremental cost per patient rises by [redacted], resulting in the estimated ICER increasing from £50,776 per QALY gained to £51,930 per QALY gained.

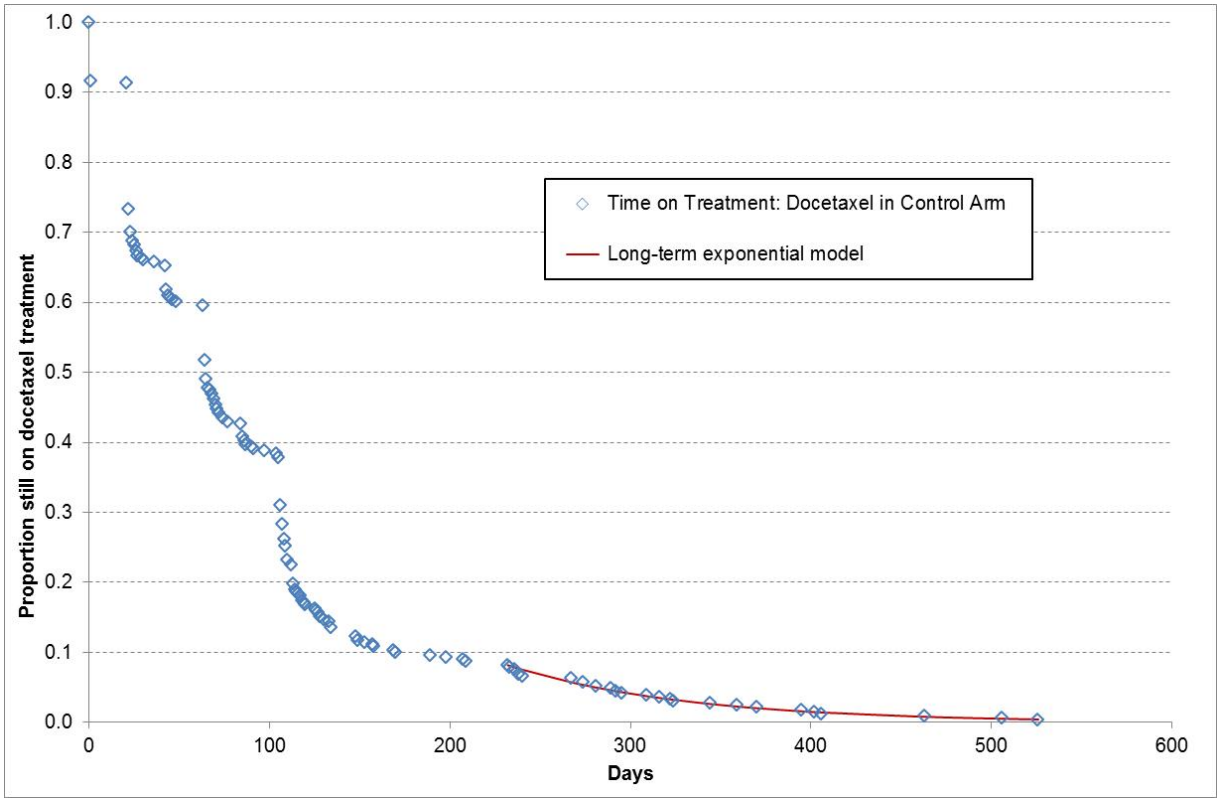


Figure 13 Time on Treatment: docetaxel in control arm with ERG long-term projection

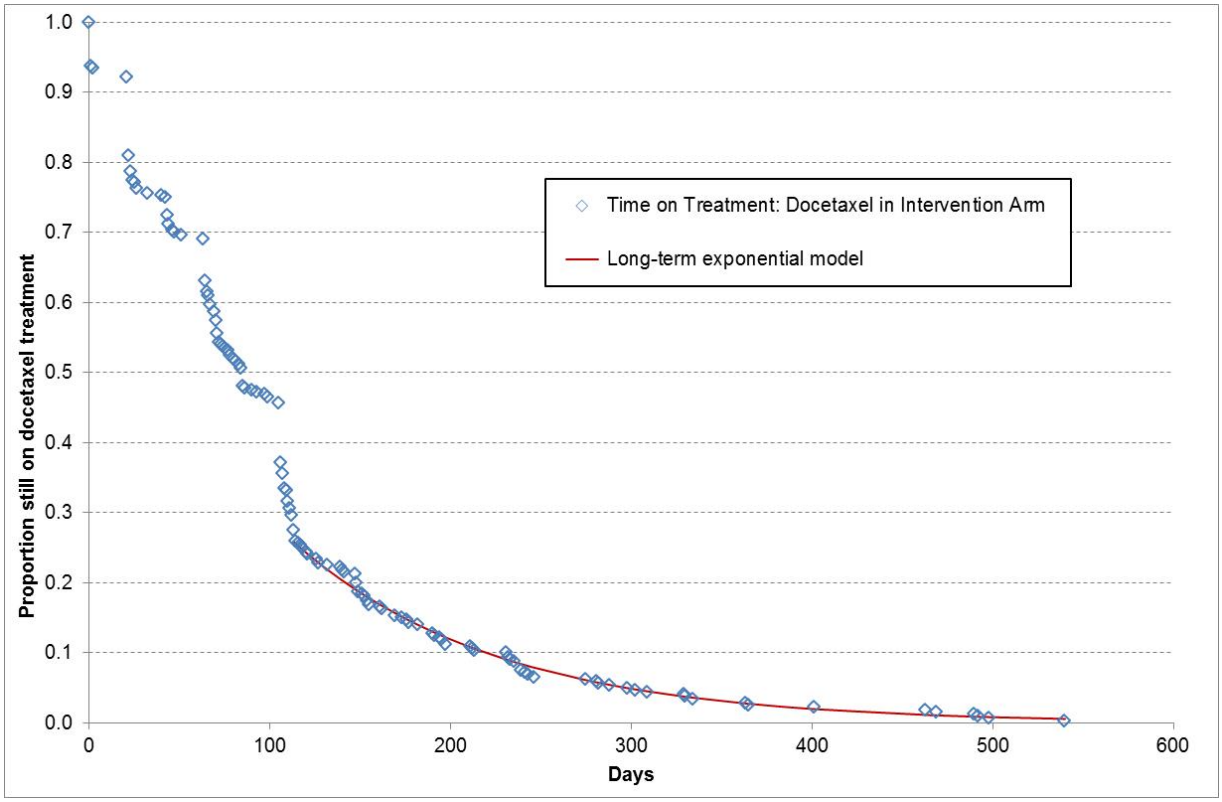


Figure 14 Time on Treatment: docetaxel in intervention arm with ERG long-term projection

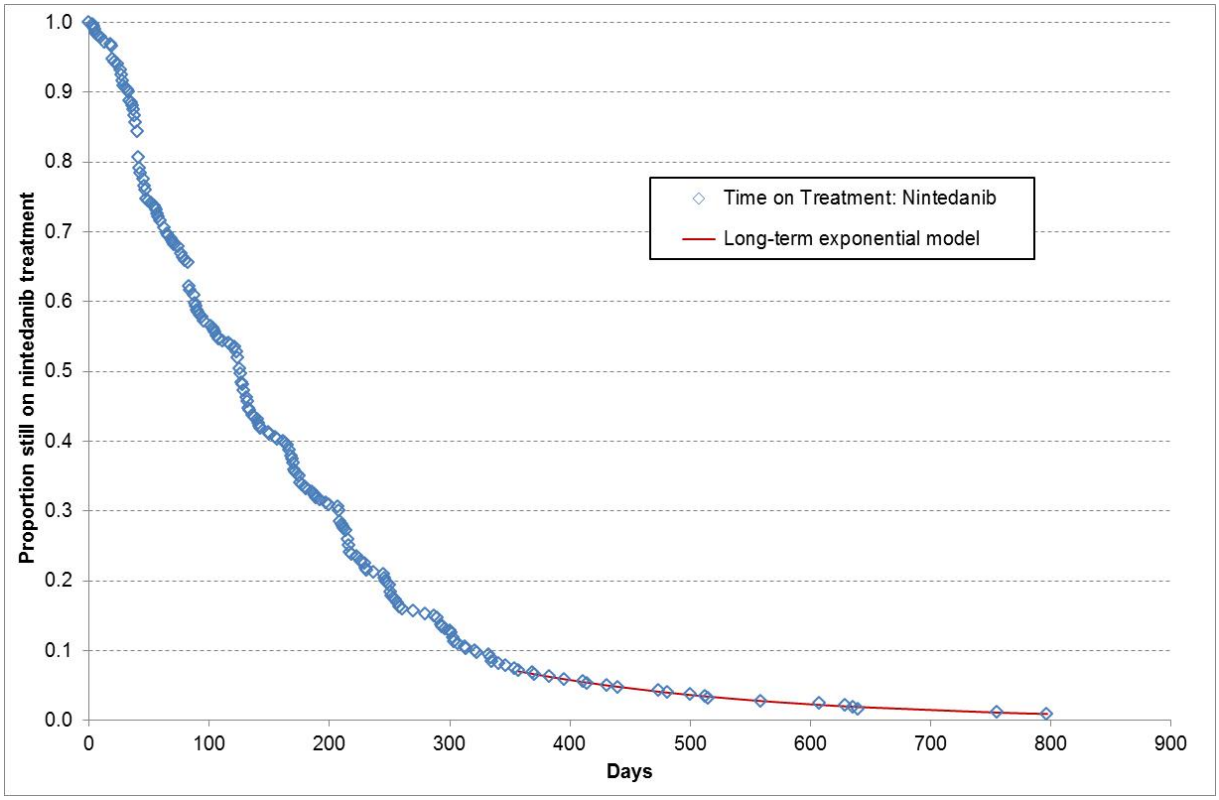


Figure 15 Time on Treatment: nintedanib in intervention arm with ERG long-term projection

5.5.5 Incorrect mid-cycle adjustment for drug costs

In the company's model the costs of both docetaxel and nintedanib are calculated for the average number of patients on treatment across each cycle. This mid-cycle adjustment for docetaxel is not accurate since three-weekly docetaxel is delivered on the first day of each cycle. Clinical advice also indicates that nintedanib doses are also dispensed on the first day of each cycle. The effect of this error is to under-estimate the quantity and cost of drugs used throughout the trial and in both arms of the comparison.

When this error is remedied the incremental discounted cost per patient increases by [redacted], and the estimated ICER increases from £50,776 per QALY gained to £53,839 per QALY gained.

5.5.6 Calculations for drug costs per dose

The average cost per dose of docetaxel delivered has been calculated by the company according to the distribution of body surface area (BSA) within the relevant UK population as a whole, though neglecting the important distinction between males and females whose mean BSA differs sufficiently to affect the overall estimated cost per dose. In addition, only the cost of the full 75mg/m^2 dose is estimated and adjusted using a relative dose intensity (RDI) index from trial data. It is more accurate to estimate the cost of a reduced dose (60mg/m^2) and then create a weighted average cost based on the balance between full and reduced doses recorded in the trial. The ERG has therefore re-estimated the overall average cost per dose of docetaxel using separate male and female subgroups, and also re-estimated the RDI multiplier to match the balance of full and reduced doses.

In addition, the ERG received clinical advice from a centre currently using nintedanib indicating that in practice nintedanib tablets are dispensed to patients at the time of docetaxel administration in blister packs sufficient to self-treat until the date of the next docetaxel dose (i.e. for days 2 to 21 of each cycle). Any missed doses are unlikely to alter the dispensing pattern, and thus missed doses will not alter the amount and cost of product dispensed. Therefore a reduction in cost through a RDI index is inappropriate. The company's method of calculating the cost per dose of nintedanib does not take account of the effect of three separate doses used (full dose, and two reduced doses) when part packs are dispensed as required at each cycle visit. Using data from LUME-Lung 1[24] of the differing balance between dose levels at each cycle, it has been possible to estimate an overall mean cost of treatment with nintedanib per cycle.

Applying these revised ERG parameter values to the company's base-case model, results in a [redacted] increase in the incremental cost per patient, and raises the estimated ICER from £50,776 per QALY gained to £52,587 per QALY gained.

5.5.7 Cost of treating febrile neutropenia

The company's model includes an estimated cost of treatment for grade 3/4 febrile neutropenia of £2,012.10 per patient affected, based on clinical advice. This figure is substantially lower than the average cost estimated by the NICE Decision Support Unit in 2007[74] which was revised for the recent MTA of second-line chemotherapy in NSCLC.[75] The ERG further updated the DSU estimate using National Reference costs[69] for 2012/13, to a mean cost per episode of £5,240.40 and mean cost per patient of £7,352.54 (assuming 1.4 episodes per patient).

Using these revised cost estimates in the company's model increases the incremental cost of nintedanib plus docetaxel vs docetaxel by £130 per patient, and raises the base-case ICER from £50,776 per QALY gained to £51,372 per QALY gained.

5.5.8 Monitoring cost

In the company's model the ERG has observed that there is a discrepancy between the cost of disease monitoring in patients who are on active treatment but who have not yet suffered disease progression (i.e. patients with stable disease). The model assigns a cost of £188 per cycle to patients in the nintedanib plus docetaxel arm and assigns a value of £205 per cycle to patients in the docetaxel arm, when the only difference in treatment relates to self-administered nintedanib tablets. On examination, it appears that the advice given by the company's clinical expert, concerning additional physician monitoring every 2 to 3 months for patients who have completed active treatment but who have not yet suffered disease progression, has been wrongly applied to patients still on active treatment with docetaxel. Moreover the unit cost employed is erroneously that of a GP consultation not an oncology out-patient visit.

When this misallocation is corrected, the incremental cost per QALY gained for nintedanib plus docetaxel vs docetaxel increases by £364, and the base-case ICER increases from £50,776 per QALY gained to £51,140 per QALY gained.

5.5.9 Discounting method

The submitted model applies discounting at a different rate for every 3-week model cycle based on the time elapsed. By convention in the UK, in line with the use of annual public sector budgets, discounting is applied annually considering the first 12 month period as involving current costs and each subsequent 12 month period requiring discounting for an additional year's delay. In some models with extended survival and multiple future events the choice of discounting method may have a large impact on the modelled ICER. However, using annual discounting in the company's model for this appraisal has only a minor effect, reducing the estimated base-case ICER from £50,776 per QALY gained to £50,532 per QALY gained.

