Olaratumab in combination with doxorubicin for treating advanced soft tissue sarcoma

A Single Technology Appraisal

Produced by	Peninsula Technology Assessment Group (PenTAG) University of Exeter St Luke's Campus, Heavitree Road, Exeter, EX1 2LU
Authors	Irina Tikhonova,[1] *Research Fellow* Tracey Jones-Hughes,[1] *Research Fellow* James Dunham,[1] *Graduate Research Assistant* Fiona Warren,[1] *Lecturer* Sophie Robinson,[1] *Information Scientist* Peter Stephens,[2] *Consultant Medical Oncologist* Martin Hoyle,[1] *Associate Professor* [1] Peninsula Technology Assessment Group (PenTAG), Exeter, UK [2] Royal Devon and Exeter NHS Foundation Trust, Exeter, UK
Correspondence to	Tracey Jones-Hughes, South Cloisters, St Luke's Campus, Heavitree Road, Exeter, EX1 2LU
Date completed	09/03/2017
Source of funding	This report was commissioned by the NIHR HTA Programme as project number
Declared competing interests of the authors	None
Acknowledgments	We acknowledge the excellent administrative support of Sue Whiffin and Jenny Lowe (both of University of Exeter).
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:	Tikhonova I, Jones-Hughes T, Dunham J, Warren, F, Robinson S, Stephens P, Hoyle M. Olaratumab in combination with doxorubicin for treating advanced soft tissue sarcoma: A Single Technology Appraisal. Peninsula Technology Assessment Group (PenTAG), 2017.

Contributions of authors

Irina Tikhonova	Led the critique of the review of economic evaluation studies and wrote the Cost-effectiveness and End-of-life sections. Contributed to the critique of the economic model.
Tracey Jones-Hughes	Led the critique of the company's decision problem and clinical effectiveness evidence. Wrote the Summary, Background, Decision problem, Clinical effectiveness and Overall conclusions. Compiled the report. Provided overall project management.
James Dunham	Led the critique of the economic model and wrote the respective section.
Fiona Warren	Led the critique of the network meta-analysis and wrote the respective section.
Sophie Robinson	Led the critique of the company's literature searching for this submission. Wrote the review of the literature searches for the report. Contributed to the writing and editing of the report.
Peter Stephens	Provided clinical advice on soft tissue sarcoma and its management within the NHS. Reviewed and revised a draft version of the report.
Martin Hoyle	Provided occasional advice on the critique of the economic evaluation, commented on drafts of the report and is the guarantor of the report.

Contents

List of tables

List of figures

Abbreviations

AESI	Adverse Events of Special Interest
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
BSA	Body Surface Area
BSC	Best Supportive Care
CDF	Cancer Drug Fund
CEA	Cost Effectiveness Analysis
CEAC	Cost Effectiveness Acceptability Curve
CRD	University of York Centre for Reviews and Dissemination
CI	Confidence Interval
CrI	Credible Interval
CRUK	Cancer Research UK
Doc	Docetaxel
Dox	Doxorubicin
ECOG	Eastern Cooperative Oncology Group
EED	National Health Service Economic Evaluations Database
eMIT	Electronic Market Information Tool
EPAR	European Public Assessment Report
EQ5D	EuroQol-5 Dimensions
ERG	Evidence Review Group
ESMO	European Society for Medical Oncology
FDA	Federal Drug Administration
GIST	Gastrointestinal stromal tumour
HR	Hazard Ratio
HRQoL	Health-Related Quality of Life
ICER	Incremental Cost-Effectiveness Ratio
Ifo	Ifosfamide
IfoDox	Ifosfamide in combination with Doxorubicin
ITT	Intent-to-treat
IV	Intravenous
KM	Kaplan Meier
LYG	Life-Years Gained
MedDRA	Medical Dictionary for Regulatory Activities
NICE DSU	National Institute for Health and Care Excellence Decision Support Unit
Ola	Olaratumab
OlaDox	Olaratumab in combination with Doxorubicin
ORR	Objective Response Rate
OS	Overall Survival
OSA	One-way Sensitivity Analysis
PFS	Progression Free Survival
PPS	Post-Progression Survival
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PSA	Probabilistic Sensitivity Analysis
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit
QALY	Quality-Adjusted Life Year
RCT	Randomised Controlled Trial
RECIST	Response Evaluation Criteria in Solid Tumours
SAE	Serious Adverse Event
SPC	Summary of Product Characteristics
STS	Soft Tissue Sarcoma
TEAE	Treatment-Emergent Adverse Event
TSD	Technical Support Document
WHO	World Health Organisation

1. Summary

1.1 Critique of the decision problem

The company defined the population to include only adults who have not been previously treated with doxorubicin, and had no prior line of systemic treatment. This differs slightly to the NICE scope, where the population specified have not been previously treated with doxorubicin). This is in keeping with Ola being a first line treatment.

The intervention in the decision problem was olaratumab in combination with doxorubicin, as in the NICE Scope.

The comparators described in the company submission match those stated in the NICE scope, which are Dox monotherapy and IfoDox.

The outcomes in the company submission match those in the Scope. With regard to the outcome 'response rates' as stated in the scope, the company have investigated objective response rates (ORR), which is the proportion of patients achieving a best overall response of partial response (PR) or complete response (CR).

Although the NICE scope did not define any subgroups, preplanned investigations by the company include stratification factors of PDGFRα, number of lines of previous treatment, histological tumour type and ECOG score.

1.2 Summary of the clinical effectiveness evidence submitted by the company

The primary focus of the company's submission was the JGDG study. Patients were randomised to OlaDox (N=66) or to Dox monotherapy (N=67). Baseline characteristics were reported as being balanced between arms, other than OlaDox having slightly more females than the Dox arm (a difference of 7 patients, 9.9%). Outcome results were as follows:

Progression free survival

OlaDox was numerically superior with an investigator-assessed improvement in median PFS of 2.5 months in the OlaDox arm over Dox alone (6.6 months [95% confidence interval (CI): 4.1, 8.3] versus 4.1 months [95% CI: 2.8, 5.4], and was statistically significant at α 0.1999 (as opposed to the more usual significance level of 0.05).

Secondary endpoints

The median overall survival (OS) was 26.5 months (95% CI: 20.9, 31.7) in the OlaDox arm and 14.7 months (95% CI: 9.2, 17.1) in the Dox arm giving a median OS increase of 11.8

months. The stratified hazard ratio (HR) gives a statistically significant result of 0.463 (95%CI 0.301 to 0.710).

Further analysis was performed on the first-line population (65% of participants) which also displayed improved OS in the OlaDox arm. The median OS was 29.1 months (95% CI: 16.3, NE) in the OlaDox arm and 14.7 months (95% CI: 8.0, 18.7) in the Dox arm (stratified HR = 0.47 [95% CI: 0.27, 0.81]; p = 0.0051).

Eli Lilly report PR, CR and ORR following both investigator and independent assessment. The results were similar with no statistically significant result seen in either case.

No data were collected for the health-related quality of life (HRQoL).

Adverse events

More patients in the Dox arm than in the OlaDox arm discontinued study treatment for reasons of adverse events (AEs) (16.4% vs. 7.6%).However, the median number of cycles of doxorubicin received was greater in the OlaDox arm (median: 21.3 weeks or approximately 7.1 cycles) compared to the Dox arm (median: 12.3 weeks or approximately 4.1 cycles).

For OlaDox, the most common Grade 3 to 4 TEAEs was neutropenia at 53%, as compared to 33% for Dox. Febrile neutropenia was, however, similar between arms (OlaDox 13% vs. Dox 14%).

Network meta-analysis

A network meta-analysis (NMA) was performed in order to compare the effectiveness (PFS and OS) of OlaDox with IfoDox. The NMA includes data from six studies including seven treatments. The main analyses included data from patients who received OlaDox as a first line treatment only, and produced a hazard ratio for overall survival that was significantly lower for OlaDox compared with Dox monotherapy and for one of the four IfoDox treatments. For two of the IfoDox treatments, there was a borderline significant reduction in the HR for OS (p-values 0.06) and for the fourth IfoDox treatment there was no significant difference in HR. Sensitivity analyses using any line of OlaDox produced comparable results.

Result for equivalent analyses using PFS found no statistically significant improvement in PFS when comparing OlaDox against Dox monotherapy and the two IfoDox regimens. Only one result produced a borderline significant result, which was the stratified ITT analysis comparing OlaDox against Dox monotherapy. Some further results from the main analysis of first line treatment only are presented, stating that for OS, OlaDox had the highest probability of being the best treatment (85.2%) with an associated SUCRA score of 0.97. For PFS, the

probability of OlaDox being the best treatment was 43.5%, with one of the IfoDox treatments having a probability of being the best treatment of 52.8%.

The predicted best response rates from the NMA are presented for three categories of response: (i) stable disease, partial response, or complete response; (ii) partial response or complete response; (iii) complete response only.

The intervention with the predicted best response rate across all three categories was Dox (75mg/m^2) + Ifo (10g/m^2). Cumulative rankograms for all treatments were provided as additional information; these indicated that the Dox (75mg/m^2) + Ifo (10g/m^2) intervention had the highest probability (0.77) of being the `best' intervention, whereas OlaDox was the third best out of six treatments, with a probability of 0.069.

The OR for discontinuation due to adverse events was below 1 for all comparisons of OlaDox versus any of the other five treatments in the decision model, and was statistically significantly lower for three of the five comparisons, with weak evidence for a lower OR in one of the comparisons.