5.5.10 Disutility of fatigue related adverse events

The key AEs identified from LUME-Lung1[24] were CTCAE grade 3 or 4 diarrhoea and fatigue. The company's analysis of EQ-5D utility data indicates that the estimated disutility for diarrhoea is low (-0.04). By contrast CTCAE grade 3 or 4 fatigue appears to have the largest effect in terms of patient disutility, amounting to an average of -0.21 across both treatment arms. However, Table 24 of the company's submitted Health Economics report[71] indicates a

large statistically significant difference between effect sizes in the two treatment arms: -0.326 for nintedanib plus docetaxel vs -0.101 for docetaxel, suggesting that patients experiencing serious fatigue on treatment are more seriously affected by the combination therapy. The company's model uses the overall average disutility estimate for both regimens. The ERG has applied a model amendment to apply the separate disutility values, resulting in a small reduction in the incremental QALY gain for nintedanib plus docetaxel vs docetaxel, and a corresponding increase of £54 per QALY gained in the base-case estimated ICER (from £50,776 per QALY gained to £50,830 per QALY gained).

5.5.11 Specification of second-line stable disease costs

Details of health care costs incurred by patients in various health states were derived from evidence provided by a panel of clinical advisors. A summary of this evidence is included in the appendices document accompanying the CS[1] (pages 70 to 77). A comparison between the details shown in the advisors evidence and the calculations used in the model to estimate average costs reveal important differences with respect to the cost of care for patients who have ceased active treatment and remain in a stable condition without evidence of further disease progression. The submitted model includes an assumption that these patients will require an hour of palliative nursing care every week and a bone scan every 3 weeks. This is in addition to a chest X-ray every 2 to 3 months and a physician visit once a year. The evidence of the clinical advisors only refers to the latter two items, and it appears that the palliative care and bone scans are included in error. Correcting this error substantially reduces the care costs per patient for any patient in a stable condition after second-line treatment. This has the effect of increasing the incremental cost per patient by [redacted] and increasing the estimated ICER for nintedanib plus docetaxel vs docetaxel from £50,776 per QALY gained to £53,470 per QALY gained.

5.5.12 Duration of docetaxel treatment

The company's base-case model follows the protocol of the LUME-Lung1[24] trial in permitting unlimited continuation of docetaxel treatment in either trial arm. One patient in the nintedanib plus docetaxel arm received 45 cycles of docetaxel, and one patient in the docetaxel monotherapy arm received 42 cycles. In the UK, standard clinical practice is to limit docetaxel to a maximum of four cycles per patient to avoid unacceptable AEs and associated poor QoL. The company's model includes an option to restrict docetaxel therapy to a maximum of four cycles. However, a formula error has been detected in the company's model which implements a limit of five rather than four cycles. The ERG has applied its own model adjustment which limits treatment to four cycles. It should be noted that this feature only affects the cost of drug acquisition and administration; it does not address the issue of

whether limiting exposure to docetaxel will impact on the prognosis of patients, nor does it attempt to adjust for consequent changes in AEs and the resulting cost and QoL effects.

This modification to the company's model reduces the base-case incremental cost per patient by [redacted], and reduces the estimated ICER for nintedanib plus docetaxel vs docetaxel from £50,776 per QALY gained to £48,060 per QALY gained.

5.5.13 Comparison with erlotinib

As noted in sections 2.2, 3.3 and 4.4.1 above, the ERG does not consider a comparison of nintedanib plus docetaxel to erlotinib is appropriate to the decision problem, a view also shared by the company. Nevertheless the company has attempted to incorporate into their model a facility to compare the relative cost-effectiveness of erlotinib and nintedanib plus docetaxel, as indicated in the NICE scope. In the absence of a trial directly comparing these regimes, it was necessary to attempt an MTC to generate estimated outcomes for patients treated with erlotinib, consistent with all relevant information in related studies. The base-case MTC includes three RCTs[56,59,62] in addition to the LUME-Lung1[24] trial: JMEI[56] which compared docetaxel with pemetrexed, WSY001[62] which compared pemetrexed with erlotinib and TAILOR[59] which compared docetaxel with erlotinib (see also Figure 1, page 39). This provides two connection pathways linking nintedanib plus docetaxel to erlotinib:

1) LUME-Lung1[24] ⇒ JMEI[56] ⇒ WSY001[62]
2) LUME-Lung1[24] ⇒ TAILOR[59]

In principle, it is desirable to employ this network to generate HRs for each time-to-event outcome as a basis for estimating the corresponding survival profiles for erlotinib, consistent with that obtained for nintedanib plus docetaxel in the LUME-Lung1[24] trial.

Time on Treatment

Employing a network may be possible for OS and PFS, but is not feasible for ToT of erlotinib, since none of the connecting trial reports (for JMEI,[56] WSY001[62] and TAILOR[59]) report results for this outcome. Instead the company has assumed that a simple exponential function is appropriate for ToT in all treatments and have calibrated this function for each trial based on an estimated mean number of treatments per patient. It has already been demonstrated in section 5.5.4 that such an assumption is not correct in the case of the LUME-Lung1[24] trial and there is no reason to believe that it would be any more successful in the other trials in the network. In particular, the company modellers have assigned a parameter value for erlotinib consistent with a mean number of erlotinib cycles (i.e. 28 days) taken from the ERG report for NICE assessment TA162,[19] without recognising that this

figure was obtained indirectly from PFS trial data (which may overstate ToT) and that the ERG on that occasion employed a 2-phase exponential model with a high risk of discontinuation in the first 11 weeks, and a lower risk thereafter. Without access to detailed patient-level ToT data for each of the studies in the MTC, it is not possible to rectify the substantial uncertainty associated with the estimation of drug acquisition costs in the company model.

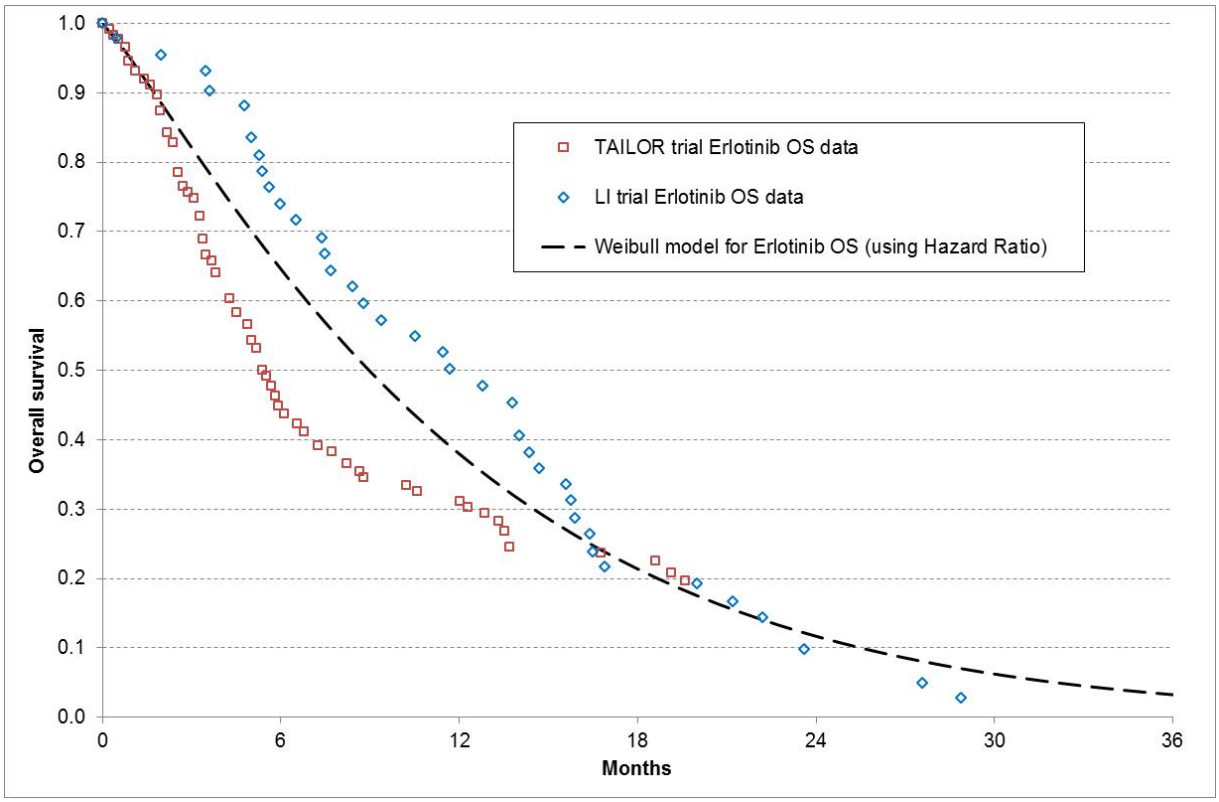
Overall survival and progression-free survival

Meta-analysis of time-to-event data in a network relies on a number of conditions being met:

- Within each trial the assumption of proportional hazards should apply
- Between trials featuring the same treatment at nodes in the MTC, treatment outcomes should be equivalent (i.e. both proportional hazards and very similar outcomes at all time points)
- If a parametric survival function is to be propagated through the network then it should be inherently proportional hazard compliant (i.e. Weibull or Exponential)

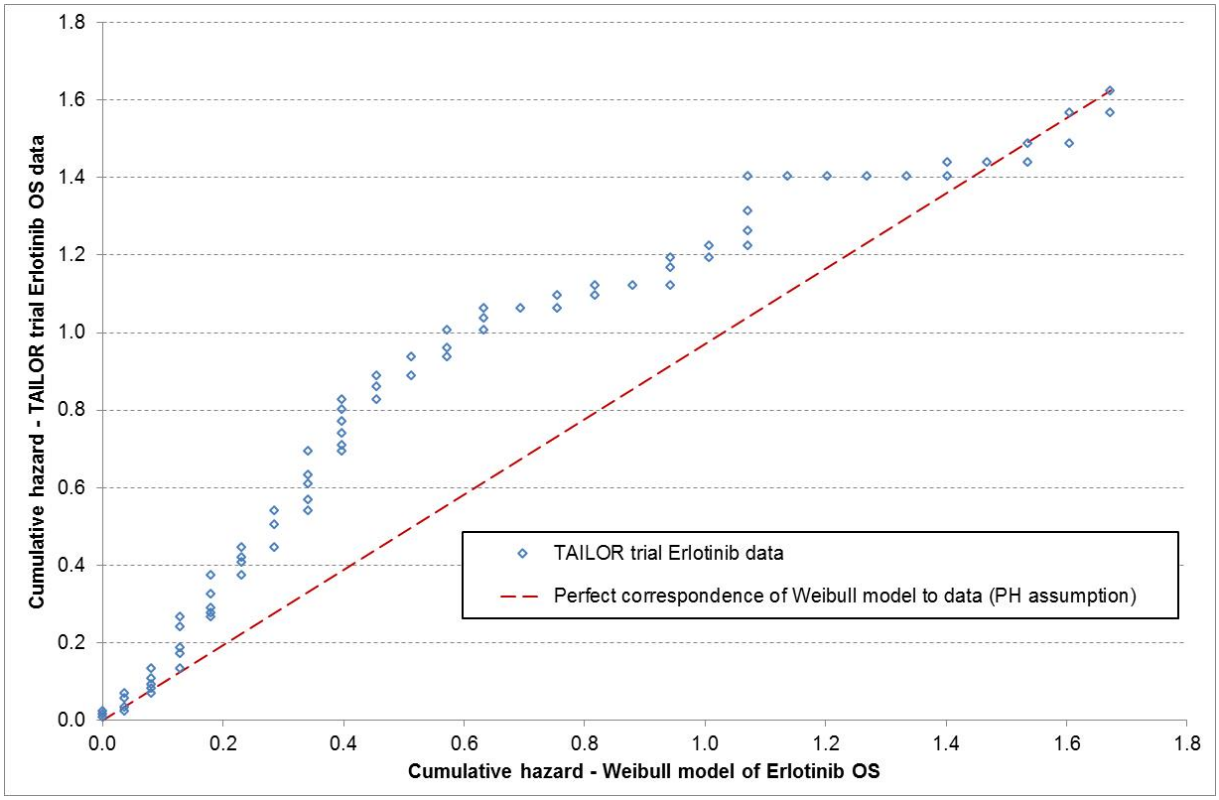
For the company's MTC of OS, a Weibull formulation was therefore used, despite this not appearing to give the best match to the nintedanib plus docetaxel LUME-Lung1[24] trial OS data. If all the above criteria are met, the resulting time-to-death profile should be a Weibull curve adjusted by an overall HR (0.64 for nintedanib plus docetaxel vs erlotinib in OS) so that it is consistent with the corresponding profiles for erlotinib in both the TAILOR[59] and WSY001[62] trials.

Figure 16 compares the fitted Weibull model for erlotinib with the erlotinib Kaplan-Meier data from the TAILOR[59] and WSY001[62] trials. It is apparent that during the first 18 months there are large differences between the three profiles. It is also possible to test whether the proportional hazards assumption is violated in both arms of the network. Figures 17 and 18 show plots of cumulative hazard data from each erlotinib trial arm against the cumulative hazard at the same time points from the Weibull OS model. The proportional hazards assumption is confirmed if the data points (corresponding to trial events) all lie close to and evenly spaced around the diagonal 'proportionality' line. It is clear that for both the erlotinib trials (TAILOR[59] and WSY001[62]) the proportional hazards assumption is seriously violated. This is likely to have been caused by multiple problems, including non-proportional hazards results in LUME-Lung1[24] trial OS data (as discussed in Appendix 7), proportional hazards violations in one or more of the other three trials in the MTC and non-equivalence of trial arms at network nodes.



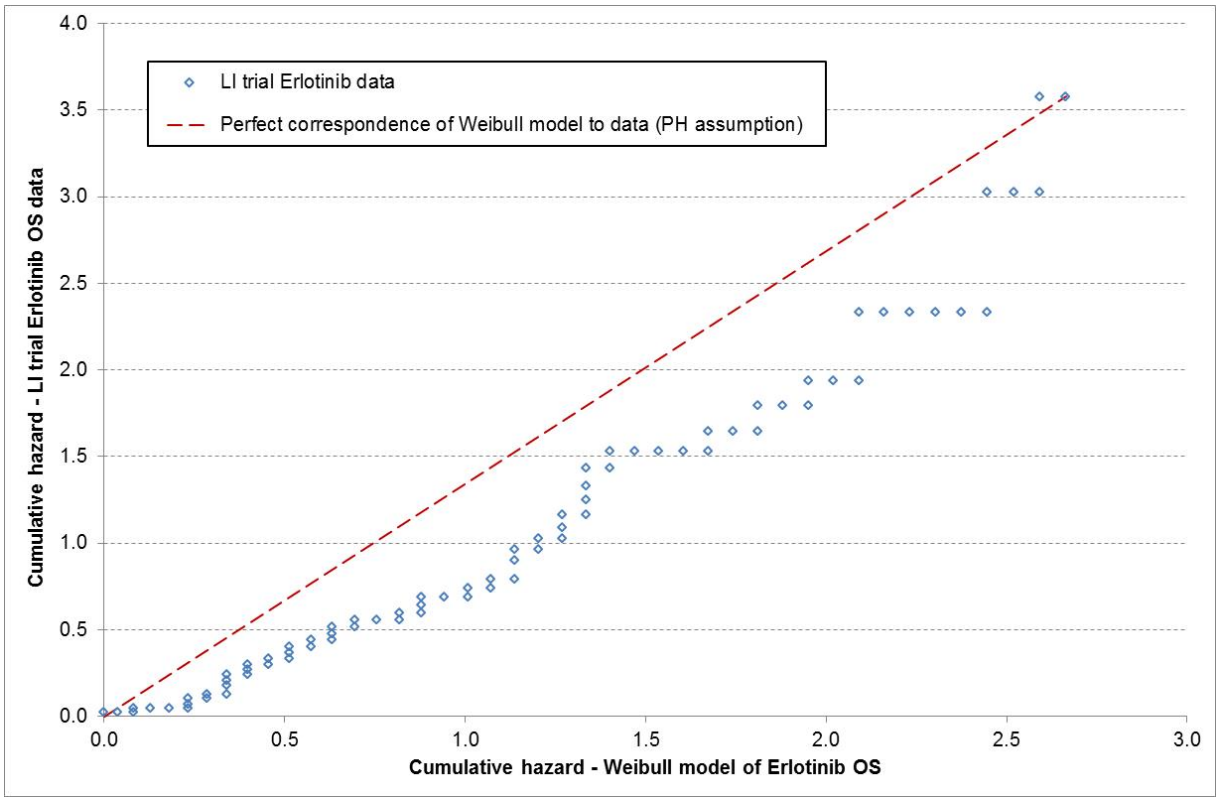
LI=WSY001; OS=overall survival

Figure 16 Weibull OS model for erlotinib-treated patients compared with original trials data



OS=overall survival; PH=proportional hazards

Figure 17 Proportionality check of Weibull erlotinib OS model vs erlotinib data from TAILOR trial



LI=WSY001; OS=overall survival; PH=proportional hazards

Figure 18 Proportionality check of Weibull erlotinib OS model vs erlotinib data from WSY001 trial

These diagnostic checks indicate not only that the estimated OS model estimates are inconsistent within the evidence network, but that the Weibull functional form calibrate from LUME-Lung1[24] trial data when transmitted through the network does not accord with the outcome patterns seen in other network trials. This calls into question the use of both Weibull parametric form and the HR for erlotinib vs nintedanib plus docetaxel estimated from the network.

The potential impact of alterations in OS far outweigh all other aspects of the model (see 5.5.2 above and Table 40 below) and therefore the importance of this finding cannot be over-estimated. The ERG has not been able to complete a full assessment of the PFS network in a similar manner due to time limitations, but early indications are that similar inconsistencies are present. However, PFS data are more complete and have less influence on cost-effectiveness results than OS.

Unfortunately, these problems with the evidence networks are so fundamental that it is not possible to rectify them and modify the company's model to provide improved estimates of OS, PFS and the relative cost-effectiveness of nintedanib plus docetaxel and erlotinib.

5.6 Conclusions of the cost-effectiveness section

Although the structure of the economic model submitted by the company is generally appropriate, the ERG is concerned by the number of implementation errors that have come to light with important consequences for the economic results generated. The ERG has identified eleven specific aspects of the submitted base-case model that are subject to challenge, or involve implementation errors. In each case an appropriate amendment has been introduced into the company's model with results ranging from minor changes to important and substantial changes to the estimated ICER per QALY gained.

Neither the company nor the ERG considers a comparison of nintedanib plus docetaxel to erlotinib to be appropriate to the decision problem. Nevertheless, this was specified in the NICE scope and the company has therefore undertaken such a comparison. However, the ERG considers that this is seriously flawed due to inconsistencies apparent in the available time-to-event data leading to conflicting results from the MTC. The ERG has applied other relevant amendments to the submitted model for this comparison, but the uncertainty in OS, PFS and ToT probably far outweighs all other effects but cannot be quantified.

6 ADDITIONAL WORK UNDERTAKEN BY THE ERG

Table 40 summarises the effects of the various ERG amendments made to the company's decision model (see also Appendix 8). The consequence of applying each proposed amendment is shown separately for comparison with the company's base-case analysis. The joint effect of applying all ERG changes to the model simultaneously is included. In addition, a second summary result is provided excluding the limitation of docetaxel treatment to four cycles because this change reflects an issue of principle (clinical evidence vs UK practice), and because the impact of applying a model revision is necessarily incomplete (the ERG cannot estimate what the effect might be on outcomes of restricting treatment).

Generally these amendments result in increased costs (both absolute and incremental) and/or reduced outcomes (survival and QALYs) and hence lead to increases in the estimated ICER per QALY gained. The company's base-case ICER (£50,776 per QALY gained) is increased to either £85,292 per QALY gained with all revisions applied, or to £82,995 per QALY gained if no limit is placed on the number of cycles of docetaxel treatment allowed.

The most influential change is the application of the ERG OS estimates. If this revision is not accepted, the revised ICER using the other ten revisions becomes £62,719 per QALY gained. The ERG's estimate of the gain in undiscounted mean OS is 3.05 months.

Cost-effectiveness results of applying the non-Time To Event ERG amendments are detailed in Table 41, with a full sensitivity analysis for a range of possible patient access scheme discounts on the list price of erlotinib in Table 42. It should be borne in mind that were it possible to estimate the mean OS for patients treated with erlotinib rather than docetaxel monotherapy in second-line chemotherapy, it is quite likely that the estimated incremental gain in life-years would diminish and the estimated ICER rise substantially.

Table 40 Cost-effectiveness results for nintedanib plus docetaxel vs docetaxel with ERG revisions to company's base-case comparison in the adenocarcinoma population

Model scenario & ERG revisions	Nintedanib + docetaxel			Docetaxel			Incremental			ICER	ICER
	Cost	QALYs	Life years	Cost	QALYs	Life years	Cost	QALYs	Life years	£/QALY	Change
Company's base-case	[redacted]	[redacted]	**1.810**	[redacted]	[redacted]	**1.419**	**+ £11,051**	**+ 0.218**	**+ 0.391**	**£50,776**	-
1) ERG OS estimates	[redacted]	[redacted]	1.493	[redacted]	[redacted]	1.238	+ £10,497	+ 0.153	+ 0.255	£68,587	**+ £17,811**
2) ERG PFS estimates	[redacted]	[redacted]	1.810	[redacted]	[redacted]	1.419	+ £11,527	+ 0.220	+ 0.391	£52,445	**+ £1,669**
3) ERG ToT estimates	[redacted]	[redacted]	1.810	[redacted]	[redacted]	1.419	+ £11,298	+ 0.218	+ 0.391	£51,930	**+ £1,154**
4) Mid-cycle adjustment	[redacted]	[redacted]	1.810	[redacted]	[redacted]	1.419	+ £11,717	+ 0.218	+ 0.391	£53,839	**+ £3,062**
5) Cost of treatment doses	[redacted]	[redacted]	1.810	[redacted]	[redacted]	1.419	+ £11,445	+ 0.218	+ 0.391	£52,587	**+ 1,811**
6) Febrile neutropenia cost	[redacted]	[redacted]	1.810	[redacted]	[redacted]	1.419	+ £11,180	+ 0.218	+ 0.391	£51,372	**+ £595**
7) Monitoring cost	[redacted]	[redacted]	1.810	[redacted]	[redacted]	1.419	+ £11,130	+ 0.218	+ 0.391	£51,140	**+ £364**
8) Discounting method	[redacted]	[redacted]	1.810	[redacted]	[redacted]	1.419	+ £11,189	+ 0.221	+ 0.391	£50,532	**-£244**
9) Disutility of fatigue	[redacted]	[redacted]	1.810	[redacted]	[redacted]	1.419	+ £11,051	+ 0.217	+ 0.391	£50,830	**+ £54**
10) Stable disease costs	[redacted]	[redacted]	1.810	[redacted]	[redacted]	1.419	+ £11,637	+ 0.218	+ 0.391	£53,470	**+ £2,693**
11) Docetaxel ≤4 cycles	[redacted]	[redacted]	1.810	[redacted]	[redacted]	1.419	+ £10,452	+ 0.217	+ 0.391	£48,060	**-£2,716**
Base-case + revisions 1-10	[redacted]	[redacted]	**1.493**	[redacted]	[redacted]	**1.238**	**+ £13,087**	**+ 0.158**	+ 0.255	**£82,995**	**+ £32,219**
Base-case + all revisions	[redacted]	[redacted]	**1.493**	[redacted]	[redacted]	**1.238**	**+ £13,437**	**+ 0.158**	**+ 0.255**	**£85,292**	**+ £34,516**

Costs and QALYs discounted; Life years undiscounted
OS=overall survival; PFS=progression-free survival; QALYs=quality adjusted life years; ToT=time on treatment

Table 41 Cost-effectiveness results for nintedanib plus docetaxel vs erlotinib with ERG revisions to company's base-case comparison in the adenocarcinoma population

Model scenario & ERG revisions	Nintedanib + docetaxel			Erlotinib			Incremental			ICER	ICER
	Cost	QALYs	Life years	Cost	QALYs	Life years	Cost	QALYs	Life years	£/QALY	Change
Company's base-case	[redacted]	[redacted]	**1.445**	[redacted]	[redacted]	**0.979**	**£7,571**	**0.280**	**0.465**	**£27,008**	-
1) ERG OS estimates	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
2) ERG PFS estimates	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
3) ERG ToT estimates	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
4) Mid-cycle adjustment	[redacted]	[redacted]	1.445	[redacted]	[redacted]	0.979	£7,815	0.280	0.465	£27,878	**+ £870**
5) Cost of treatment doses	[redacted]	[redacted]	1.445	[redacted]	[redacted]	0.979	£7,926	0.280	0.465	£28,275	**+ £1,267**
6) Febrile neutropenia cost	[redacted]	[redacted]	1.445	[redacted]	[redacted]	0.979	£7,897	0.280	0.465	£28,173	**+ £165**
7) Monitoring cost	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**
8) Discounting method	[redacted]	[redacted]	1.445	[redacted]	[redacted]	0.979	£7,679	0.285	0.465	£26,927	**-£81**
9) Disutility of fatigue	[redacted]	[redacted]	1.445	[redacted]	[redacted]	0.979	£7,571	0.280	0.465	£27,020	**+ £12**
10) Stable disease costs	[redacted]	[redacted]	1.445	[redacted]	[redacted]	0.979	£7,576	0.280	0.465	£27,027	**+ £19**
11) Docetaxel ≤4 cycles	[redacted]	[redacted]	1.445	[redacted]	[redacted]	0.979	£7,069	0.283	0.465	£24,975	**-£2,033**
Base-case + revisions 4-11	[redacted]	[redacted]	**1.445**	[redacted]	[redacted]	**0.979**	**£8,147**	**0.288**	**0.465**	**£28,307**	**+ £1,299**

Costs and QALYs discounted; Life years undiscounted
NA = not applicable; OS=overall survival; PFS=progression-free survival; QALYs=quality adjusted life years; ToT=time on treatment; X = meaningful amendments for time-to-event estimates are not possible due to unreliable data network or absence of data

Table 42 Cost-effectiveness results for nintedanib plus docetaxel vs erlotinib with ERG revisions to company's base-case comparison in the adenocarcinoma population: sensitivity of ICER to different patient access scheme discount levels for erlotinib.

Model scenario & ERG revisions	**Patient access scheme discount for erlotinib**										
	0%	**5%**	**10%**	**15%**	**20%**	**25%**	**30%**	**35%**	**40%**	**45%**	**50%**
Company's base-case	**£27,008**	**£27,939**	**£28,870**	**£29,802**	**£30,733**	**£31,664**	**£32,596**	**£33,527**	**£34,458**	**£35,390**	**£36,321**
1) ERG OS estimates	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
2) ERG PFS estimates	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
3) ERG ToT estimates	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
4) Mid-cycle adjustment	£27,878	£28,902	£29,926	£30,950	£31,975	£32,999	£34,023	£35,047	£36,071	£37,095	**£38,119**
5) Cost of treatment doses	£28,275	£29,206	£30,138	£31,069	£32,000	£32,932	£33,863	£34,794	£35,726	£36,657	**£37,588**
6) Febrile neutropenia cost	£28,173	£29,104	£30,035	£30,967	£31,898	£32,830	£33,761	£34,692	£35,624	£36,555	**£37,486**
7) Monitoring cost	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**
8) Discounting method	£26,927	£27,851	£28,775	£29,699	£30,623	£31,547	£32,471	£33,395	£34,319	£35,243	**£36,167**
9) Disutility of fatigue	£27,020	£27,951	£28,883	£29,815	£30,747	£31,678	£32,610	£33,542	£34,474	£35,405	**£36,337**
10) Stable disease costs	£27,027	£27,958	£28,890	£29,821	£30,752	£31,684	£32,615	£33,546	£34,478	£35,409	**£36,340**
11) Docetaxel ≤4 cycles	£24,975	£25,897	£26,820	£27,742	£28,664	£29,587	£30,509	£31,431	£32,354	£33,276	**£34,198**
Base-case + revisions 4-11	**£28,307**	**£29,314**	**£30,320**	**£31,327**	**£32,334**	**£33,341**	**£34,348**	**£35,354**	**£36,361**	**£37,368**	**£38,375**

Costs and QALYs discounted; Life years undiscounted
NA = not applicable; OS=overall survival; PFS=progression-free survival; QALYs=quality adjusted life years; ToT=time on treatment; X = meaningful amendments for time-to-event estimates are not possible due to unreliable data network or absence of data

7 END OF LIFE

The company makes a case that nintedanib plus docetaxel meets the criteria set by NICE for end of life treatment. Namely:

- The life expectancy of the patient population was short (< 24 months)
- The number of patients who would be eligible for the treatment is small
- The increase in OS is >3 months

The company states on page 288 of the CS:[1]

- Patients with advanced NSCLC have a short life expectancy of less than 24 months on average. Using the extrapolated results from the LUME-Lung 1[24] trial data implemented in the cost-effectiveness model, the median OS of patients on docetaxel monotherapy (current standard of care) is 10.23 months and the mean OS is 15.96 months.
- The total eligible population for nintedanib plus docetaxel in England is 703.
- Extension to life due to nintedanib plus docetaxel vs docetaxel monotherapy in the target population with the base-case assumptions within the model is a mean of 3.96 months. The extension in OS over erlotinib is a mean of 5.16 months.

The ERG agrees that patients with advanced NSCLC have a short life expectancy of less than 24 months. It also agrees that the patients who would be eligible for the treatment is small. As noted in section 5.5.2, by applying the K-M trial results using the AUC method until the long-term OS trends were established and then projecting remaining estimated survival using the exponential trends, the ERG calculated the mean extension in OS to be 3.05 months for the base-case analysis of nintedanib plus docetaxel vs docetaxel. The ERG were only able to carry out a partial comparison of nintedanib plus docetaxel to erlotinib for reasons outlined in section 5.5.13 (excluding the time-to-event outcomes known to be subject to the most uncertainty) and were therefore unable to derive a mean estimate for OS for nintedanib plus docetaxel vs erlotinib.