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

The method for the systematic review was poorly described by the company. However, their searches were adequate and the ERG concludes that the company did not miss any evidence.

The primary focus of the company's submission was the JGDG study. This was generally an appropriately-designed RCT, although a small population due to the rarity of STS. The patients were representative of the UK population, with 65% receiving first-line therapy.

As noted in the company submission, the protocol for doxorubicin monotherapy differs from typical UK practice. The trial population were able to receive up to 8 cycles of olaratumab/doxorubicin whereas a maximum of 6 cycles of doxorubicin monotherapy typically used in UK clinical practice

The use of subsequent therapies following treatment assignment was permitted (45% of participants in the dox arm subsequently received Ola monotherapy), which may be a limitation to the study design due to confounding outcome measures.

The open-label design of the trial, although unavoidable as the treatments generally require different levels of medical intervention, increases the risk of bias.

Network Meta-analysis

The NMA includes data from six studies with some variation in patient demographic characteristics.

The six studies include seven treatments, six of which are included in the decision problem. The network is connected through a common reference intervention, Dox monotherapy, included in all except one of the trials. The final trial compares two different IfoDox regimens, one of which is connected to the network via Dox monotherapy, through a connection with one of the same IfoDox regimens used in a different trial. Each pairwise comparison in the network has only one trial providing data.

With regard to the survival analysis modelling, the company opts for the fractional polynomials method. However, this method requires individual patient data, which was only available for the JGDG trial. Hence, the authors used a method to reconstruct individual patient data using digitised KM plots and numbers at risk/numbers of events for the remaining studies. The HRs and confidence intervals that were reported in the publications alongside those derived from the reconstructed data were in general very close. However, no sensitivity analyses were reported. Also, the quality of the KM plot images is not discussed, nor the perceived quality of the reconstructed data. The use of fixed effect models is appropriate due to the nature of the network.

1.4 Summary of the cost-effectiveness evidence submitted by the company

1.4.1 Company's systematic review of economic evaluations

The cost-effectiveness systematic review identified 19 publications that reported cost-effectiveness and/or cost-utility analyses in STS. The only publication pertaining to interventions in this submission was a publication by Guest et al. (2013) comparing the cost-effectiveness of IfoDox to trabectedin.

Since HRQoL data was not reported in the JGDG study, Lilly conducted a systematic literature review to identify published health-state utility estimates. The company identified the only prior NICE TA in this area, trabectedin TA185.(2010)[1] Three publications, Reichardt et al. (2012),[2] Amdahl et al. (2014)[3] and Delea et al. (2014),[4] fulfilled the requirements of the NICE reference case since quality of life was measured directly from STS patients using the EQ-5D or EORTC-QLQ-C30 mapped to EQ-5D, with valuation based on a UK population.

The study by Reichardt[2] reports health state utility values (HSUVs) by line of chemotherapy and health state, progression-free and progressed (Table 60, p 142). HSUVs from this study were, therefore, considered most appropriate for inclusion in the model, although the study had some limitations. The values were based on a small sample (the number of assessments in each health state ranged from 12 to 35) and represented a mixed population of patients with STS (n=94) or bone sarcoma (n=20), although the majority of assessments were for STS patients. In addition, the study's requirement that *patients provide a response*

may have excluded patients with early disease progression. The company acknowledged that this may have resulted in higher utility values than would be expected for all patients with disease progression.

Lilly did not conduct any separate searches for adverse event literature. Resource use, costs and health utility estimates relevant to mSTS patients experiencing adverse events were identified as part of the systematic review of economic evidence (including economic evaluations, resource use, cost and utility estimates relevant to mSTS).

1.4.2 Company's submitted economic evaluation

1.4.2.1 Methods

The company presented a model-based economic evaluation to address the decision problem.

The model presented by Eli Lilly is a cohort-based partitioned survival model. It is informed by the JGDG study, a systematic review of country-specific resource use, costs and utilities, the company's own observational study of resource cost and use, and multiple oncologists' and external consultants' advice on STS and model implementation.

The company's evidence is submitted with the intention that Olaratumab would be used a first-line treatment. There are 3 health states in the model; progression free survival, post-progression survival and death, with those experiencing disease progression having up to 3 further lines of therapy available and best supportive care (BSC).

In the base case analysis, patients enter the model upon commencement of receiving first line treatment, OlaDox or the comparator, Dox/IfoDox. The patients can then remain progression free, during which time they continue their first-line therapy until the completion of treatment, come off it for another reason (toxicity, physician/patient decision) or their disease progresses, whichever is soonest. Whilst Dox has a capped number of administrations, Ola administrations were taken until disease progression in the baseline study, which the model mirrors. Those who have progressed are then placed in one of the post-progression survival (PPS) lines of treatments.

The proportions of patients in each state at time t is calculated through the use of using PFS and OS data, and hazard is a function of t. PFS survival in the base-case is directly estimated from JGDG data using a Kaplan-Meier fit, whereas OS survival data are calculated from JGDG data and external data to provide parametric fittings beyond the trial timeline. PPS survival is then the difference between PFS and OS.

The cycle length is one week, which does not directly correspond to a treatment cycle. The treatment cycle proposed is a 21-day cycle in which Dox is administered once (day 1) for all

arms and Ola is administered on days 1 and 8 for the OlaDox arm. Due to the short cycle time in relation to model horizon (25 years in the base case), no half-cycle correction is implemented. Despite Ola monotherapy being offered to Dox monotherapy patients post-treatment and post-progression in the JGDG trial, the company's base case does not allow for Dox patients to receive Ola monotherapy.

In the model, the perspective on costs was the NHS and personal social services perspectives, and the perspective on health effects was the direct health effects on patients, in accordance with the NICE reference case.

The baseline model time horizon was 25 years, which is justified as a lifetime horizon based on the OS data from the JGDG study extrapolated to beyond the censoring.

Treatment effectiveness was estimated from the JGDG trial and from post-hoc analyses conducted on the data collected.

The economic model used the following clinical endpoints:

- Overall survival (OS), the time from entering the model until death from any cause.
- Progression free survival (PFS), the time from entering the model until disease progression
- Post progression survival (PPS), the time from disease progression until death.

PFS was modelled using a Kaplan-Meier approach. Investigating parametric approaches was explored in the sensitivity analysis.

1.4.2.2 Results

OlaDox compared to Dox monotherapy is estimated to cost [redacted] per QALY (ICER), and OlaDox compared to the IfoDox combination is estimated to cost [redacted] per QALY.

ICERs from the base case PSA were [redacted] for OlaDox versus Dox and [redacted] for OlaDox versus IfoDox. OlaDox had a [redacted] probability of being cost-effect against Dox monotherapy and a [redacted] chance of being cost-effect against IfoDox at list price.

With regard to deterministic sensitivity analysis, most of the changes had relatively small impacts on the ICERs (within 15% of the baseline estimates). The changes which had the greatest impacts were the choice of parametric survival functions in the OlaDox vs Dox arm (not reported for IfoDox), changing PPS utilities values and changing drug administrations in the IfoDox arm. Using a Weibull or Gompertz survival function had the greatest impact, increasing the ICER to [redacted].

The company's UK specific scenario analysis shows that ICERs change from [redacted] for OlaDox vs Dox and [redacted] for OlaDox vs IfoDox.

1.5 Summary of the cost-effectiveness evidence submitted by the company, and the ERG's critique of the cost-effectiveness

Eli Lilly conducted a systematic review for cost-effectiveness evidence. The searches did not identify any studies directly relating to the decision problem. The company, therefore, developed a *de novo* economic model to address the decision problem.

In their model, Eli Lilly consider all treatments from the NICE Scope, and their base-case ICERs are:

- OlaDox vs. Dox [redacted] per QALY
- OlaDox vs. IfoDox [redacted] per QALY

In Eli Lilly's analysis, olaratumab is considered a 1st-line therapy which is not in line with NICE Scope.

1.5.1 Model checking

We checked Eli Lilly's model and found no major errors.

1.5.2 Model structure

Eli Lilly developed a three-state partitioned survival model with a standard model structure that has been used in numerous HTAs: pre-progression, post-progression and death. In this model, patients receive study treatments until progression, occurrence of severe adverse events or other causes leading to discontinuation of treatment. In progressive disease, patients receive up to 4 lines of subsequent therapy.

We consider the overall model structure appropriate.

1.5.3 Method of PFS estimation

Progression free survival (PFS) for OlaDox vs. Dox was modelled using Kaplan-Meier data from the JGDG trial. No extrapolation of PFS was performed since PFS data was mature.

The ERG agrees with the choice of KM curves for the base-case analysis.

For OlaDox vs. IfoDox comparison, PFS was derived from a network meta-analysis, based on fractional polynomials (Section 4.2.1, p 48). To estimate hazard function, Eli Lilly used median estimates of the coefficients of fractional polynomials from the NMA, while in our base case we assume mean estimates, which constitutes Item 2 in the PenTAG base case (Section 5.3.5.2, p 137.).

1.5.4 Method of OS estimation

1.5.4.1 OlaDox vs. Dox

Due to a small number of patients and events in the 1st-line subgroup of JGDG study (40/21 and 47/36 patients/events in the OlaDox and Dox arms, respectively), the "arm together" approach was used, i.e., parametric survival models were fitted to the ITT dataset with line of therapy as a covariate.

Since OS data from the pivotal trial was immature, Lilly extrapolated patient survival up to 25 years after mean age at the diagnosis of advanced STS (which is about 58.5 years in JGDG trial). The company then used external data from Van Glabbeke et al. (1999)[5] for validation of extrapolated OS for Dox patients.