8 DISCUSSION

The NICE scope for this STA stipulates the population should be adults with locally advanced or metastatic NSCLC that has progressed following prior chemotherapy. The decision problem differs in that it is restricted to NSCLC with adenocarcinoma histology. It also includes patients with locally recurrent disease. The ERG considers both differences to be appropriate since they reflect the relevant population stipulated by the anticipated licensed indication for nintedanib plus docetaxel.[28] Based on the LUME-Lung 1[24] trial, the majority (94.2%) of these patients will have metastatic disease at the time of second-line treatment. The majority (85% to 90%) of such patients in England would be expected to have EGFR-negative disease,[30-32] [redacted].

The NICE scope also states that docetaxel and erlotinib are relevant comparators. The company notes the preliminary recommendation issued by NICE in February 2014 is that erlotinib should not be recommended for treating locally advanced or metastatic NSCLC that has progressed after non-targeted chemotherapy in people with tumours that are EGFR-negative. The ERG also notes the same recommendation from August 2014.[21] Furthermore, in current clinical practice in England, the majority of patients with EGFR-positive NSCLC receive erlotinib (or another TKI) as first-line treatment.[13-15] These patients would, therefore, be unlikely to receive erlotinib as a second-line therapy. Because nintedanib is administered in combination with docetaxel, patients in receipt of nintedanib must be fit enough to receive docetaxel. Such patients are, therefore, likely to be assessed as ECOG PS 0 to 1. The general opinion of clinical advisors to both the company and ERG is that patients who are sufficiently fit to tolerate treatment with docetaxel will receive docetaxel rather than erlotinib. In view of these factors, while the decision problem does include erlotinib as a comparator for secondary analyses, this is nevertheless considered by the company to be an irrelevant comparator to nintedanib plus docetaxel. The ERG agrees with the company that erlotinib is not a relevant comparator for the same reasons.

Evidence for the relative effectiveness of nintedanib plus docetaxel is derived from the LUME-Lung 1[24] trial which compares nintedanib plus docetaxel to placebo plus docetaxel. This, therefore, provides direct evidence for the clinical effectiveness of nintedanib plus docetaxel vs docetaxel alone. The trial appears to be of good quality and low risk of bias and reports that nintedanib plus docetaxel is superior to placebo plus docetaxel in terms of OS (median improvement of 2.3 months) and PFS (median improvement of 1.4 months). However, the ERG does not consider that the assumption of proportional hazards is consistent with the trial data, and therefore use of these results in cost-effectiveness

modelling should not be based implicitly or explicitly on this assumption. The reported gain in efficacy is accompanied by an increase in CTCAE grade ≥3 AEs and SAEs but these AEs are reported to be generally manageable. Some differences in HRQoL between treatment arms have been reported but none result in differences between arms in terms of overall global health status/QoL. The AEs of greatest concern are fatal AEs. More fatal AEs have been reported in the nintedanib plus docetaxel arm than in the placebo plus docetaxel arm. However, the numbers are small and the company is using ongoing surveillance to monitor this issue. Neutropenia and sepsis have also been identified as important risks.

One potential limitation with regard to the generalisability of the findings from LUME-Lung 1[24] to clinical practice in England relates to three of the exclusion criteria that the trial employed. First, patients with major pleural effusion were excluded. Second, patients with evidence of cavitary or necrotic tumours were excluded. Third, patients receiving therapeutic anticoagulation (except low dose heparin) or antiplatelet therapy (except for chronic low-dose therapy with acetylsalicylic acid ≤325mg/day) were excluded. Pleural effusions[50,51] and venous thromboembilsm [52] appear to predict poor prognosis; evidence of cavitary or necrotic tumours may also result in a worse prognosis, although cavitation may be a less strong prognostic factor.[49] These exclusion criteria may partially explain why, in LUME-Lung 1,[24] a higher proportion of patients than would be expected in clinical practice also received third-line treatment on disease progression. This may in turn also be an indicator that patients included in this trial were fitter than those generally seen in NHS clinical practice.

In order to derive an estimate for cost-effectiveness of nintedanib plus docetaxel to docetaxel alone, the company have developed a de novo partitioned survival Markov model, which incorporates data from LUME-Lung 1[24] alongside other published sources. The company's estimate of cost-effectiveness for nintedanib plus docetaxel vs docetaxel is £50,766 per QALY gained. However, the ERG identified a number of weaknesses in the company's model and is concerned about the number of implementation issues that it identified. The most important area in terms of its impact on the ICER related to OS estimation. Here inadequate information was provided about specific data sources used for SEER and LUCADA used to validate the long-term extrapolation of OS. Furthermore OS projection was based on the flawed assumption that there is constant hazard over time.

In total the ERG made 11 changes to the company's model. These related to: inappropriate methods used to project time-to-event outcomes (OS, PFS and time-on-treatment); mid-cycle adjustment error; inappropriate methods used to estimate cost of treatment doses; underestimate of true cost of febrile neutropenia; monitoring costs; non-UK standard approach to discounting; overall average disutility estimate for fatigue used for both

regimens; error in stable disease costs and erroneous restriction of docetaxel to four cycles. When all of the ERG's alterations are implemented, the ERG's revised estimate of cost-effectiveness for the comparison of nintedanib plus docetaxel with docetaxel is £85,292 per QALY gained. Independently, implementing each of the ERG's changes in the model results in ICERs ranging from £50,532 to £68,587 per QALY gained. The change which has the largest impact on the size of the ICER is the method used to estimate OS. If all of the other changes in the model are implemented, except replacement of the company's OS model, the ICER increases to £62,719 per QALY gained.

There is no direct evidence for the relative effectiveness of nintedanib plus docetaxel compared with erlotinib. In order to compare the relative clinical effectiveness for these two regimens, the company conducted a number of MTCs. However, the ERG has identified a number of uncertainties and weaknesses in relation to these MTCs. Relating to the generalisability of the trials to clinical practice, the ERG notes that only patients in one trial, LUME-Lung 1,[24] had received pemetrexed as a first-line treatment and even then, this was only a minority (18.8%). Pemetrexed is now the treatment of choice for adenocarcinoma patients with EGFR-negative disease who, as noted above, constitute the majority of adenocarcinoma patients in England.

There are also a number of methodological weaknesses with the conduct of the MTCs, the most important being that proportional hazards are presumed to hold throughout the MTC networks for both PFS and OS. As discussed above, the ERG has found that this is not the case within the LUME-Lung 1[24] and, as a consequence, any results generated comparing nintedanib plus docetaxel with erlotinib cannot be considered reliable. Important differences in trial and patient characteristics of trials included in the MTCs have also been observed which question the validity of the base-case, scenario and sensitivity analyses.

To compare nintedanib plus docetaxel to erlotinib, the results from the MTCs have been incorporated into the company's model. The company's estimate of cost-effectiveness for nintedanib plus docetaxel vs erlotinib is £27,008 per QALY gained. However, as discussed above, there are a number of methodological issues with the conduct of the MTCs which undermine any confidence in this estimate. Furthermore, in addition to those discussed above, additional problems have been identified in relation to ToT where again the assumption for proportional hazards is assumed. It is impossible to ascertain whether this is true for any trial other than LUME-Lung 1[24] as these data were not available for any other trial. However this assumption did not hold for LUME-Lung 1.[24] Furthermore, the ERG also established that not only is the assumption of proportional hazards for OS violated for LUME-Lung 1[24] but this is also violated for OS reported in two other trials (WSY001[62] and

TAILOR[59]) included in the MTC. Thus because of concerns about the relevance of erlotinib as a comparator and the appropriateness of the analyses conducted, the ERG only considers it feasible to estimate a reliable ICER per QALY gained using the direct trial data from LUME-Lung 1[24] for patients with locally advanced, metastatic or locally recurrent NSCLC of adenocarcinoma tumour histology after first-line chemotherapy.

9 OVERALL CONCLUSIONS

The ERG agrees that LUME-Lung 1[24] is a high quality trial that demonstrates the efficacy of nintedanib plus docetaxel over docetaxel for patients with adenocarcinoma after first-line chemotherapy. Based on the clinical and cost-effectiveness evidence available, the ERG only considers it feasible to estimate an ICER per QALY gained using the direct trial data from LUME-Lung 1[24] for this population. The ERG concludes that the comparison of nintedanib plus docetaxel vs docetaxel yields an ICER that is higher than £50,000 per QALY gained.

10 REFERENCES

1. Boehringer Ingelheim. Single technology appraisal (STA) submitted by Boehringer Ingelheim Ltd: Nintedanib in combination with docetaxel for the treatment of adult patients with locally advanced, metastatic or recurrent non-small cell lung cancer (NSCLC) of adenocarcinoma tumour histology after first-line chemotherapy. 2014.
2. American Cancer Society. Lung Cancer (Non-Small Cell). 2013; Available from: http://www.cancer.org/cancer/lungcancer-non-smallcell/detailedguide/non-small-cell-lung-cancer-what-is-non-small-cell-lung-cancer (Accessed 19/06/2013).
3. NHS Choices. Lung cancer. 2013 [updated 02/08/2013]; Available from: http://www.nhs.uk/Conditions/Cancer-of-the-lung/Pages/Introduction.aspx (Accessed 16/09/2014).
4. Roy Castle Lung Cancer Foundation. Lung Cancer Facts and Figures. 2014; Available from: http://www.roycastle.org/lung-cancer/Lung-Cancer-Facts-and-Figures (Accessed 16/09/2014).
5. Cancer Research UK. Lung Cancer Key Facts. 2014 [updated 22/02/13 ('What causes lung cancer?' section)]; Available from: http://www.cancerresearchuk.org/cancer-info/cancerstats/keyfacts/lung-cancer/ (Accessed 16/09/2014).
6. Jett JR, Schild SE, Keith RL, Kesler KA. Treatment of non-small cell lung cancer, stage IIIB: ACCP evidence-based clinical practice guidelines (2nd edition). Chest. 2007; 132:266s-76s.
7. Socinski MA, Crowell R, Hensing TE, Langer CJ, Lilenbaum R, Sandler AB*, et al.* Treatment of non-small cell lung cancer, stage IV: ACCP evidence-based clinical practice guidelines (2nd edition). Chest. 2007; 132:277S-89S.
8. Bottomley A, Efficace F, Thomas R, Vanvoorden V, Ahmedzai SH. Health-related quality of life in non-small-cell lung cancer: methodologic issues in randomized controlled trials. J Clin Oncol. 2003; 21:2982-92.
9. Degner LF, Sloan JA. Symptom distress in newly diagnosed ambulatory cancer patients and as a predictor of survival in lung cancer. J Pain Symptom Manage. 1995; 10:423-31.
10. Maione P, Rossi A, Bareschino MA, Sacco PC, Schettino C, Falanga M*, et al.* Factors driving the choice of the best second-line treatment of advanced NSCLC. Rev Recent Clin Trials. 2011; 6:44-51.
11. The Healthcare Quality Improvement Partnership (HQIP) The Health and Social Care Information Centre (HSCIC) The Royal College of Physicians (RCP). National Lung Cancer Audit Report. Report for the audit period 2011. 2012; Available from: http://www.hqip.org.uk/assets/NCAPOP-Library/NCAPOP-2012-13/Lung-Cancer-National-Audit-Report-pub-2012.pdf (Accessed 16/09/2014).
12. National Institute for Health Research. MK-3475 for advanced or recurrent PD-L1 positive non-small cell lung cancer – second line. 2013; Available from: http://www.hsc.nihr.ac.uk/files/downloads/2210/2508.10efb920.MK3475_nscl_Nov13.pdf (Accessed 10 October 2014).
13. National Institute for Health and Care Excellence (NICE). Erlotinib for the first-line treatment of locally advanced or metastatic EGFR-TK mutation-positive non-small cell lung cancer. NICE technology appraisal guidance 258. 2012; Available from: http://guidance.nice.org.uk/TA258 (Accessed 16/09/2014).
14. National Institute for Health and Care Excellence (NICE). Gefitinib for the first-line treatment of locally advanced or metastatic non-small-cell lung cancer. NICE

technology appraisal guidance 192. 2010; Available from: http://www.nice.org.uk/guidance/ta192 (Accessed 16/09/2014).

15. National Institute for Health and Care Excellence (NICE). Afatinib for treating epidermal growth factor receptor mutation-positive locally advanced or metastatic non-small-cell lung cancer. NICE technology appraisal guidance 310. 2014; Available from: http://www.nice.org.uk/guidance/ta310 (Accessed 16/09/2014).

16. Leighl NB. Treatment paradigms for patients with metastatic non-small-cell lung cancer: first-, second-, and third-line. Curr Oncol. 2012; 19:S52-8.

17. Stinchcombe TE, Socinski MA. Considerations for second-line therapy of non-small cell lung cancer. Oncologist. 2008; 13 Suppl 1:28-36.

18. Boehringer Ingelheim. Cytotoxic regimens by line of treatment for the UK. Data On File. ONC 13-01. 2013.

19. National Institute for Health and Care Excellence (NICE). Erlotinib for the treatment of non-small-cell lung cancer. NICE technology appraisal guidance 162. 2008; Available from: http://www.nice.org.uk/guidance/ta162 (Accessed 16/09/2014).

20. National Institute for Health and Clinical Excellence (NICE). The diagnosis and treatment of lung cancer (update of Clinical Guideline 24). NICE clinical guideline 121. 2011; Available from: http://www.nice.org.uk/guidance/CG121 (Accessed 16/09/2014).

21. National Institute for Health and Care Excellence (NICE). Erlotinib and gefitinib for treating non-small-cell lung cancer that has progressed following prior chemotherapy (Review of TA162 and TA175): appraisal consultation document. 2014; Available from: http://www.nice.org.uk/guidance/gid-tag347/resources/erlotinib-and-gefitinib-for-treating-nonsmallcell-lung-cancer-that-has-progressed-following-prior-chemotherapy-review-of-ta162-and-ta175-appraisal-consultation-document (Accessed 16/09/2014).

22. Hilberg F, Roth GJ, Krssak M, Kautschitsch S, Sommergruber W, Tontsch-Grunt U*, et al.* BIBF 1120: triple angiokinase inhibitor with sustained receptor blockade and good antitumor efficacy. Cancer Res. 2008; 68:4774-82.

23. Reck M, Kaiser R, Eschbach C, Stefanic M, Love J, Gatzemeier U*, et al.* A phase II double-blind study to investigate efficacy and safety of two doses of the triple angiokinase inhibitor BIBF 1120 in patients with relapsed advanced non-small-cell lung cancer. Ann Oncol. 2011; 22:1374-81.