We consider the patients in JGDG trial to be similar to those in clinical practice. However, as explained in Section 5.3.5.1.2, we disagree with Lilly's selection of gamma function as the best fit based on Van Glabbeke et al. (1999)[5] used for external validation. The patient population in that study was substantially younger than the population in JGDG (with 75.5% of patient $\leqslant$ 60 years old), which might overestimate the long-term survival of the patient population relevant to this appraisal, and bias cost-effectiveness results.

In our base-case, we utilised a long-normal distribution, presented in the company's submission among other candidate models, which, based on our expert's opinion, provides clinically reasonable prediction of 5 and 10 year survival after diagnosis of advanced STS. This constitutes Item 1 in the PenTAG base case (Section 5.3.5.1.2, p 128).

In the company's base case, no treatment effect was assumed after 32 months (of note, the length of the observational period in JGDG trial was 47 months); alternative assumptions (of tapering over 12 months period, and of treatment effect as observed in the trial) were examine in sensitivity analyses. We identified an error in the model related to one of those analyses, which is described in Section 5.3.5.1.2 (p 128). The error has no effect on the base case.

1.5.4.2 OlaDox vs. IfoDox

For the OlaDox vs. IfoDox comparison, the company used OS curve derived from a NMA. As with PFS, Lilly used median values of the coefficients of fractional polynomials from the NMA to estimate overall survival. The effect of using mean estimates is explored in our base case (item 2 in the PenTAG base case, (Table 1, p 25 and Table 70, p 157).

1.5.5 Costs

1.5.5.1 Drug acquisition and administration costs

Olaratumab and G-CSF (filgrastim) are dosed based on patient's weight. The doses of other drugs, considered in this appraisal, are given proportional to body surface area (BSA). Lilly assumed a mean weight of 77.3 kg referencing GeDDiS trial which was conducted mainly in the UK. A mean BSA of 1.91m^2 was taken from Health Survey of England (2013). We could not verify the mean weight in the reference provided by the company and assumed the mean weight of 82.5 kg from JGDG trial, which constitutes item 3 of the PenTAG base case (Table 1, p 25 and Table 70, p 157).

1.5.5.1.1 OlaDox

In the JGDG trial, olaratumab was given intravenously (IV) on days 1 and 8 of 21-days treatment cycles until disease progression. Dox was administered IV once per 21-day treatment cycle for up to 8 cycles or disease progression. Even though in the UK practice Dox is usually given up to 6 treatment cycles, or the total dose up to 450 mg/m^2, for consistency, we cost Dox treatment on the basis of the pivotal trial. Importantly, the mean total cumulative dose of Dox reported in JGDG trial was less than 450 mg/m^2 (Section 5.3.7.1.1, p 146).

1.5.5.1.2 IfoDox

Since Eli Lilly could not identify a study which would report the IfoDox regimen most commonly used in the UK, Dox 60mg/m^2 + Ifo 9g/m^2 , they assumed that the regimes, Dox 60mg/m2 + Ifo 9g/m2 and Dox 75mg/m^2 + Ifo 10g/m^2, have similar efficacy. Our clinical expert advised us that, in terms of efficacy, these regimens would not be significantly different.

There was no data available to estimate the extent of dose reduction on IfoDox. Therefore, Eli Lilly modelled the planned dose of IfoDox, while costing of OlaDox was bases on the dose reported in JGDG trial. A sensitivity analysis conducted by the company showed that 20% decrease in IfoDox dose increased the ICER only slightly (by about £1,000).

Acquisition costs for Ifo and Mesna in Lilly's model were based on BNF 2015. Our analysis incorporates current prices for these drugs from BNF February 2017. These are items 4 and 5 in our base case (Table 1, p 25 and Table 70, p 157).

1.5.5.1.3 Vial sharing

In the company's model, no vial sharing was assumed for all intravenously administered drugs. Therefore, all calculations were based on the assumption of full drug wastage. We consider this assumption reasonable since, with a rare cancer, vial sharing is unlikely.

1.5.5.1.4 Availability of vial sizes for Ola

Conditional marketing authorisation has been granted only for the 500mg vial of olaratumab. The company, however, assumed availability of 500mg and 190mg vial sizes in anticipation of marketing authorisation for the 190mg vial of Ola. In our base-case, we assume that only 500 mg vial of olaratumab is available. This constitutes item 6 of the PenTAG base case

1.5.5.1.5 Drug administration costs

We believe that administration costs were underestimated in Lilly's model. In particular, it was assumed that the length of OlaDox administration (with premedication for both drugs) is less than 2 hours. According to our expert's opinion, OlaDox administration may take 2.5-3 hours (Section 5.3.7.2, p 149). This assumption constitutes item 7 in our base case (Table 1, p 25 and Table 70, p 157).

1.5.5.2 Disease management costs

Disease management costs including disease monitoring, tests and other health state costs were informed by Lilly observational study. We consider assumptions in sensitivity analyses.

1.5.5.3 Adverse event costs

Costs of adverse event of Grade $\geqslant$3 were calculated by combining the proportion of events likely to require hospitalisation based on data from the JGDG trial with estimates of costs per event (including outpatient visit costs).

In the base case, the costs of hospitalisation were estimated from NHS reference costs, which reflect the length of stay specific to UK practice (but not specific to STS patients). The costs of AEs were accounted for in the first year of the model.

We are generally satisfied with the approach taken by Lilly to costing of treatment associated with AEs. However, we identified some inconsistencies in unit costs reported in the submission, which have negligible effect on the ICERs.

1.5.6 Utilities

As the JGDG trial did not collect any health-related quality of life (HRQoL) data, the company conducted a systematic literature review to identify published health-state utility estimates. Three studies were judged to provide consistent utility estimates. Estimates used in the company's model were either directly measured using the EQ-5D or by mapping the EORTC-QLQ-C30 to the EQ-5D.

The utility values of 0.72 and 0.56 were assumed in the base case for pre- and post-progression states, respectively. They were based on the study by Reichardt et al. (2012) which reported health state UV for patients with mSTS and metastatic bone sarcoma with

favourable response to chemotherapy. The mean age of patients at metastatic disease diagnosis was 49.5 (SD = 17.1), while in the JGDG trial the mean age of patients from OlaDox arm was 56.8 (SD=12.53) and 58.3 (SD=12.50) in the Dox arm. The company acknowledged that the study selection criterion may have resulted in higher utility values than would be expected for all patients. We believe that age heterogeneity may bias the results further since patients in Reichardt et al. (2012) were substantially younger than in JGDG study. However, we are unable to find a source of utilities that is clearly superior.

1.5.7 End of Life criteria

Eli Lilly argues that their presented evidence supports inclusion into NICE's End of Life category. The ERG, however, notes that using median life expectancy, on which Lilly's analysis was based, is an incorrect interpretation of the NICE criteria, and that mean life expectancy should have been considered in their EoL analysis.

In their base-case their comparators Dox/IfoDox have an (undiscounted) mean life expectancy of 2.32/2.67 years, respectively (calculated from Lilly model). As such, their base case would not qualify for the EoL category and the standard £20,000-£30,000 ICER threshold would be applicable.

In the ERG base case, the mean undiscounted life expectancy for Dox is 1.83 years (IfoDox unchanged). Based on the above criteria, the OlaDox vs. Dox arm would then qualify for End of Life.

1.5.8 Eli Lilly's model results

In the OlaDox vs. Dox comparison, the OlaDox arm accrues the most QALYs (2.11), with 0.46 in the pre-progression state and 1.66 in progressed disease. Dox has 1.22 QALYs, with 0.36 QALYs in progression free and 0.86 QALYs in progressed disease (PD).

In the OlaDox vs. IfoDox comparison, the OlaDox arm accrues the most QALYs (2.18), with 0.63 in the pre-progression state and 1.55 in progressed disease. IfoDox treatment resulted in 1.43 QALYs, with 0.56 QALYs in progression free and 0.86 QALYs in progressed disease (PD).

Costs in PFS are split into drug acquisition and administration, disease management and treatment of adverse events costs. The OlaDox arm has the largest costs in almost all these categories, totalling [redacted] in the comparison with Dox, and [redacted] when compared to IfoDox, while the total costs in Dox and IfoDox arms are [redacted] and [redacted], respectively.

Costs in PD are driven by drug costs, and they are similar across treatments: [redacted] in OlaDox arm and [redacted] in Dox arm (OlaDox vs. Dox); [redacted] in OlaDox and [redacted] in IfoDox arm for the relevant comparison.

1.5.9 Critique of Eli Lilly's analysis

In this section, we highlight our key areas of disagreement with Eli Lilly's analysis. As a result of our critique of their model, we have developed PenTAG base case ICERs (Table 1, p 25 and Table 70, p 157) by adjusting the following items in Lilly's model:

1. Parametric survival function for OS in Dox and OlaDox arms (we use log-normal function)
2. Coefficients of fractional polynomials estimated in NMA for OlaDox vs. IfoDox comparison (we use mean values instead of medians)
3. Ifo and Mesna prices (we use current prices)
4. Patients' mean weight (we use patients' mean weight from JGDG trial)
5. Availability of vial sizes of olaratumab (we assume that only 500mg vial is available)

Drug administration costs (we base costing on corrected HRG codes)

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

1.6.1 Strengths

- Multicentre, appropriately randomised design of the RCT JGDG
- The population recruited to study JGDG was representative of the typical UK patient population
- Eli Lilly's analysis was clearly described in their report.
- The structure of Lilly's model is appropriate and consistent with the natural history of mSTS.
- We found no major errors in the model code, although we found a number of errors of minor importance.
- The clinical effectiveness evidence for OlaDox vs. Dox is of good quality, taken from an RCT.