24. Reck M, Kaiser R, Mellemgaard A, Douillard JY, Orlov S, Krzakowski M*, et al.* Docetaxel plus nintedanib versus docetaxel plus placebo in patients with previously treated non-small-cell lung cancer (LUME-Lung 1): a phase 3, double-blind, randomised controlled trial. Lancet Oncol. 2014; 15:143-55.

25. Ellis LM, Hicklin DJ. VEGF-targeted therapy: mechanisms of anti-tumour activity. Nat Rev Cancer. 2008; 8:579-91.

26. Folkman J. Seminars in Medicine of the Beth Israel Hospital, Boston. Clinical applications of research on angiogenesis. N Engl J Med. 1995; 333:1757-63.

27. Bergers G, Benjamin LE. Tumorigenesis and the angiogenic switch. Nat Rev Cancer. 2003; 3:401-10.

28. European medicines Agency (EMA). Vargatef. 2014; Available from: http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/human/medicines/002569/smops/Positive/human_smop_000727.jsp&mid=WC0b01ac058001d127 (Accessed 07/10/2014).

29. Zimmermann FB, Molls M, Jeremic B. Treatment of recurrent disease in lung cancer. Semin Surg Oncol. 2003; 21:122-7.

30. Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E*, et al.* Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial. Lancet Oncol. 2012; 13:239-46.

31. Peters S, Adjei AA, Gridelli C, Reck M, Kerr K, Felip E. Metastatic non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2012; 23 Suppl 7:vii56-64.

32. Cancer Research UK. Biological therapy for lung cancer. Available from: http://www.cancerresearchuk.org/about-cancer/type/lung-cancer/treatment/biological-therapy-for-lung-cancer (Accessed 07/10/2014).

33. European medicines Agency (EMA). Taxotere 20 mg/0.5 ml concentrate and solvent for solution for infusion. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/000073/WC500035264.pdf (Accessed 30/09/2014).

34. Mellemgaard A, Kaiser R, Douillard JY, Orlov SV, Krzakowski MJ, Von Pawel J*, et al.* Analysis of overall survival in adenocarcinoma NSCLC patients receiving 2nd line combination treatment with nintedanib (BIBF 1120) + docetaxel in the LUME-Lung 1 trial: A randomized, double-blind, placebo-controlled phase 3 study. Eur J Cancer. 2013; 49:S798.

35. Reck M, Mellemgaard A, Douillard J, Orlov S, Krzakowski M, Von Pawel J*, et al.* Nintedanib (BIBF 1120) + docetaxel as 2nd-line therapy in patients with stage IIIB/IV or recurrent NSCLC: Results of the Phase III, randomised, double-blind LUME-Lung 1 trial. (For the LUME-Lung 1 Study Group). J Thorac Oncol. 2014; 9:S39-S40.

36. Reck M, Mellemgaard A, Orlov SV, Krzakowski MJ, Von Pawel J, Gottfried M*, et al.* Antiangiogenic-specific adverse events (AEs) in patients with non-small cell lung cancer (NSCLC) treated with nintedanib (N) and docetaxel (D). J Clin Oncol. 2014; 32:15_suppl:8100.

37. Reck M, Novello S, Mellemgaard A, Orlov S, Kaiser R, Barrueco J*, et al.* Impact of tumor burden on the overall survival analysis of the lume-lung 1 study: A randomized, double-blind phase 3 trial of nintedanib (BIBF 1120) + docetaxel in NSCLC patients progressing after first-line chemotherapy. J Thorac Oncol. 2013; 8:S196.

38. The Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. [updated 21/03/2011]; Available from: http://www.cochrane.org/training/cochrane-handbook (Accessed 16/09/2014).

39. Hanna NH, Kaiser R, Sullivan RN, Aren OR, Ahn MJ, Tiangco B*, et al.* Lume-lung 2: A multicenter, randomized, double-blind, phase III study of nintedanib plus pemetrexed versus placebo plus pemetrexed in patients with advanced nonsquamous non-small cell lung cancer (NSCLC) after failure of first-line chemotherapy. J Clin Oncol. 2013; 31:15_suppl:8034.

40. Boehringer Ingelheim. Multicentre, randomised, double-blind, Phase III trial to investigate the efficacy and safety of oral BIBF 1120 plus standard docetaxel therapy compared to placebo plus standard docetaxel therapy in patients with stage IIIB/IV or recurrent non small cell lung cancer after failure of first line chemotherapy (LUME Lung 1). Clinical Trial Report (Primary PFS analysis). Data on file. ONC 14-19. 2012.

41. Boehringer Ingelheim. Multicentre, randomised, double-blind, Phase III trial to investigate the efficacy and safety of oral BIBF 1120 plus standard docetaxel therapy compared to placebo plus standard docetaxel therapy in patients with stage IIIB/IV or

recurrent non small cell lung cancer after failure of first line chemotherapy (LUME Lung 1). Clinical Trial Report (Final OS Analysis). Data on file. ONC 14-18. 2013.

42. Boehringer Ingelheim. LUME-Lung 1 CTR (1199.13) (final OS analysis). Appendix. Data on file. ONC 14-24. 2013.

43. Boehringer Ingelheim. Multicentre, randomised, double-blind, Phase III trial to investigate the efficacy and safety of oral BIBF 1120 plus standard docetaxel therapy compared to placebo plus standard docetaxel therapy in patients with stage IIIB/IV or recurrent non small cell lung cancer after failure of first line chemotherapy. Trial Statistical Analysis Plan. ONC 14-20. 2011.

44. Boehringer Ingelheim. Multicentre, randomised, double-blind, Phase III trial to investigate the efficacy and safety of oral BIBF 1120 plus standard docetaxel therapy compared to placebo plus standard docetaxel therapy in patients with stage IIIB/IV or recurrent non small cell lung cancer after failure of first line chemotherapy. Trial Statistical Analysis Plan addendum. Data on file. ONC 14-21. 2011.

45. Boehringer Ingelheim. Summary of Clinical Efficacy. Data on file. ONC 14-22. 2013.

46. Kaiser R, Barrueco JR, Reck M, Hanna N, Gann C, Glomb P*, et al.* Identification of a clinical biomarker for 2nd line combination with nintedanib in adenocarcinoma non-small cell lung cancer (NSCLC) patients in two phase III trials. Eur J Cancer. 2013; 49:S822.

47. Novello S, Mellemgaard A, Kaiser R, Douillard JY, Orlov S, Krzakowski M*, et al.* Analysis of patient-reported outcomes from the lume-lung 1 trial: A randomized, double-blind, placebo-controlled phase 3 study in second-line advanced non-small cell lung cancer (NSCLC) patients. J Thorac Oncol. 2013; 8:S1207.

48. Reck M, Kaiser R, Mellemgaard A, Douillard JY, Orlov S, Krzakowski MJ*, et al.* Nintedanib (BIBF 1120) plus docetaxel in NSCLC patients progressing after first-line chemotherapy: LUME Lung 1, a randomized, double-blind phase III trial. J Clin Oncol. 2013; 31:abstract LBA8011.

49. Phernambucq EC, Hartemink KJ, Smit EF, Paul MA, Postmus PE, Comans EF*, et al.* Tumor cavitation in patients with stage III non-small-cell lung cancer undergoing concurrent chemoradiotherapy: incidence and outcomes. J Thorac Oncol. 2012; 7:1271-5.

50. Morgensztern D, Waqar S, Subramanian J, Trinkaus K, Govindan R. Prognostic impact of malignant pleural effusion at presentation in patients with metastatic non-small-cell lung cancer. J Thorac Oncol. 2012; 7:1485-9.

51. Ryu JS, Ryu HJ, Lee SN, Memon A, Lee SK, Nam HS*, et al.* Prognostic impact of minimal pleural effusion in non-small-cell lung cancer. J Clin Oncol. 2014; 32:960-7.

52. Hicks LK, Cheung MC, Ding K, Hasan B, Seymour L, Le Maitre A*, et al.* Venous thromboembolism and nonsmall cell lung cancer: a pooled analysis of National Cancer Institute of Canada Clinical Trials Group trials. Cancer. 2009; 115:5516-25.

53. Boehringer Ingelheim. Multicentre, randomised, double-blind, Phase III trial to investigate the efficacy and safety of oral BIBF 1120 plus standard docetaxel therapy compared to placebo plus standard docetaxel therapy in patients with stage IIIB/IV or recurrent non small cell lung cancer after failure of first line chemotherapy. Clinical Trial Protocol. Data on file. ONC 14-30. 2008.

54. National Institute for Health and Clinical Excellence (NICE). Guide to the methods of technology appraisal 2013. 2013; Available from: http://publications.nice.org.uk/pmg9 (Accessed 03/10/2014).

55. Kim ST, Uhm JE, Lee J, Sun JM, Sohn I, Kim SW*, et al.* Randomized phase II study of gefitinib versus erlotinib in patients with advanced non-small cell lung cancer who failed previous chemotherapy. Lung Cancer. 2012; 75:82-8.

56. Scagliotti G, Hanna N, Fossella F, Sugarman K, Blatter J, Peterson P*, et al.* The differential efficacy of pemetrexed according to NSCLC histology: a review of two Phase III studies. Oncologist. 2009; 14:253-63.

57. Sun JM, Lee KH, Kim SW, Lee DH, Min YJ, Yun HJ*, et al.* Gefitinib versus pemetrexed as second-line treatment in patients with nonsmall cell lung cancer previously treated with platinum-based chemotherapy (KCSG-LU08-01): an open-label, phase 3 trial. Cancer. 2012; 118:6234-42.

58. Lee DH, Lee JS, Kim SW, Rodrigues-Pereira J, Han B, Song XQ*, et al.* Three-arm randomised controlled phase 2 study comparing pemetrexed and erlotinib to either pemetrexed or erlotinib alone as second-line treatment for never-smokers with non-squamous non-small cell lung cancer. Eur J Cancer. 2013; 49:3111-21.

59. Garassino MC, Martelli O, Broggini M, Farina G, Veronese S, Rulli E*, et al.* Erlotinib versus docetaxel as second-line treatment of patients with advanced non-small-cell lung cancer and wild-type EGFR tumours (TAILOR): a randomised controlled trial. Lancet Oncol. 2013; 14:981-8.

60. Ciuleanu T, Stelmakh L, Cicenas S, Miliauskas S, Grigorescu AC, Hillenbach C*, et al.* Efficacy and safety of erlotinib versus chemotherapy in second-line treatment of patients with advanced, non-small-cell lung cancer with poor prognosis (TITAN): a randomised multicentre, open-label, phase 3 study. Lancet Oncol. 2012; 13:300-8.

61. Maruyama R, Nishiwaki Y, Tamura T, Yamamoto N, Tsuboi M, Nakagawa K*, et al.* Phase III study, V-15-32, of gefitinib versus docetaxel in previously treated Japanese patients with non-small-cell lung cancer. J Clin Oncol. 2008; 26:4244-52.

62. Li N, Ou W, Yang H, Liu QW, Zhang SL, Wang BX*, et al.* A randomized phase 2 trial of erlotinib versus pemetrexed as second-line therapy in the treatment of patients with advanced EGFR wild-type and EGFR FISH-positive lung adenocarcinoma. Cancer. 2014; 120:1379-86.

63. Hanna N, Shepherd FA, Fossella FV, Pereira JR, De Marinis F, von Pawel J*, et al.* Randomized phase III trial of pemetrexed versus docetaxel in patients with non-small-cell lung cancer previously treated with chemotherapy. J Clin Oncol. 2004; 22:1589-97.

64. National Confidential Enquiry into Patient Outcome and Death. For better, for worse? A review of the care of patients who died within 30 days of receiving systemic anti-cancer therapy. London: National Confidential Enquiry into Patient Outcome and Death 2008.

65. Hirsh V. Is the Evaluation of Quality of Life in NSCLC Trials Important? Are the Results to be Trusted? Front Oncol. 2014; 4:173.

66. Chouaid C, Agulnik J, Goker E, Herder G, Lester J, Vansteenkiste J*, et al.* Health-related quality of life and utility in patients with advanced non-small-cell lung cancer: a prospective cross-sectional patient survey in a real-world setting. J Thorac Oncol. 2013; 8:997-1003.

67. Nafees B, Stafford M, Gavriel S, Bhalla S, Watkins J. Health state utilities for non small cell lung cancer. Health Qual Life Out. 2008; 6. doi: 10.1186/1477-7525-6-84

68. Department of Health. Electronic Market Information Tool. Commercial Medicines Unit; 2014; Available from: http://cmu.dh.gov.uk/electronic-market-information-tool-emit (Accessed 7 October 2014).

69. Department of Health. NHS Reference costs guidance 2012-13. London 2013; Available from: https://www.gov.uk/government/publications/reference-costs-guidance-for-2012-13--2 (Accessed 15/10/2014).

70. MIMS. Tarceva (erlotinib). 2013; Available from: http://www.mims.co.uk/Drugs/cancer/antineoplastics/tarceva/ (Accessed 7 October 2014).

71. Boehringer Ingelheim. Health Economic Report for LUME-Lung 1. Data on file. ONC 14-29. 2014.

72. National Institute for Health and Care Excellence (NICE). Pemetrexed for the first-line treatment of non-small-cell lung cancer. NICE technology appraisal guidance 181. London 2009; Available from: http://guidance.nice.org.uk/TA181 (Accessed 16/09/2014).

73. Curtis L. Unit Costs of Health and Social Care (PSSRU) 2012. 2012; Available from: http://www.pssru.ac.uk/project-pages/unit-costs/2012/ (Accessed 14/10/2014).

74. Morgan A, Sutton A, Wailoo A. The risk and costs of febrile neutropenia in patients with non small cell lung cancer treated with docetaxel. Sheffield: NICE Decision Support Unit, ScHARR; 2007; Available from: http://www.nicedsu.org.uk/PDFs%20of%20reports/Erlotinib%20DSU%20final%20report1.pdf (Accessed 14/10/2014).

75. Greenhalgh J, Bagust A, Boland A, Dwan K, Beale S, Hockenhull J*, et al.* Erlotinib and gefitinib for treating non-small cell lung cancer that has progressed following prior chemotherapy (review of NICE technology appraisals 162 and 175). Health Technol Assess. 2014 (in press).

11 APPENDICES

Appendix 1: Detailed critique of the company's search strategy

The ERG's critique of the company's search strategy was undertaken in two parts: (i) An examination of the sources searched and the terms used to make a judgement whether the strategy appeared to be sufficient; (ii) The conduct of its own search strategy to determine if any additional relevant studies were identified. The sources searched by the company and the ERG are summarised in Table 43.