1.6.2 Weaknesses and areas of uncertainty

- The JGDG trial is a Phase 2 study intended to provide preliminary data, therefore the population is small and the significance level set at 0.1999 (rather than the more conventional 0.05 for larger trials). The increased significance level reduces the region of acceptance i.e., there is more likelihood of rejecting the null hypothesis (that there is no significant difference between arms).
- The open-label design introduces the risk of bias

- The maximum number of cycles of doxorubicin in the JGDG is eight, whereas standard UK clinical practice is 6.
- The use of subsequent therapies following treatment assignment

There is substantial uncertainty in Lilly's economic model:

- The pivotal trial, JGDG, had a relatively small patient population and immature patients' survival data. Therefore, the clinical effectiveness and cost-effectiveness of OlaDox vs. Dox is highly uncertain.
- Cost-effectiveness of OlaDox vs. IfoDox is highly uncertain since it is based on an indirect comparison.
- Since overall survival data for patients in the JGDG trial was immature, external data was used for validation of extrapolated post-progression survival. The evidence source, used for validation, reported survival of a substantially younger patient population. Besides, the longest reported follow-up was 10 years after disease diagnosis, while the company extrapolated survival up to 25 years post diagnosis. Therefore, Lilly's estimates of OS for all treatments are highly uncertain.
- In the OlaDox vs. Dox comparison, the effect on survival outcome in the Dox arm to subsequent treatment with olaratumab was examined. However, the conclusion of no effect of subsequent treatment with olaratumab on survival in the Dox arm may not be valid due to a number of reasons such as small sample size, immaturity of data, and other assumptions made, which are not supported by data from JGDG. This may further contribute to uncertainty in survival and, therefore, cost-effectiveness. .
- HRQL data was not collected in JGDG trial. Health state utilities, used in Lilly's analyses, were from studies with substantially younger patient populations, which may contribute to uncertainty via selection bias.
- Eli Lilly underestimate drug administration costs in their analysis by assuming shorter drug administration time for OlaDox.
- The cost of post-progression treatment assumed in the model is not in line with the results of Lilly observational study, which, as the company stated, were used to parameterise the cost.
- Lilly assumed availability of both vial sizes of olaratumab, 190 and 500 mg, in anticipation of marketing authorisation for the 190 mg vial of Ola. However, conditional marketing authorisation has been granted only for the 500mg vial.

1.6.3 Summary of exploratory and sensitivity analyses undertaken by the ERG

1.6.3.1 PenTAG base case

Table 1. Derivation of PenTAG base case ICERs (£ per QALY)

					OlaDox vs. Dox	OlaDox vs. IfoDox
		PenTAG's assumption in the base case	**Lilly's base case**	**Reference**	[redacted]	[redacted]
1	Parametric survival function for OS	Log-normal	Gamma	Section 5.3.5.1.2, p128.	[redacted]	NA as the model uses a fractional polynomial function for the indirect comparison
2	Coefficients of fractional polynomials estimated in NMA	Mean values	Median values	Section 5.3.5.2 p137.	NA	[redacted]
3	Patients' mean weight	82.5 kg	77.3 kg	Section 5.3.7.1, p145	[redacted]	[redacted]
4	Ifo prices	£66.08 and £130.04 for 1g and 2g vials, respectively	£91.32 and £179.88 for 1g and 2g vials, respectively	Section 5.3.7.1.2**.** p146	[redacted]	[redacted]
5	Mesna prices	£9.77 and £3.95 for 1000mg and 400mg vial, respectively	£29.41, and £13.41 for 1000mg vial and 400mg vial, respectively	Section 5.3.7.1.2**.** p148	[redacted]	[redacted]
6	Availability of vial sizes for Ola	Only 500 mg vial available	Both vial sizes, 190 and 500 mg, are available	Section 5.3.7.1.5, p148.	[redacted]	[redacted]
7	HRG codes and unit costs	Corrected		Section 5.3.7.2, p149.	[redacted]	[redacted]
Overall: 1+2+3+4+5+6+7	**PenTAG base case**				[redacted]	[redacted]

1.6.3.2 Sensitivity analyses

Sensitivity analyses were performed on the most important sources of uncertainty in the company's cost-effectiveness analysis of OlaDox vs. Dox and OlaDox vs. IfoDox :

- Health state utilities
- Extrapolated overall survival (due to immature survival data in JGDG trial)
- Treatment costs post-progression
- Availability of vial sizes for Ola

The ERG believes that the results of the cost-effectiveness analysis are highly uncertain due to the small number of patients included in the pivotal trial (JGDG), and immaturity of patient survival data. Therefore, conclusions on the cost-effectiveness of Ola should be treated with caution.

We also believe that EQ-5D-5L data being collected in the ongoing phase 3 trial (ANNOUNCE),comparing OlaDox vs. Dox, will help to reduce the uncertainty of the cost-effectiveness of Ola compared to the treatments currently available in the NHS in England and Wales.

2. Background

2.1 Critique of company's description of underlying health problem

Eli Lilly describe soft tissue sarcoma (STS) as a '*rare and heterogeneous group of malignant tumours that develop from cells in the soft, supporting tissues connecting and surrounding other organs of the body including muscle, fat and blood vessels.*' (Source: Eli Lilly submission, Section 3, p. 27).

Most soft tissue sarcomas are derived from a mesodermal cell origin, with the exception of the malignant peripheral nerve sheath tumours, which are neuroectodermal in origin. Soft tissue sarcomas are grouped according to their presumed cell of origin (e.g., liposarcoma from lipo-cyte cell line, rhabdomyosarcoma from primitive skeletal muscle cell line). [6] However, some soft tissue sarcomas, such as malignant fibrous histiocytoma, have no known cell of origin.[7]

According to www.cancerresearchuk, the largest proportion of soft tissue sarcoma cases occur in the limbs, with slightly smaller proportions in the connective tissue of the trunk, and a much smaller proportion in the gynaecological organs (2008-2010). The percentage distribution is as follows:

- Limbs 25%
- Other sites 16%
- Connective tissue of trunk 15%
- Gynaecological organs 10%
- Skin 8%
- Gastrointestinal tract 7%
- Head, face and neck 7%
- Organs within trunk 6%
- Retroperitoneum 5%
- Male genitals 2%

More than 50 histological subtypes of STS exist according to the World Health Organisation, which are classified according to the originating cell rather than the site in which the sarcoma has developed.[7]

These are generally subdivided into gastrointestinal stromal tumour (GIST) and non-GISTs. GISTs are outside the scope of this technology appraisal.

Some common examples of non-GIST in adults include:

- Fat tissue sarcoma – liposarcoma
- Synovial sarcoma
- Smooth muscle sarcoma – leiomyosarcoma
- Peripheral nerve sarcoma – malignant peripheral nerve sheath tumour (MPNST)
- Fibrous tissue sarcoma – fibrosarcoma
- Pleomorphic
- Fibromatosis
- Blood and lymph vessel – angiosarcoma
- Blood vessel – haemangiosarcoma
- Lymph vessel – lymphangiosarcoma

Eli Lilly have included all histological subtypes of STS other than GIST and Kaposi sarcoma, which are considered to have distinct aetiologies and/or treatments.

The ERG believes the description given is appropriate.

2.1.1 Epidemiology

Eli Lilly give the following estimates of the incidence of STS (Source: Eli Lilly submission, Section 3, p. 27):

> *The true incidence of STS is challenging to evaluate. Among the main limitations are the substantial heterogeneity of the disease, with many rare histological subtypes and changing histopathological classification. According to Cancer Research UK (20) in 2010 there were 3,272 new cases of STS (all subtypes combined) in the UK: 1,660 (51%) in males and 1,612 (49%) in females of all ages. That is equivalent to about 9 people being diagnosed with STS per day. In England there were 2,740 new cases of STS in 2010 (20). The crude incidence rate shows that there are 54 new STS cases for every million males in the UK, and 51 for every million females (20). Males in the UK have a similar risk of developing STS compared with females.*

The incidence statistics provided by Eli Lily are well-sourced, however, it may be helpful to note that 'all subtypes combined' will include approximately 20% (661 cases) of sarcoma 'not otherwise specified' where the pathologist did not enter a specific morphological subtype. The figure of 3,272 is likely to include GIST, of which www.GISTsupportuk estimate 900 new cases a year and 152 cases of Kaposi sarcoma, which are excluded in the population under

investigation by Eli Lilly. The most common subtypes of soft tissue sarcoma in the UK in 2008-2010 were leiomyosarcoma (18%), fibroblastic sarcoma (14%) and liposarcoma (13%).[8]

The aetiology of STS remains largely unknown. In rare cases, associations have been made with certain risk factors such as environmental and genetic influences, however, in general, no cause is identified.[7]

2.1.2 Diagnosis

The symptoms of STS are vague, since they may or may not be experienced by people. Furthermore, the cause of the symptoms may be a result of a different medical condition: (Source: Eli Lilly submission, Section 3, p28):

> *STS commonly present as painless, incidentally observed tumours that often do not influence function or general health despite their often large volume. Combined with their rarity, this often leads to their misdiagnosis as benign conditions (22).*