Table 43 Databases searched

Databases searched by company	Databases searched by ERG
Bibliographic databases: • MEDLINE and MEDLINE In-Process (PubMed) • EMBASE (Interface not stated) • Cochrane Library (Wiley Interscience): o Cochrane Database of Systematic Reviews o Cochrane Central Register of Controlled Trials	Bibliogrphic databases: • MEDLINE and MEDLINE In-Process (OvidSP) • EMBASE (OvidSP) • Cochrane Library (Wiley Interscience): o Cochrane Database of Systematic Reviews o Cochrane Central Register of Controlled Trials o Database of Abstracts of Reviews of Effects (DARE) o Health Technology Assessment Database (HTA) o NHS Economic Evaluation Database (NHS EED)
The following sources were searched for grey literature: • ClinicalTrials.gov (www.clinicaltrials.gov) • American Society for Clinical Oncology (ASCO) annual meeting (www.asco.org) • European Society for Medical Oncology (ESMO) annual meeting (www.esmo.org/) • National Guidelines Clearinghouse In addition, reference lists of identified systematic reviews were assessed for additional relevant studies	The following online sources were searched for grey literature: • ClinicalTrials.gov (www.clinicaltrials.gov) • American Society for Clinical Oncology (ASCO) annual meeting (www.asco.org) • European Society for Medical Oncology (ESMO) annual meeting (www.esmo.org/) • National Institute for Health and Care Excellence (www.nice.org.uk) • metaRegister of Controlled Trials (http://www.controlled-trials.com/mrct/) • US Food and Drug Administration (www.fda.gov/) • European Medicines Agency (www.ema.europa.eu/) • National Institute for Health and Clinical Excellence (http://www.nice.org.uk/) • International Society for Pharmacoeconomics and Outcomes Research (www.ispor.org) • Scottish Medicines Consortium (https://www.scottishmedicines.org.uk/) • Summary of Product Characteristics (www.medicines.org.uk/emc/medicine/20929/SPC/tyverb) • Medicines and Healthcare products Regulatory Agency (http://www.mhra.gov.uk/) • The European Union Clinical Trials Register (https://www.clinicaltrialsregister.eu/)

Direct evidence

Five databases were searched by the company on 28 February 2014. These are the minimum specified by NICE and the ERG considers would be sufficient to identify relevant studies. The same search strategy was run across all databases and included free text and MeSH terms of lung cancer, relapsed and second line search terms and randomised controlled trial. The search was limited to humans. The company limited online grey literature searching to the past four years (from January 2011 to February 2014) as they stated conference proceedings older than four years of high quality can be expected to be published in peer viewed journals and therefore picked up in the search results. In addition to the databases searched, the citation lists of relevant systematic reviews published since 2009 were also examined to identify other relevant studies. The ERG considers this search to be adequate although some cancer synonyms have been missed and combining search terms with 'AND' as opposed to 'adjacency' reduces the precision of the search.

The ERG conducted its own searches on 8th August 2014. The ERG search strategy also included free text and MeSH terms, drug search terms and a search term filter to identify RCTs. It did not identify any additional studies.

Indirect evidence

The company completed MTC searches on the same date as the systematic review searches using the same search terms and the same databases. The ERG conducted searches on 21st August 2014 and searched the same databases as its previous search. The search terms included free text and MeSH search terms. An RCT filter was used. The strategy also included a drug comparison concept combined as follows:

- Nintedinab + docetaxel vs docetaxel
- Docetaxel vs gefitinib
- Docetaxel vs erlotinib
- Docetaxel vs pemetrexed
- Pemetrexed vs gefitinib
- Pemetrexed vs erlotinib
- Pemetrexed vs pemetrexed + erlotinib
- Pemetrexed + erlotinib vs erlotinib
- Erlotinib vs gefitinib

No additional studies were identified by the ERG that met the company's eligibility criteria for inclusion into the MTC.

Appendix 2: Eligibility criteria for study inclusion into the company's systematic review and MTC

Table 44 describes the eligibility criteria employed by the company for inclusion into its systematic review. In addition, all non-nintedanib studies were subsequently excluded from the results of the search.

Table 44 Eligibility criteria for inclusion into the company's systematic review

Parameter	Inclusion criteria	Exclusion criteria
Population	**Relapsed or refractory NSCLC** Adults with histologically or cytologically confirmed, locally advanced and/or metastatic NSCLC of stage IIIB or IV (according to American Joint Committee on Cancers) or recurrent NSCLC (all histologies): • Squamous-cell carcinoma • Adenocarcinoma • Large cell carcinoma	Any patient population other than relapsed or refractory NSCLC
Interventions	Any second-line pharmacological treatment for relapsed or refractory NSCLC • Monotherapy • Combination chemotherapy	Patients who were treatment-naïve, had received more than first-line therapy, or had received only non-pharmacological interventions
Outcomes	Relevant outcomes for full-text inclusion: • Overall survival and progression-free survival • Time to relapse • Time to death • Adverse events (all CTCAE grades and CTCAE grade 3 to 4) • Withdrawals • Mean dose and number of cycles of therapy received	No outcomes of interest
Study design	Randomised controlled trials (RCTs) only	Not an RCT (e.g. observational)
Language restrictions	Any language‡	
Date	2000 onwards*	Prior to 2000*
Country	Any	None

NSCLC=non-small cell lung cancer
‡ Non-English-language publications were identified for the efficacy review but none met the inclusion criteria.
*Abstracts published prior to the year 2011 and systematic reviews published prior to the year 2009 were excluded.
Source: Table 6 of the CS[1]

Table 45 describes the eligibility criteria employed by the company for inclusion, with rationale, into its MTC. The search was also limited to include only results with abstracts.

For both the systematic review and MTC, all abstracts obtained from the database search were each examined manually by two researchers applying the predefined eligibility criteria. Following this, a random sample of excluded abstracts was checked for accuracy by a third researcher to confirm the exclusion decisions. Any discrepancy in the decision to include or exclude a study was reviewed by and resolved between researchers. The full-text articles for

abstracts deemed potentially relevant during this first level of screening were retrieved in order to confirm their inclusion in the review. All full-text publications were independently reviewed by two researchers, with all disagreements being resolved by consensus.

Table 45 Eligibility criteria for inclusion in the company's MTC

Parameter	Inclusion criteria	Exclusion criteria	Rationale
Population	**Relapsed or refractory NSCLC (RR NSCLC)** Adults with histologically or cytologically confirmed, locally advanced and/or metastatic NSCLC of stage IIIB or IV (according to American Joint Committee on Cancers) or recurrent NSCLC (all histologies, including patients with mixed histology): • Squamous-cell carcinoma • Adenocarcinoma • Large cell carcinoma Additional inclusion criteria applied during feasibility assessment: • Study must report data for adenocarcinoma subgroup, or 75% or more of participants should have adenocarcinoma	Studies not assessing patients with locally advanced or metastatic, stage IIIB, or IV/recurrent NSCLC Additional exclusion criteria applied during feasibility assessment: • Study does not report data for an adenocarcinoma subgroup • Fewer than 75% of participants overall had adenocarcinoma	The patient population evaluated in our MTC matches the population for which nintedanib is being considered for approval.
Interventions	Any second-line pharmacological treatment for RR NSCLC: • Monotherapy • Combination therapy with other pharmacological agents Additional inclusion criteria applied during feasibility assessment: • Intervention should be licensed for use as second-line treatment for NSCLC	• Trials evaluating non-second-line treatment (e.g., first-, third- or subsequent-line therapy) without subgroup data provided for second-line treatment only • Dose comparison studies without a placebo or control arm • Studies evaluating maintenance treatment	To evaluate nintedanib vs currently available licensed interventions for the second-line treatment of RR NSCLC.
Comparators	Any pharmacotherapy or no treatment: • Other second-line pharmacological treatment • Usual care/no additional intervention • Placebo	None in addition to the above criteria	To compare included interventions with common comparators currently available for the second-line treatment of RR NSCLC, as well as usual care/no intervention and placebo.

Parameter	Inclusion criteria	Exclusion criteria	Rationale
Outcomes	Outcomes relevant to clinical efficacy and safety which were reported in the LUME-Lung 1 study, including: • OS • PFS • ORR • AEs Additional inclusion criteria applied during feasibility assessment: • Study must report relevant data from at least one outcome that has been reported for other studies, thus enabling a comparison across treatments	• Study protocols without outcome data presented • Studies with only patient baseline characteristics reported	We considered outcomes for which an MTC comparing nintedanib + docetaxel with other second-line treatments was feasible, and only included studies with published results for these outcomes.
Study design	Randomised controlled trials (RCTs) only	Non-RCTs Pooled analyses of RCTs	RCTs provide the highest quality clinical trial data.
Language restrictions	Any language		To minimise bias, RCTs published in languages other than English were included in the search, but no relevant non-English language papers were identified
Date	2000 onwards If a study is an abstract only (for example, from a conference), it was only included if it was published in 2011 or onwards	Primary studies published prior to 2000, systematic literature reviews published before 2010 and conference abstracts published prior to 2011 were also excluded	Limiting the review to studies published from 2000 enabled us to focus on the latest trials evaluating the second-line treatment of NSCLC that reflect current clinical practice and patient populations. Conference abstracts were limited to those presented in 2011 onwards, as full text publications of earlier abstracts reporting on studies of a high quality would be expected to have been published. Systematic reviews were limited to those published in the previous 4 years, as these were used only to identify additional relevant primary research papers and therefore needed to be as up-to-date as possible.

AE=adverse event; NSCLC= non-small-cell lung cancer; ORR=overall response rate; OS=overall survival; PFS=progression-free survival; RCT=randomised controlled trial
Source: Table 25 of the CS[1]

Appendix 3: Eligibility criteria for patient inclusion into LUME-Lung 1

Table 46 Inclusion and exclusion criteria for selection of the trial population in LUME-Lung 1

Eligibility criteria for LUME-Lung 1	
Inclusion criteria	• Male or female patient aged 18 years or older • Histologically or cytologically confirmed, locally advanced and/or metastatic NSCLC of stage IIIB or IV or recurrent NSCLC • Relapse or failure of one first-line prior chemotherapy • At least one target tumour lesion that has not been irradiated within the past 3 months and that can accurately be measured • Life expectancy of at least 3 months • ECOG PS of 0 or 1 • Patient has given written informed consent
Exclusion criteria	• More than one prior chemotherapy regimen for advanced and/or metastatic or recurrent NSCLC • More than one chemotherapy treatment regimen (either neoadjuvant or adjuvant or neoadjuvant + adjuvant) prior to first-line chemotherapy • Previous therapy with other VEGFR inhibitors (other than bevacizumab) or docetaxel for treatment of NSCLC • Persistence of clinically relevant therapy related toxicities from previous chemotherapy and/or radiotherapy • Treatment with other investigational drugs or other anti-cancer therapy, or treatment in another clinical trial within the past 4 weeks before start of therapy or concomitantly with this trial • Radiotherapy (except extremities and brain) within the past 3 months prior to baseline imaging • Active brain metastases or leptomeningeal disease • Radiographical evidence of cavitary or necrotic tumours • Centrally located tumours with radiographical evidence (CT or MRI) of local invasion of major blood vessels • History of clinically significant haemoptysis within the past 3 months • Therapeutic anticoagulation (except low dose heparin) or antiplatelet therapy (except for chronic low-dose therapy with acetylsalicylic acid ≤325mg/day) • History of major thrombotic or clinically relevant major bleeding event in the past 6 months • Known inherited predisposition to bleeding or thrombosis • Significant cardiovascular diseases • Inadequate safety laboratory parameters • Significant weight loss (>10 %) within the past 6 weeks • Current peripheral neuropathy greater than CTCAE grade 2 except due to trauma • Pre-existing ascites and/or clinically significant pleural effusion • Major injuries and/or surgery within the past 10 days prior to randomisation with incomplete wound healing • Serious infections requiring systemic antibiotic therapy • Decompensated diabetes mellitus or other contraindication to high-dose corticosteroid therapy • Gastrointestinal disorders or abnormalities that would interfere with absorption of the study drug • Active or chronic hepatitis C and/or B infection • Serious illness or concomitant non-oncological disease or laboratory abnormality that may increase the risk associated with study participation or study drug administration • Patients who are sexually active and unwilling to use a medically acceptable method of contraception during the trial and for at least 12 months after end of active therapy • Pregnancy or breast feeding • Psychological, familial, sociological, or geographical factors potentially hampering compliance with the study protocol and follow-up schedule • Patients unable to comply with the protocol

Eligibility criteria for LUME-Lung 1	
	• Active alcohol or drug abuse • Other malignancy within the past 3 years other than basal cell skin cancer, or carcinoma in situ of the cervix • Any contraindications for therapy with docetaxel • History of severe hypersensitivity reactions to docetaxel or other drugs formulated with polysorbate 80 (Tween 80) • Hypersensitivity to nintedanib and/or the excipients of the trial drugs • Hypersensitivity to contrast media

CT=computerised (or computed) tomography, CTCAE=Common Toxicity Criteria for Adverse Events. ECOG PS=Eastern Cooperative Oncology Group Performance Status; MRI=magnetic resonance imaging, NSCLC=non-small-cell lung cancer, VEGFR=vascular endothelial growth factor receptor
Source: adapted from Table 9 of the CS[1]

Appendix 4: Clinical endpoints and statistical analyses plan in LUME-Lung 1

Outcomes measured are summarised in Table 47. The TSAP[43] is summarised in Table 48.