STS rarely causes symptoms in the early stages. The first sign of a sarcoma in an arm, leg, or torso may be a painless lump or swelling. Since STS can develop in flexible, elastic tissues or deep spaces in the body, the tumour can often push normal tissue out of its way as it grows. Therefore, a sarcoma may grow quite large before it causes symptoms. Eventually, it may cause pain as the growing tumour begins to press against nerves and muscles.[9]

2.1.3 Prognosis and burden of disease

Eli Lilly have provided the following information regarding prognosis (Source: Eli Lilly submission, Section 3, p. 28)

> *Prognosis depends on several factors, including a patient's age and the size, depth, histologic grade and stage of the tumour (23). Delays in the diagnosis of STS are common: a UK study found that the median time for a patient to be referred to a specialist centre from first presentation to a medical professional was 25.0 weeks (mean 83.1 weeks)(24).*
>
> *Unfortunately, approximately 50% of patients develop distant metastases and eventually die of disseminated disease (21).*
>
> *The median overall survival of patients with metastatic STS treated with the existing standard of care, doxorubicin is 12 -16 months [10]*
>
> *The 5-year relative survival for STS (all stages) diagnosed in England have improved from 48% in 1985-1989 to 56% in 2000-2004 [11]*

Specific factors associated with a poorer prognosis include:[12]

- Age older than 60 years.
- Tumours larger than 5 cm in greatest dimension.
- High-grade histology with high mitotic activity.
- Positive margins after resection.[13]

With regard to burden of disease and the role of palliative care, information is lacking for locally advanced and metastatic STS. However, pain, including neuropathic pain and dyspnoea appear to be the most common problems.[14]

2.2 Critique of company's overview of current service provision

2.2.1 Health inequalities

Eli Lilly discusses the potential health inequalities experienced by people with STS as opposed to people with more common cancers (Source: Eli Lilly submission, Section 3, p. 28):

> *[…] because of the rarity of sarcomas, patients may receive inappropriate treatment by non-skilled practitioners before the diagnosis has been made and the need for specialised care has been recognised (29). According to Sarcoma UK, a third of patients diagnosed with sarcoma each year are not referred appropriately to sarcoma specialised services (28). Of these, 10% will be treated inappropriately for another condition and 10% will be informed that their symptoms are not serious enough to warrant a return to their GP (28).*

The ERG agrees that this may lead to a delayed diagnosis with a potentially more advanced sarcoma.

2.2.2 Current UK STS treatment pathway

The company highlights the complication of an overall treatment pathway due to the large number of histological subtypes. This is compounded by the lack of RCTs as a result of the rarity of the disease and therefore treatment is based on smaller, phase 2 trials.

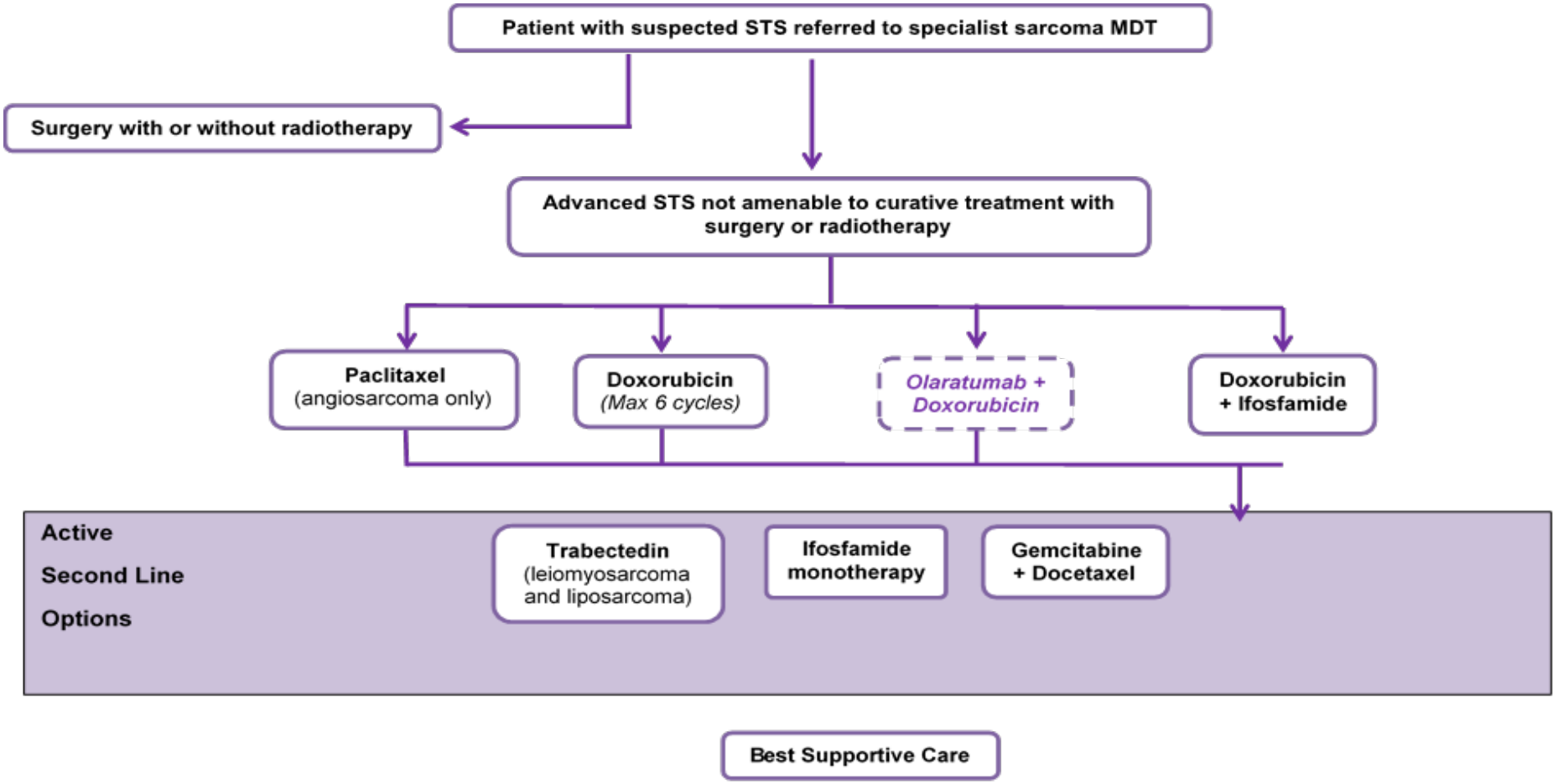
The company provides an overview, with surgery as the standard treatment and radiotherapy may be recommended pre- or post-operatively. It is then noted that STS have a tendency to recur as inoperable or metastatic disease. At this point, chemotherapy is the treatment of choice. (Source: Eli Lilly submission, Section 3, p. 30):

> *Therefore chemotherapeutic agents are generally used with palliative intent in advanced STS. The published STS chemotherapy response rates vary from 10% to 50%, depending*

on the drugs used, methods of assessment, patient selection and histological subtype (21). Good performance status, young age and the absence of liver metastases have a significant, favourable influence on both survival time and response rate (36).

Given the variation in clinical practice across the UK, the company provides a pathway for locally advanced and metastatic STS which takes into account NICE guidance, NHS England service specification, guidelines published by the British Sarcoma Group (BSG) and the views of UK sarcoma specialists.

Figure 1: Current UK STS Treatment Pathway (excluding GIST and Kaposi Sarcoma)



Source: Eli Lilly submission, Section 3, p. 30

2.2.2.1 First line chemotherapy

Eli Lilly gives a description of current UK practice for first line chemotherapy (Source: Eli Lilly submission, Section 3, p. 32):

> *In the UK, doxorubicin has been used as standard first-line treatment for over three decades (1). According to a real world treatment patterns study conducted in the UK (38), 47% of patients in the first-line advanced STS setting are currently treated with doxorubicin. It is used at a dose of 75mg/m2 every three weeks. The duration of treatment depends on response but a maximum of 6 cycles is recommended because of the risk of cumulative cardiotoxicity (21). Doxorubicin has an objective response rate of between 10% and 30% (39) (40) (41) and approximately 45% of patients derive some clinical benefit (42). However, despite the improvements in STS patient outcomes, the median survival of patients with metastatic sarcoma is still only 12 to 16 months (2).*

The company notes that ifosfamide monotherapy is not routinely administered for this patient group in the NHS setting. With regard to IfoDox, they note that although the response rates may be higher, there is no improvement in OS and toxicity is increased. (Source: Eli Lilly submission, Section 3, p. 32)

The company discusses other first line treatments for specific STS subtypes such as gemcitabine with docetaxel, where the efficacy and toxicity profile does not appear to be an improvement on doxorubicin, and paclitaxel, which is indicated for angiosarcoma.

2.2.2.2 Second line chemotherapy

Second line chemotherapy is described by the company as follows (Source: Eli Lilly submission, Section 3, p. 33):

> *In patients of good performance status, second-line chemotherapy options following failure of first-line anthracycline-based chemotherapy include ifosfamide, trabectedin, dacarbazine and gemcitabine, alone or in combination with docetaxel. Reported response rates are in the range of 5% to 25% and the choice of agent depends on histology, toxicity profile and patient preference (21). According to NICE TA185 (48), trabectedin is recommended as a treatment option for patients with advanced STS if treatment with anthracyclines and ifosfamide have failed or they cannot tolerate anthracyclines and ifosfamide, or anthracyclines and ifosfamide are unsuitable.*

The company notes that although other second-line options are available, these are not routinely administered via clinical practice in the NHS. However, a UK real world treatment patterns study shows (Source: Eli Lilly submission, Section 3, p. 33):

> *[…] that 35% of patients in the second-line advanced STS setting received GemDoc, 24% receive pazopanib, 9% received doxorubicin, 8% received ifosfamide, 5% received trabectedin and 19% received other regimens (38)*.

2.2.3 Anticipated place of olaratumab in clinical practice

The company anticipates that OlaDox will be an option for first line chemotherapy, where the other choices are Dox monotherapy or IfoDox (Eli Lilly submission, Section 3, p 34):

> *The UK standard of care for advanced STS has shown consistent efficacy across a broad range of histological subtypes and is associated with a median OS of 12-16 months, a median PFS of 2 to 5 months and response rates of between 10% and 30% (2) (39, 40, 46, 49)*
>
> *The phase 1b/2 study JGDG demonstrated that OlaDox increases the efficacy of Dox monotherapy, with a generally manageable tolerability profile. In this study, OlaDox extended median OS by a further 11.8 months (26.5 months OlaDox vs. 14.7 months Dox; HR = 0.463; p = 0.0003) relative to Dox monotherapy.*

It should be noted that the maximum cumulative lifetime dose of doxorubicin is 450mg/m^2 (6 cycles at a dose of 75mg/m^2), therefore, patients who have already received Dox in the first-line setting would not be eligible for treatment with OlaDox in subsequent lines of treatment.

3. Critique of company's definition of decision problem

The company presented their decision problem within the Executive Summary chapter, under the subheading 'statement of the decision problem' (Eli Lilly submission, Section 1.1, p. 12-13). A summary table of the NICE Scope, the company's decision problem and the ERG's critique is presented below (Table 2).[15] Further comments to the decision problem follow the table.

Table 2: Summary table of decision problem critique

Decision problem	NICE Scope	Company's decision problem	ERG notes
Population	Adults with advanced soft tissue sarcoma that is not amenable to curative treatment with surgery or radiotherapy, and who have not been previously treated with doxorubicin.	Adults with advanced soft tissue sarcoma that is not amenable to curative treatment with surgery or radiotherapy, and who have not been previously treated with doxorubicin, and where patients have had no prior line of systemic treatment (excludes adjuvant/neoadjuvant chemotherapy). (i.e. olaratumab in combination with doxorubicin as a first-line treatment in advanced or metastatic disease)	The company have specified patients with no previous line of treatment, since this is the anticipated place for OlaDox in the UK STS treatment pathway.
Intervention	Olaratumab in combination with doxorubicin	As per Scope	No comments.
Comparator	Doxoxrubicin monotherapy Doxorubicin with ifosfamide	As per Scope	No comments
Outcome	The outcome measures to be considered include: Overall survival PFS Response rates Adverse effects of treatment Health-related quality of life Cost per quality-adjusted life year.	As per Scope	The company have specified objective response rate (ORR)

Key: OlaDox, olaratumab + doxorubicin; ORR, objective response rate; PFS, progression free survival; STS, soft tissue sarcoma
Source: NICE Scope [8] and Eli Lilly submission, Table 1, p. 14–15

3.1 Population

The defined population in the company's submission (where adults have not been previously treated with doxorubicin, and had no prior line of systemic treatment), is similar to the population specified in the NICE Scope (adults who have not been previously treated with doxorubicin).[8] However, the company have included a further exclusion of no previous systematic treatment. Eli Lilly explain this addition by stating that (Source: Eli Lilly submission, Section 1.1, Table 1, p. 14):

> *In study JGDG, patients with no previous line of treatment comprised 65% of the patient population (metastatic and excluding adjuvant/neoadjuvant chemotherapy). In view of the anticipated place of olaratumab in combination with doxorubicin (OlaDox) in the UK STS treatment pathway, the base case cost-effectiveness analysis presented in this submission is for OlaDox as a first-line treatment.*

Overall, the ERG agrees that the population considered by the company's submission is appropriate.

3.2 Intervention

The company's decision problem specified the intervention as 'olaratumab in combination with doxorubicin', which matches the NICE Scope.[15]

The NICE Scope describes olaratumab as follows; "Olaratumab (brand name unknown, Eli Lilly) is a monoclonal antibody, which acts as a platelet-derived growth factor receptor alpha (PDGFRα) antagonist. It prevents the formation of new blood vessels and limits nutrient supply to the tumour causing death of tumour cells. It is administered intravenously."

The Committee for Medicinal Products for Human Use (CHMP) reviewed olaratumab (Lartruvo, Eli Lilly) under the EMA's accelerated assessment program and recommended conditional approval for the medicine:

> *As part of the conditional marketing authorisation, the applicant for Lartruvo must provide results from an ongoing Phase III study in order to confirm the previous results. The study compares how long patients receiving doxorubicin plus Lartruvo survive compared with patients who only receive doxorubicin. The study is ongoing and the data will be provided by the applicant. The CHMP will review the benefits and risks of Lartruvo annually to determine whether the conditional marketing authorisation can be maintained until full data are available.*

Because soft tissue sarcoma is rare, Lartruvo received an orphan designation from the Committee for Orphan Medicinal Products (COMP) in 2015. Orphan designation is the key instrument available in the European Union (EU) to encourage the development of medicines for patients with rare diseases. Orphan-designated medicines qualify for ten years' market exclusivity. In addition orphan designation gives medicine developers access to incentives, such as fee reductions for marketing authorisation applications and for scientific advice.

The EMA have recommended the following guidance for administering olaratumab:

The recommended dose of Lartruvo is 15 mg per kilogram body weight, given twice over a period of three weeks, on days 1 and 8. These three-week cycles should be repeated until the disease gets worse or side effects become unacceptable. Lartruvo is given in combination with doxorubicin for up to 8 cycles of treatment, followed by Lartruvo alone in patients whose disease has not got worse. Doxorubicin is given on day 1 of each cycle, after the Lartruvo infusion.

Our clinical advisor (PS) commented that they would typically only administer doxorubicin for a total of 6 cycles, rather than 8.

3.3 Comparators

The comparators described in the company submission match those stated in the NICE scope, which are Dox monotherapy and IfoDox.[8] However, our clinical advisor (PS) commented that IfoDox is usually a neoadjuvant treatment, rather than palliative, due to the associated toxicity. Furthermore, it displays no benefit for overall survival, but does improve response rate.

3.4 Outcomes

The outcomes in the company submission match those in the Scope. With regard to the outcome 'response rates' as stated in the scope, the company have investigated objective response rates (ORR), which is the '*proportion of patients achieving a best overall response of partial response (PR) or complete response (CR), according to RECIST'* (Source: Eli Lilly submission, Section 4.3, Table 8, p. 37). The ORR is an appropriate measure. However, this combined outcome often does not show the rate of CR, which may be very low but tends to provide the greater benefit.

3.5 Other relevant factors

In response to special considerations relating to equity and equality, the company consider OlaDox to fulfil the two criteria specified in section 6.2.10 of the NICE guide to the methods of technology appraisal under 'life-extending treatment at the end of life'. The company's justification is shown in Table 3.

Table 3: Response to end of life criteria

Criterion	Data available
The treatment is indicated for patients with a short life expectancy, normally less than 24 months	The median overall survival of patients with metastatic STS treated with the existing standard of care, doxorubicin is 12 -16 months [10]
There is sufficient evidence to indicate that the treatment has the prospect of offering an extension to life, normally of a mean value of at least an additional 3 months, compared with current NHS treatment.	**JGDG study:** Median OS benefit of 11.8 months vs. Dox (the UK standard of care)

Source: Eli Lilly submission: Section 4.13, Table 37, p. 115

4. Clinical effectiveness

4.1 Critique of the methods of review(s)

4.1.1 Searches

4.1.1.1 Clinical effectiveness

Eli Lilly presented a literature search protocol to support its review of clinical effectiveness. This protocol included systematic searches of key biomedical databases using a literature search strategy, searching of conference websites and a search of clinical trials.gov. The literature search was last updated in September 2016.

The bibliographic database searching used a search strategy that took the following form:

1. (controlled index terms for sarcoma and various sub-types of sarcoma) OR
2. (free-text terms for sarcoma and various sub-types of sarcoma) AND
3. (controlled index terms for olaratumab or an extensive list of comparators including non-hormonal or hormonal treatments OR
4. (free-text terms to olaratumab or an extensive list of comparators including non-hormonal or hormonal treatments) AND
5. (a range of search terms for study design (RCTs, clinical trials, meta-analysis and systematic reviews), limits to remove studies conducted on animals and studies on children)

NOT

6. (a range of search terms to exclude comment, letters, editorials, case reports, reviews, guidelines, cross-sectional, cohort, case control, observational, prospective and retrospective studies, guidelines as publication type) NOT
7. (a range of controlled index and free text terms – it is not clear what the rationale is for these – registries, chart review, administrative data, claims data, electronic medical record, electronic health record, real world, medicare claims, SEER, surveillance, epidemiology and end results) NOT
8. (controlled index and free-text terms for HIV or AIDS) AND
9. (limited to 2004 onwards).

The search strategy was applied in the following bibliographic databases: Medline and Medline-in-Process (PubMed), Embase (Elsevier at embase.com) and The Cochrane Library.

The following conference websites were searched: American Society of Clinical Oncology (ASCO) in Jan-May 2015, European Society for Medical Oncology (ESMO) in Feb 2015, and the Connective Tissue Oncology Society (CTOS) in July 2016. Finally, clinicaltrials.gov was searched for relevant, unpublished studies in March 2015.