Table 47 LUME-Lung 1 Outcomes measured

Endpoint/ assessment	Details
Primary outcome	
PFS	• PFS by central review, using modified RECIST (version 1.0) criteria. Tumour assessments performed at baseline (within 4 weeks of randomisation), and every 6 weeks after first docetaxel administration • PFS was defined as time from date of randomisation to date of disease progression, or to date of death, whichever occurred earlier • Disease progression was defined as: o new lesions, including new lesions in a previously irradiated field o an unequivocal increase in a tumour within a previously irradiated field o an increase in sum of longest diameter (SLD) of the target lesions of 20% from nadir (lowest value measured since treatment started) • Patients who experienced a 30% reduction from baseline in SLD of target lesions and a single instance of a 20% increase in SLD from nadir were considered as having progressed • The primary PFS analysis considered all data collected until the cut-off date for the efficacy analysis, which was the date of the 713th PFS event • The stratified log-rank test was used to test for the effect of nintedanib at the 2-sided alpha-level of 0.05. The log-rank test included the four stratification factors used at randomisation.
Secondary outcomes	
OS	• OS was the key secondary endpoint • OS was defined as the time from date of randomisation to date of death (irrespective of cause of death). Patients who stopped active trial treatment were followed until death or lost to follow-up • Stratified log-rank test and a two-look Lan-DeMets group sequential design with an O'Brien-Fleming-type boundary at a two-sided cumulative 5% level of significance.
PFS by local investigator review	PFS by local investigator review
Tumour response evaluation	Tumour response by central independent review and local investigator assessment, according to modified RECIST (version 1.0) criteria was assessed at baseline (within 4 weeks of randomisation) and every 6 weeks after first docetaxel administration, and categorised into one of the following categories: • complete response (CR) - disappearance of all target lesions and non-target lesions • partial response (PR) - at least a 30% decrease in the SLD of target lesions, taking as reference the baseline SLD • stable disease (SD) - neither sufficient shrinkage of target lesions to qualify for PR nor sufficient increase to qualify as PD; persistence of one or more non-target lesions • progressive disease (PD): o new lesions, including new lesions in a previously irradiated field o an unequivocal increase in a tumour within a previously irradiated field o an increase in SLD of the target lesions of 20% from nadir (lowest value measured since treatment started) • unknown (UNK)

Endpoint/ assessment	Details
	Evaluation of tumour response was based on radiological tumour assessments (CT or MRI) • Tumour images were centrally reviewed by a panel of central independent radiologists. Following radiological review, all patient information was presented to an oncologist. The radiologists and the oncologist were blinded to treatment • Best overall response: o represents the best response a patient has had during their time in the study up until progression, last evaluable assessment in the absence of progression or the start of subsequent anti-cancer therapy. o for patients whose progression event is death, best objective response will be calculated based on data up until the last evaluable RECIST assessment prior to death. • Confirmed objective response o A patient was considered to have a confirmed objective response if a CR or PR was confirmed by imaging no earlier than 28 days after the first occurrence of the response • Disease control o Disease control was defined as a best overall response of CR, PR, or SD recorded at least 6 weeks after the date of randomisation • Time to confirmed objective response o Time from randomisation to first documented confirmed response (CR or PR) recorded at least 6 weeks after the date of randomisation • Duration of confirmed objective response o Time from first documented confirmed response (CR or PR) to progression, or death in the absence of progression • Duration of disease control o Time from randomisation to progression, or death in the absence of progression (whichever occurs earlier) amongst patients with disease control • Change in tumour size o The best change in size (i.e. SLD) of target lesions from baseline was analysed. The maximum SLD decrease from baseline (or the minimum increase in SLD for patients with no reduction in target lesion size) was considered as the best change of the target lesion size in a patient
Clinical improvement	• Clinical improvement quantified the maintenance of body weight and ECOG PS, by measuring the time from randomisation to deterioration in body weight of more than 10% from baseline, and/or increase in ECOG performance score of at least 1 category from baseline, whichever occurred earlier. Patients who died without prior deterioration were considered as having deteriorated at the time of death. • Clinical improvement was analysed until end-of treatment only
HRQoL	• HRQL was measured at the screening visit, at 21-day intervals during treatment, at the end of active treatment, and at the first follow-up visit by the following standardised self-assessment questionnaires: o EQ-5D health status self-assessment questionnaire o EORTC Quality of Life Questionnaire (EORTC QLQ-C30) o EORTC lung cancer specific supplementary module (EORTC QLQ-LC13) • The EQ-5D includes the following two questionnaires, which were analysed descriptively: o Five dimensions of health (mobility, self-care, usual activities, pain/discomfort and anxiety/depression), which are analysed descriptively. Each dimension comprised three levels (no problems, some problems, severe problems)

Endpoint/ assessment	Details
	○ A visual analogue scale (VAS) recorded the respondents self-rated health status on a vertical graduated (0 to 100) scale • The EORTC QLQ-C30 questionnaire includes a global health status/HRQL scale, 5 functional scales, 3 symptom scales, and 6 single items to assess dyspnoea, insomnia, appetite loss, constipation, diarrhoea and financial difficulties. The QLQ-LC13 supplementary module was designed to be used by patients receiving chemotherapy or radiotherapy. It incorporates a multi-item scale to assess dyspnoea, and a series of single items to assess pain, coughing sore mouth, dysphagia, peripheral neuropathy, alopecia and haemoptysis. • The main HRQL endpoints were the time to deterioration for cough (QLQ-LC13, question 1), dyspnoea (QLQ-LC13, questions 3 to 5) and pain (QLQ-C30, Questions 9 and 19) and were evaluated as follows: ○ Distribution of patients with improved, stable, or worsened scores. Improvement was defined as scores that improve by ≥10 points (0 to 100 point scale) at any time during study. Worsening was defined as a worsening in EORTC scores of ≥ 10 points at any time in patients with no improvement. Otherwise, a patient was considered stable. ○ Time to deterioration: defined as time from randomisation to the first 10-point increase (i.e. worsening) from baseline score
Pharmacokinetics	• Pharmacokinetics of nintedanib and of its clinical relevant metabolites BIBF1202 and BIBF1202 glucuronide were determined from blood samples taken at Visit 2 of Treatment Course 2 and 3; both prior to and after the administration of nintedanib.
Safety	• Incidence and intensity of AEs according to the CTCAE version 3.0 • Changes in safety laboratory parameters • The safety analysis included data collected until the safety cut-off date

CR=complete response; CT=computed tomography; CTCAE=Common Terminology Criteria for Adverse Events; ECOG=Eastern Cooperative Oncology Group; EORTC QLQ LC=European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire (Lung Cancer Module); EMA=European Medicines Agency; EORTC= European Organisation for Research and Treatment of Cancer; EQ-5D=European Quality of Life-5 Dimensions; HRQL=health related quality of life; MRI=Magnetic resonance imaging; OS=overall survival; PD=progressive disease; PFS=progression-free survival; PR=partial response; QLQ=quality of life questionnaire; PRO=patient reported outcome; RECIST=Response Evaluation Criteria in Solid Tumours; SD=stable disease; SLD=sum of longest diameters; VAS=visual analogue scale
Source: Table 13 of the CS[1]

Table 48 Trial statistical analysis plan for LUME-Lung 1

Stage of analysis	Description
Futility analysis	A pre-planned futility analysis was to be performed by the central independent DMC after approximately 50% of the PFS events needed for the primary PFS analysis had occurred (~356 events), for the purpose of advising the sponsor as to whether or not the study should continue as planned. The sponsor was blinded to the results of this analysis. Although PFS by central independent review was the primary endpoint, PFS as assessed by the local investigator was used for the futility analysis because of the logistical complexity and the time it took to complete the central independent review of patients' imaging data.
Primary PFS analysis	The primary PFS analysis was to be performed when 713 patients had experienced a centrally independently assessed PFS event (cut-off date 2 November 2010). At this time, a protocol-defined interim analysis of OS was also to be performed. The primary analysis was based on the ITT population.
Final OS analysis	The final analysis of OS was performed when 1,151 patients had died (cut-off date 15 February 2013). At the time of the final OS analysis an updated analysis of all available PFS events was also performed.

DMC=data monitoring committee; ITT=intention-to-treat; OS=overall survival; PFS=progression-free survival
Source: TSAP[43]

Appendix 5: Subgroup and sensitivity analyses in LUME-Lung 1

Subgroup analyses

A number of subgroup analyses for the primary endpoint of PFS assessed by central review and for the secondary outcome OS were pre-specified in the protocol:

- tumour histology (squamous vs non-squamous)
- baseline ECOG PS (0 vs 1)
- presence of brain metastases at baseline (yes vs no)
- prior treatment with bevacizumab (yes vs no)
- sex (male, female)
- age (<65years, ≥65 years)
- race (Asian vs non-Asian patients; information was derived from the race categories as documented on the CRF)
- smoking status (never smoked vs currently smokes/ex-smoker)

The following subgroup analyses were added post-hoc:

- geographical region (Asia, Europe, South Africa; based on country of enrolment)
- best response to first-line therapy (CR/PR/SD, PD, unknown/missing/NA)
- sum of longest diameters at baseline (<7.5cm vs ≥7.5cm)
- time since first-line therapy (<9 months vs ≥9 months)

The company lists a number of baseline characteristics (CS[1], p.77), which were also investigated for subgroup effects. However, neither the protocol[53] or CTR[40] specified whether these were pre-specified or post-hoc analyses. Therefore, the ERG asked for clarification on this issue and the company responded stating that three variables were pre-specified in the protocol[53] and three were included in an amendment to the TSAP[44] for the final OS analysis of LUME-Lung 1.[24]

- presence of liver metastases (yes vs no) (a priori).
- disease stage at diagnosis (<IIIB/IV, IIIB, IV) (a priori).
- concomitant therapy with biphosphonates at baseline (yes vs no) (a priori).
- presence of adrenal metastases (yes vs no) (included in amendment)
- number of metastatic organs at baseline (≤2 metastatic organs, >2 metastatic organs, not centrally reviewed) (included in amendment)
- lactate dehydrogenase (LDH) level at baseline (LDH ≤1, LDH >1) (included in amendment)

Sensitivity analyses

The following sensitivity analyses were pre-specified in the protocol for PFS:

- Analysis using a Cox proportional hazards model fitting the four stratification factors as covariates
- Analysis using a stepwise variable selection method to identify covariates that might be relevant to efficacy
- Analysis replacing actual tumour imaging dates with the originally scheduled dates of radiological assessments
- Analysis using an interval-censoring approach

However, the ERG found that the list of covariates included in the model for the second sensitivity analysis were listed in the CS[1] (pages 74 to 75), but were not all pre-specified in the protocol. The ERG asked for clarification on whether these factors were pre-specified, and the company responded with the following information, stating that only four out of twelve were pre-specified:

- Brain metastases at baseline: predefined strata and also for this analysis in interim TSAP LUME-Lung 1 before unblinding of primary PFS data (a priori).
- Prior treatment with bevacizumab: predefined strata and also for this analysis in interim TSAP LUME-Lung 1 (a priori).
- Body- surface area: (post-hoc).
- Age: Predefined in interim TSAP LUME-Lung 1 in subgroup section (a priori).
- Duration of first-line chemotherapy: (post-hoc).
- Time to first progression: specified in TSAP amendment (post-hoc).
- Time since first histological diagnosis: (post-hoc).
- Presence of ipsilateral metastases in the lung at baseline: (post-hoc).
- Presence of contralateral metastases in the lung at baseline: (post-hoc).
- Bone metastases at baseline: (post-hoc).
- Adrenal metastases at baseline: specified in TSAP amendment (post-hoc).
- Sum of target lesions at baseline: predefined in interim TSAP LUME-Lung 1 (a priori).

The following sensitivity analyses were pre-specified in the protocol for OS:

- Analysis using a Cox proportional hazards model with three of the stratification factors used at randomisation as covariates(ECOG PS at baseline, prior bevacizumab treatment, presence of brain metastases at baseline)
- Analysis using a model which included the stratification factors and the baseline sum of the longest diameters (SLD) of the target lesions (mm) as covariates.

Appendix 6: Methods utilised by the company for making indirect comparisons and mixed treatment comparisons

Mixed treatment comparisons

MTCs were performed using the Markov chain Monte Carlo software package OpenBUGs. The company ran all analyses using fixed-effects models, which assume there is no heterogeneity in relative effects. Random-effects models were also performed if sufficient data was available to estimate a random-effects coefficient, i.e. there were comparisons in the network with evidence from more than one trial. The company chose not to fit random-effects models in situations where the data was sparse, as the estimate of random-effects variation would be too reliant on the choice of prior. The company chose to use vague (non-informative) priors for study and treatment effects, in order to enable a moderate amount of random-effects variation.

Three chains were used to run the analyses, and in all cases, the first 50,000 burn-in simulations were discarded to allow for convergence. Estimates were then obtained from a further 50,000 iterations. The company performed several validation checks to ensure that the models had converged sufficiently and that the estimates produced were reliable. These included examining the Brooks-Gelman-Rubin (BGR) plots and inspection of the values of the Monte Carlo error (Monte Carlo standard error of the mean) to assess validity.

Bucher indirect comparisons

A Bucher indirect comparison is a simple method of comparing two treatments for which there is no direct evidence. In order to obtain an estimate of the treatment effect of A vs C, it is possible to look at two trials which have a common comparator, i.e. Trial 1 considering A vs B, and Trial 2 considering B vs C. The Bucher method does not incorporate random-effects variance from trials elsewhere in the evidence network, i.e. trials which consider C vs D.

Wherever possible, the company conducted Bucher indirect comparisons.

Appendix 7: Assessment of proportional hazards assumption in LUME-Lung 1 Trial

Both indirect comparisons and MTCs require the trials included in the analysis to conform to the assumption of proportional hazards for meaningful and robust results to be generated. This means that the hazard (i.e. the risk of an event occurring at a particular time) is in a constant ratio between the patterns of events observed in the two treatment arms, independent of the time since randomisation. This is a strong assumption which is frequently violated, and it is important that its validity is confirmed prior to carrying out any meta-analysis of outcomes from multiple clinical trials.

In this appraisal a single trial (LUME-Lung 1[24]) compares nintedanib plus docetaxel treatment with erlotinib through a network of trials in which the only links are trials which feature docetaxel monotherapy as a treatment arm. If the proportional hazards assumption is not supported by the LUME-Lung 1[24] trial data, any estimation of the relative effectiveness of nintedanib plus docetaxel vs erlotinib (i.e. a calculated HR) will lack credibility and be effectively meaningless. In this appendix the validity of the proportional hazards assumption in LUME-Lung 1[24] is considered for two key outcomes (PFS and OS) critical to the modelling of cost-effectiveness.

PFS

Figure 19 shows clearly that the PFS survival curve LUME-Lung 1[24] trial arms diverge after about six weeks and then converge and cross after about one year, indicating that the patient PFS advantage from nintedanib plus docetaxel treatment is limited to the first year after treatment. To test the proportional hazards assumption in this data set the HR has been calculated at each event time in either arm of the trial and are shown in Figure 19. If the proportional hazards assumption is supportable the HR values should vary randomly about a horizontal line corresponding to the conventional estimated HR for the trial. Clearly this is not the case as a strong upward trend is apparent following the initial fluctuations (which are due to the small numbers of events recorded in the first few weeks of the trial).

On this basis it must be concluded that any HR estimated from a meta-analysis aimed at comparing PFS outcomes between nintedanib plus docetaxel and any treatment other than docetaxel does not satisfy the essential requirement for validity and reliability, and cannot be considered appropriate for populating a cost-effectiveness model.