It is not clear why the conference website searches and clinical trials searches were not updated in 2016 with the other literature searches (apart from the CTOS search).

The literature searching for clinical effectiveness studies is reasonably well conducted and reported. However there are a few concerns.

- The text on p75 paragraph 3 states that no date restrictions were used, but in fact searches were limited from 2004 onwards. The decision to limit by date may be valid but it is explained in the text.
- The filter used to limit to RCTs is not the Cochrane search filter or any other validated filter that we recognize. It is unclear why a validated search filter was not used to limit to RCTs.
- The decision to exclude a large number of study types (see line 6 of the literature search description above) may have resulted in exclusion of relevant studies. It is not clear what the result of those exclusions would be or whether any relevant studies could have been inadvertently excluded as a result.
- There is a long list of excluded terms, combined to the main search with NOT (see line 7 of the literature search description above). Again it is not clear what the result of those exclusions would be or whether any relevant studies could have been inadvertently excluded as a result. It is not clear how or why these terms were chosen as a method to exclude studies.
- The effect of excluding the terms listed in lines 6-7 is that if a paper or study contained one of the excluded terms listed, as part of the wording in the abstract or elsewhere in the database record, that paper would be excluded from the search results – even if it was relevant.
- There is little information about the screening methods used for the review. It is not clear whether the papers were double screened at title and abstract and whether full text studies were double screened. Data extraction methods for included papers are not detailed.
- The literature search results in Section 4.10 page 78, paragraph 1 are not in agreement with the results in the PRISMA diagram on page 81, Figure 16. Eli Lilly

have clarified the figures on page 78 and given more detail on these but this still doesn't correspond to the figures in Figure 16. This may just be a question of not having updated the PRISMA diagram in line with the text but this is not clear.

4.1.1.2 Adverse events

Eli Lilly did not undertake separate literature searches to identify studies reporting adverse events. In clarification Eli Lilly stated that adverse events would be best reported in the clinical trials found by the systematic review.

Eli Lilly's searches were limited by study design. It is therefore possible that exclusion of cohort, case-control, cross-sectional and case series as publication types in the literature searches means that papers reporting adverse events may have been missed.

4.1.2 Inclusion criteria

Eli Lilly's inclusion criteria are given below (Table 4) with an additional column added to the right of the table, taken from the Scope for reference and comparison. Comments about the differences in inclusion criteria are outlined below the table.[15]

Table 4: Scope of the literature review: PICOS criteria for study inclusion

Criteria	From Eli Lilly Definition	From Scope
Population	Patients aged ≥ 18 years with advanced STS Disease has not responded to surgery or radiotherapy Stage III and IV	Adults with advanced soft tissue sarcoma that is not amenable to curative treatment with surgery or radiotherapy, and who have not been previously treated with doxorubicin.
Interventions/ comparators	Trials that include any of the following interventions in at least 1 study arm:[a] Axitinib Bevacizumab Brivanib Brostallicin Carboplatin Celecoxib Cediranib Cisplatin Conatumumab Crizotinib Cyclophosphamide Dacarbazine Dactinomycin Daunorubicin/daunomycin Deforolimus Docetaxel Doxorubicin Epirubicin Eribulin Etoposide Everolimus	Intervention: Olaratumab in combination with doxorubicin Comparators: • doxorubicin monotherapy • doxorubicin with ifosfamide

Criteria	From Eli Lilly Definition	From Scope
	Gemcitabine Idarubicin Ifosfamide Imatinib Interferon Irinotecan Liposomal doxorubicin Methotrexate Mitoxantrone Olaratumab Ombrabulin Paclitaxel Pazopanib Pegylated doxorubicin Progestogen Rapamycin Sirolimus Sorafenib Sulindac Sunitinib Temozolomide Temsirolimus Topotecan Trabectedin Vinblastine Vincristine Vinorelbine Aromatase inhibitors (e.g., exemestane, anastrozole, letrozole, vorozole, formestane, fadrozole) Gn-RH analogues (tamoxifen, toremifene)	
Outcomes	Specific outcomes on efficacy, safety, and health-related quality of life will be determined during data extraction.	The outcome measures to be considered include: • overall survival • progression-free survival • response rates • adverse effects of treatment • health-related quality of life.
Study Design	Phase 2, 3, and 4 clinical trials (including single-arm phase 2 trials) Randomised clinical trials Long-term follow-up (open label) studies of clinical trials Subanalyses and reanalyses of randomised clinical trials Comparative (head-to-head) clinical trials Systematic reviews and meta-analyses[b]	
Other	No table of excluded studies	

Key: AE, adverse events: STS, soft tissue sarcoma

Notes: a, Any study investigating these interventions with any control arm were included (i.e., at least 2 treatment regimens covered), such as an alternative chemotherapy or targeted agent regimen; b, Systematic reviews and meta-analyses were not included in their own right, but were used for identification of primary studies not previously identified.
Source: Eli Lilly submission, Table 24, pp. 77–78 and NICE Scope[15]

4.1.2.1 Population

The population defined in Eli Lilly's submission differs to the NICE Scope in that it does not exclude participants who have previously received doxorubicin.[15] However, for the only RCT identified in the systematic review, previous treatment with doxorubicin was an exclusion criterion and therefore this study complies with the scope.[16]

4.1.2.2 Interventions/comparators

There are many more comparators listed by the company than are included in the Scope (Dox monotherapy and IfoDox). This is with a view to performing a network meta-analysis due to the lack of evidence for olaratumab and therefore the ERG considers the expanded list of comparators appropriate (Source: Eli Lilly submission, Section 4.5):

Other than the phase 2 study JGDG, no other published head-to-head randomised clinical trials were found that provide evidence of the efficacy and safety of OlaDox versus the comparators (Dox and IfoDox) listed in the NICE final scope. Therefore, no direct meta-analysis was performed and instead the evidence networks were analysed via a network meta-analysis (NMA).

4.1.2.3 Outcomes

The outcomes are poorly defined in the inclusion criteria by the company and are stated in Table 4 to be identified during the data extraction procedure, rather than pre-specified. Therefore, these cannot be compared with the Scope. With no defined outcome criteria by the company, there is a risk that a decision to include may be made on the results of a study and the evidence presented may be biased. However, studies excluded on the basis of outcomes were examined by the ERG and none would have fulfilled the inclusion criteria for this review.

4.1.2.4 Study design

The Scope did not restrict study design. However, the NICE reference case guide to the methods of technology appraisal 2013 (Chapter 5.2.3)[17] recommends studies should be restricted to RCTs and when they are not available, non RCTs. Studies included in the company submission were one RCT in the systematic review and 6 RCTs in the network meta-analysis. We are satisfied the study designs meet the reference case.

4.1.2.5 Study selection

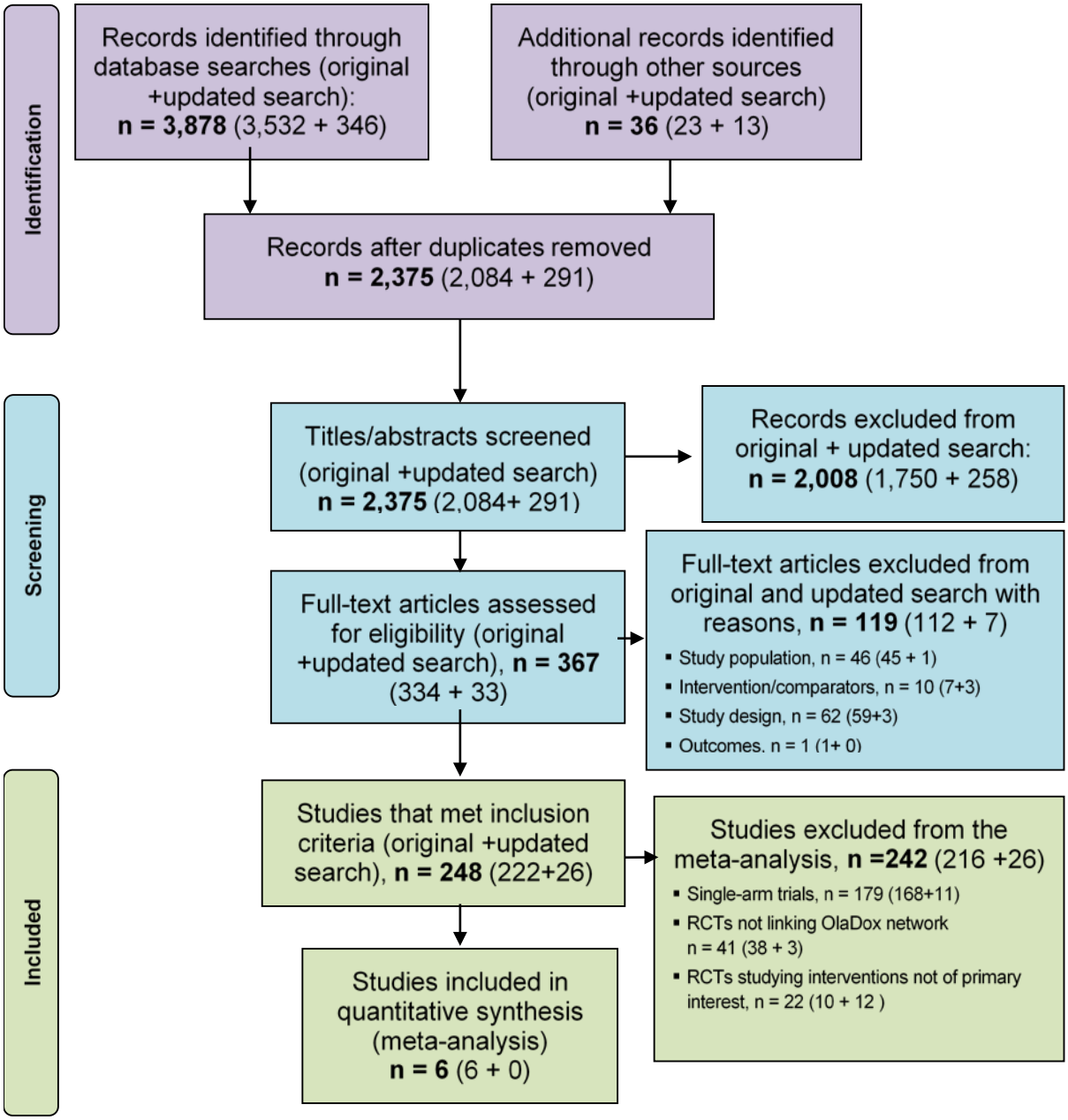
The company submission does not fully explain the process used in study selection, for example, there is no mention of the number of researchers independently reviewing the abstracts and the full-texts of studies, or how discrepancies between investigators following comparison were resolved. It is standard in systematic reviews to have two researchers independently assess studies for inclusion and for any discrepancies to be resolved by a third reviewer.[17]