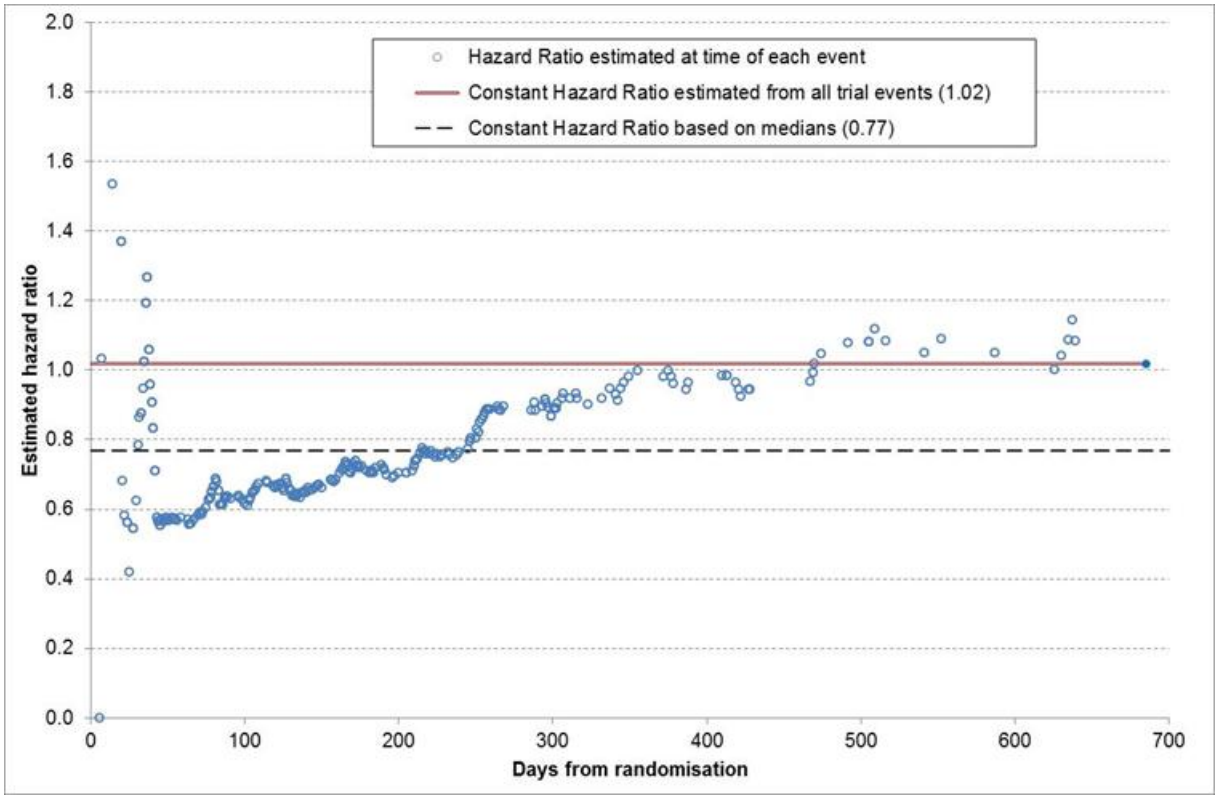


Figure 19 Variation in estimated PFS HR with time in the LUME-Lung 1 clinical trial

OS

Similarly the trend of OS HR estimates also show systematic variations over time (Figure 20): from a peak of 1.1 at four months, falling to less than 0.75 at 400 to 500 days, and increasing thereafter. This pattern is not consistent with the presumption of a steady common HR independent of time, and therefore indicates that the proportional hazards assumption cannot be applied to the LUME-Lung 1[24] OS data set with confidence.

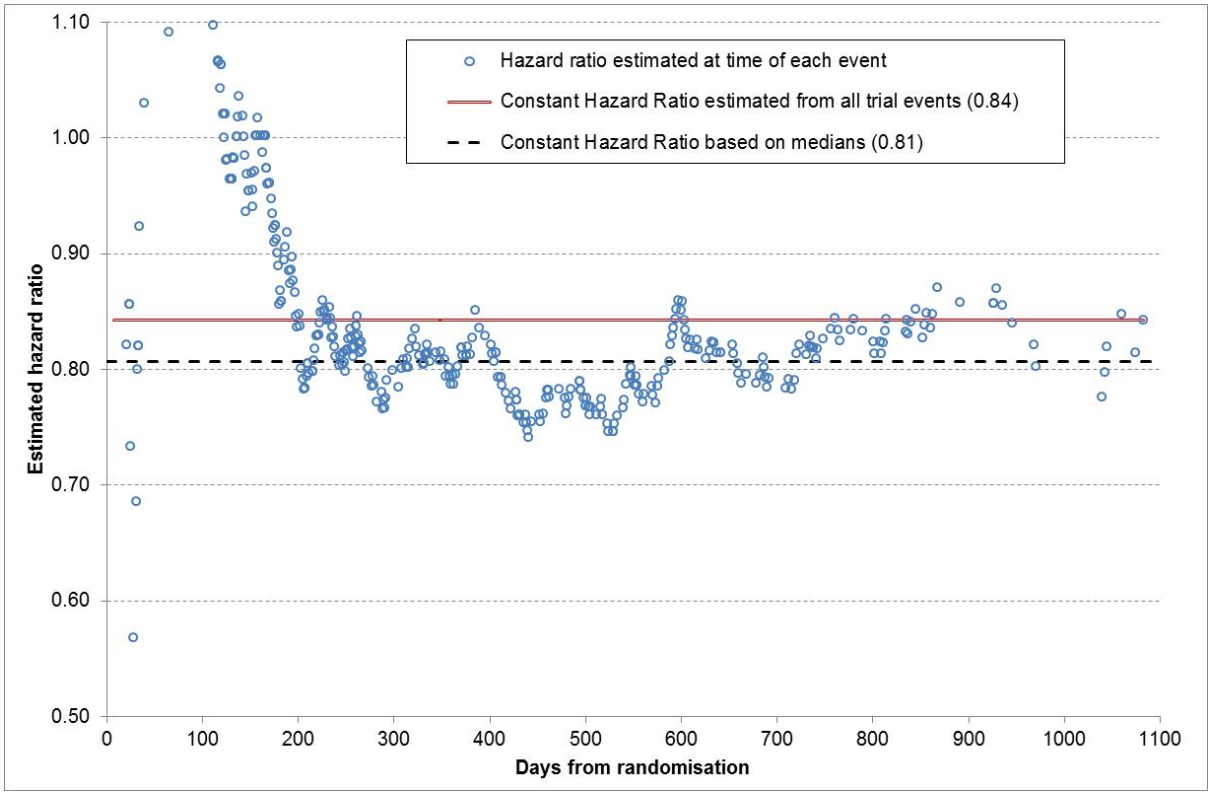


Figure 20 Variation in estimated OS HR with time in the LUME-Lung 1 clinical trial

CONCLUSION

Without a single robust time-invariant HR for either PFS or OS it is not possible to use conventional methods to link and compare the outcomes of patients treated with nintedanib plus docetaxel to patients treated with erlotinib in either the TAILOR[59] or the WSY001[62] trials, regardless of the characteristics of the other trials in the network. Without such comparison meaningful cost-effectiveness analysis involving erlotinib is not possible.

Appendix 8: ERG Revisions to company's model: Nintedanib STA

All revisions are activated by a binary logic switch with 0 = unchanged, 1 (or any non-zero number) = apply ERG modification.

Logic switches are indicated by range variables Mod_*n* where n = 1 - 12. The Mod numbers do not directly match the Table Row numbers, and one Table Row involves applying 2 similar Mod revisions simultaneously.

A menu of revisions/Mod numbers appears on the 'Results' worksheet together with summary results as used to transfer to the ERG report.

ERG Table 14 Row	Binary switch	Associated detail	Implementation instructions
		ERG_Survival_Tables.xlsx	Copy this worksheet as an additional sheet in the model. Ensure that the named ranges ERG_OS, ERG_PFS, ERG_TOT are correctly named in the model.
1. ERG OS estimates	Mod_7	LUME1_OS40-1-3.xlsx	In Sheet 'Survival', Replace formula in cell AW119 by =IF(Mod_7=0,OFFSET(AI119,0,2*ch_OS-2),VLOOKUP(B119,ERG_OS,2)) Copy formula in cell AW119 to range AW120:AW405 Replace formula in cell AX119 by =IF(Mod_7=0,OFFSET(AJ119,0,2*ch_OS-2),VLOOKUP(B119,ERG_OS,3)) Copy formula in cell AX119 to range AX120:AX405

ERG Table 14 Row	Binary switch	Associated detail	Implementation instructions
2. ERG PFS estimates	Mod_6	LUME1_PFS-1-1.xlsx	In Sheet 'Survival', Replace formula in cell M119 by =IF(Mod_6=0,IF(ch_PFS=1,G119,IF(ch_PFS=2,I119,K119)),VLOOKUP(B119,ERG_PFS,2)) Copy formula in cell M119 to range M120:M405 Replace formula in cell N119 by =IF(Mod_6=0,IF(ch_PFS=1,H119,IF(ch_PFS=2,J119,L119)),VLOOKUP(B119,ERG_PFS,3)) Copy formula in cell N119 to range N120:N405

ERG Table 14 Row	Binary switch	Associated detail	Implementation instructions
3. ERG TOT estimates	Mod_8	LUME1_TOT_DocArm_40-1-6.xlsx LUME1_TOT_NinArm_DocTx_40-1-7.xlsx LUME1_TOT_NinArm_NinTx_40-1-8.xlsx	In Sheet 'comp1Model', Replace formula in cell L15 by =IF(Mod_8=0,100%,VLOOKUP(E15,ERG_TOT,4)) Replace formula in cell L16 by =IF(Mod_8=0,L15*(1-rDiscontinuation_nine_Comp1),VLOOKUP(E16,ERG_TOT,4)) Copy formula in L16 to range L17:L301 Replace formula in cell M15 by =IF(Mod_8=0,100%,VLOOKUP(E15,ERG_TOT,3)) Replace formula in cell M16 by =IF(Mod_8=0,IF(OR(Efficacy!$F$43="no",E15<4),M15*(1-rDiscontinuation_doce_Comp1),0),VLOOKUP(E16,ERG_TOT,3))*IF(AND(Mod_4=1,E16>3),0,1) Copy formula in M16 to range M17:M301 In Sheet 'comp2Model', replace formula in cell M15 by =IF(Mod_8=0,100%,VLOOKUP(E15,ERG_TOT,2)) Replace formula in cell M16 by =IF(Mod_8=0,IF(OR(Efficacy!$F$43="no",E15<4),M15*(1-rDiscontinuation_Comp2),0), VLOOKUP(E16,ERG_TOT,2))*IF(AND(Mod_4=1,E16>3),0,1) Copy formula in M16 to range M17:M301

ERG Table 14 Row	Binary switch	Associated detail	Implementation instructions
4. Mid-cycle adjustment	Mod_10	None	In Sheet 'comp1Model', replace formula in cell BE16 by =IF(Mod_10=0,AVERAGE(S15:S16),S15)*cDrugAdmin_doxa_Comp1 Copy formula in BE16 to ranges BE17:BE19, BE21:BE301 Replace formula in cell BE20 by =IF(Efficacy!$F$43="no",IF(Mod_10=0,AVERAGE(S19:S20),S19)*cDrugAdmin_doxa_Comp1,0) In Sheet 'comp2Model', Replace formula in cell AV16 by =IF(Mod_10=0,AVERAGE(S15:S16),S15)*cDrugAdmin_Comp2 Copy formula in AV16 to ranges AV17:AV19, AV21:AV301 Replace formula in cell AV20 by =IF(Efficacy!$F$43="no",IF(Mod_10=0,AVERAGE(S19:S20),S19)*cDrugAdmin_Comp2,0)
5. Cost of treatment doses	Mod_1	DrugCalcs.xlsx Sheet 'Calcs_75mg' LUME1_MeanDoseCostEstimates(adjusted fordose reductions).xlsx Sheet: 'LUME1_DoseLevels_40_3_1'	In Sheet 'UnitCosts', Replace formula in cell H35 by =IF(Mod_1=0,Y37,98.480134%) Replace formula in cell H36 by =IF(Mod_1=0,Y38,99.08405%) Replace formula in cell H37 by =IF(Mod_1=0,Y39,99.08405%) Replace formula in cell I35 by =IF(Mod_1=0,DrugCostCalc!$S$71,37.5) Replace formula in cell I36 by =IF(Mod_1=0,G36*F36*BSA,37.5) Replace formula in cell I37 by =IF(Mod_1=0,DrugCostCalc!$S$88,37.5)
	Mod_11	LUME1_MeanDoseCostEstimates(adjusted fordose reductions).xlsx Sheet: 'LUME1_DoseLevels_40_2_1'	In Sheet 'UnitCosts', Replace formula in cell K33 by =IF(Mod_11=0,I33*J33,1409.920164)
6. Febrile neutropenia cost	Mod_2	FNcost.xlsx	In Sheet 'AdverseEvents', Replace formula in cell I195 by =IF(Mod_2=0,SUM(H196:H209),7352.543797)

ERG Table 14 Row	Binary switch	Associated detail	Implementation instructions
7. Monitoring cost	Mod_9	None	In Sheet 'ResourceUse', Replace formula in cell E91 by =100%*IF(Mod_9=0,1,0) Replace formula in cell F91 by =timeDaysInCycle/(2.5*timeDaysInMonth)*IF(Mod_9=0,1,0) Replace formula in cell M67 by =IF(Mod_9=0,0,100%) Replace formula in cell N67 by =timeDaysInCycle/(2.5*timeDaysInMonth)*IF(Mod_9=0,0,1) Replace formula in cell P67 by =IF(Mod_9=0,UnitCosts!$E$73,0) Replace formula in cell Q67 by =IF(Mod_9=0,M67*N67*P67,0)
8. Discounting method	Mod_3	None	In Sheet 'comp1Model', Replace formula in cell H16 by =IF(Mod_3=0,H15/(1 + discCc),1/(1 + iDiscCost)^INT(F16)) Replace formula in cell I16 by =IF(Mod_3=0,I15/(1 + discHc),1/(1 + iDiscHealth)^INT(F16)) Copy range H16:I16 to rows 17-301 In Sheet 'comp2Model', Replace formula in cell H16 by =IF(Mod_3=0,H15/(1 + discCc),1/(1 + iDiscCost)^INT(F16)) Replace formula in cell I16 by =IF(Mod_3=0,I15/(1 + discHc),1/(1 + iDiscHealth)^INT(F16)) Copy range H16:I16 to rows 17-301
9. Disutility of fatigue	Mod_5	None	In Sheet 'Utilities' Replace formula in cell E66 by =(SUMPRODUCT(Utilities!$E$50:$E$62,AdverseEvents!$E$34:$E$46) + IF(Mod_5=0,0,(-0.326-E55)*AdverseEvents!E39))/AdverseEvents!$E$48 Replace formula in cell E66 by =(SUMPRODUCT(Utilities!$E$50:$E$62,AdverseEvents!$F$34:$F$46) + IF(Mod_5=0,0,(-0.101-E55)*AdverseEvents!F39))/AdverseEvents!$F$48

ERG Table 14 Row	Binary switch	Associated detail	Implementation instructions
10. Stable disease costs	Mod_12	None	In Sheet 'Resource Use' Replace formula in cell M65 by =IF(Mod_12=0,100%,0%) Replace formula in cell N65 by =IF(Mod_12=0,timeWeeksInCycle,0) Replace formula in cell M78 by =IF(Mod_12=0,100%,0%) Replace formula in cell M78 by =IF(Mod_12=0,1,0)
11. Docetaxel ≤4 cycles	Mod_4	None	See details for #3