From 2,375 citations identified, 2008 were excluded and 367 were taken to full-text screening. One hundred and nineteen full-text articles were excluded prior to screening for inclusion in the NMA.

A further, 242 studies were excluded, with 179 being single arm trials, 41 RCTs not linking to OlaDox and 22 RCTs not investigating interventions of primary interest. This leaves six studies included in the meta-analysis, only one of which is an RCT comparing OlaDox versus Dox.

The PRISMA diagram reported in Eli Lilly's submission is copied below (Figure 2).

Figure 2: PRISMA study flow diagram



4.1.3 Critique of data extraction

The company submission does not provide details on the methods of data extraction, therefore the ERG are unable to comment.

4.1.4 Quality assessment

Details of the company's critical appraisal of the JGDG study,[16] alongside our critique, can be seen below in Table 5. The critical appraisal has been adapted from the CRD's assessment criteria for risk of bias in RCTs.[18]

Table 5: Critical appraisal of JGDG study

Critical appraisal criterion	Eli Lilly's Assessment	ERG Comment
Was randomisation carried out appropriately?	Yes Patients were randomised using an interactive voice response system (IVRS) or interactive web response system (IWRS). After entering the patient's information, the IVRS/IWRS assigned the patient to a treatment arm based on a dynamic randomisation algorithm, which served to minimise imbalance between treatment arms with respect to stratification factors.	The minimisation randomisation technique is an acceptable system for randomisation.
Was the concealment of treatment allocation adequate?	Yes Computer generated, centralised system (IVRS or IWRS).	The ERG agree that the centralised system of treatment allocation ensures allocation concealment.
Were the groups similar at the outset of the study in terms of prognostic factors?	Yes. Groups were well balanced between study treatment arms. Some small discrepancies were observed, such as in the slightly higher percentage of females on the experimental arm; but such minor imbalances are not unusual or unexpected in a randomized trial of this size.	As highlighted by Eli Lilly, there is a difference of 10% in females between the intervention and comparator arms. However, we agree that overall the demographics between OlaDox and Dox are well balanced.
Were the care providers, participants and outcome assessors blind to treatment allocation?	Open-label study. However, the study team were blinded to aggregate efficacy data and reviewed only blinded summary reports prior to the interim analysis of PFS. Furthermore, a blinded independent review of radiographic scans was conducted following the final PFS database lock to evaluate any potential systematic bias favouring any one of the treatment arms with respect to PFS assessment.	Since the study was open labelled, the care providers and participants could not be blinded to treatment allocation. Awareness of treatment allocation will have introduced the potential for bias within the study, particularly with reporting of adverse events and progressive disease determined by symptomatic deterioration. However, the independent radiological reviewers were blinded.
Were there any unexpected imbalances in drop-outs between groups?	No. No further details given.	The most common reason for discontinuation of study treatment was progressive disease. Within the first 8 cycles, for radiologically documented PD, this was 21 in the OlaDox arm and 27 in the Dox. However, if the symptomatic deterioration PD is also included, then the number of participants reported with PD in the Dox arm rises to 34.

Critical appraisal criterion	Eli Lilly's Assessment	ERG Comment
Is there any evidence to suggest that the authors measured more outcomes than they reported?	No.	The outcome measures listed in the protocol for the trial correspond with the outcome measures reported.
Did the analysis include an intention-to-treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Analyses were performed using all randomized patients (ITT population). Handling missing data was reported in the statistical analysis plan.	Yes, we agree the main analysis adopts 'intention to treat' principles.

Key: Dox, doxorubicin; IVRS, interactive voice response system; IWRS, interactive web response system; OlaDox, olaratumab+doxorubicin
Source: Eli Lilly submission, Section 4.6, Tables 15-16, pp. 56–57

4.1.5 Evidence synthesis

From the searches, only one RCT was identified. Therefore synthesis of the evidence was not required.

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

4.2.1 Methods

The single RCT (study name JGDG; main publication by Tap et al. 2016) identified was presented in detail within the submission.[16] No further relevant studies were identified by the ERG.

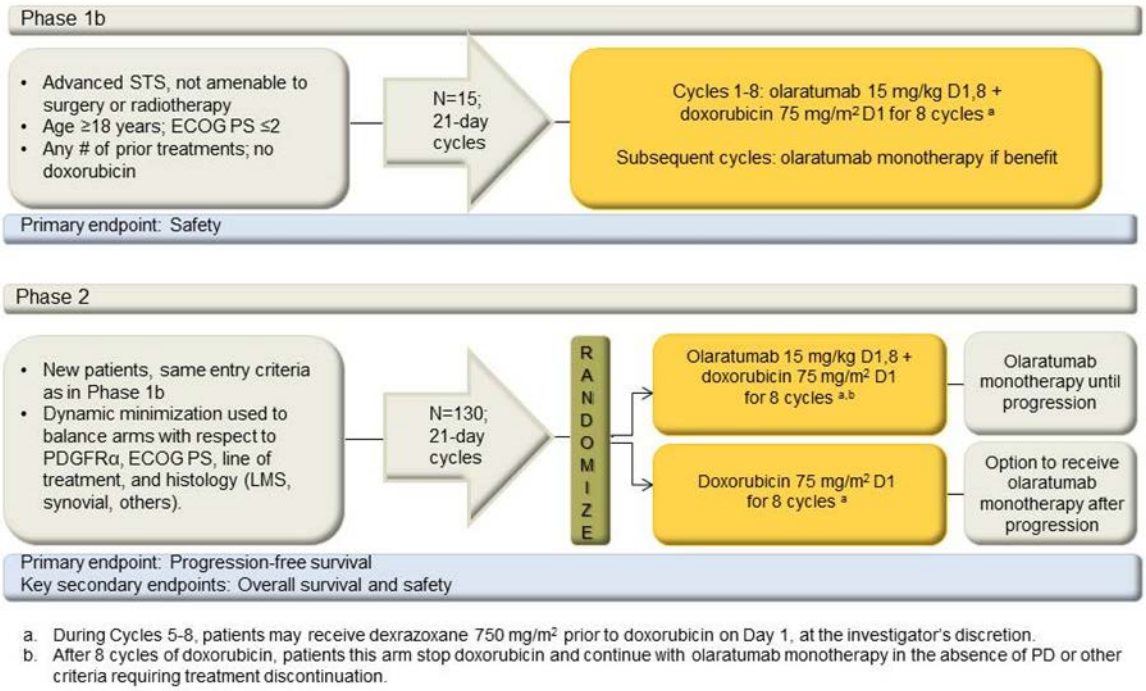
4.2.1.1 Study objective

The company's submission does not report a study objective. However, the primary outcome measure of progression-free survival and secondary outcome measures of overall survival, objective response rate and safety correspond to the outcome measures detailed in the NICE Scope.[15]

4.2.1.2 Study design and treatment

The study JGDG was a multicentre (16 sites, USA), open-label, phase two trial. Phase 1b was a single arm trial with safety as the primary endpoint. These participants were not included in phase 2, which was a randomised, parallel-group study. The company's submission mainly focuses on the safety and efficacy data generated in phase 2.

Figure 3: JGDG trial design



a. During Cycles 5-8, patients may receive dexrazoxane 750 mg/m² prior to doxorubicin on Day 1, at the investigator's discretion.
b. After 8 cycles of doxorubicin, patients this arm stop doxorubicin and continue with olaratumab monotherapy in the absence of PD or other criteria requiring treatment discontinuation.

Source: Eli Lilly submission, Section 4.3, Figure 2, p36

Randomisation

Phase 1b was non-randomised.

Prior to randomisation for phase 2, four patient baseline characteristics were pre-defined as stratification factors (Eli Lilly submission, Section 4.3, Table 8, p 37):

- PDGFRα expression (positive vs. negative assessed by IHC);
- Number of lines of previous treatment (0 vs. ≥1);
- Histologic tumour type (LMS vs. synovial sarcoma vs. other subtype); and
- ECOG PS (0-1 vs. 2).

Platelet-derived growth factor (PDGF) and PDGF receptor (PDGFR) signalling plays a significant part in mesenchymal biology, including mesenchymal stem cell differentiation, growth and angiogenesis.[16] However, the immunohistochemical method used to assess

PDGFRα expression for patient stratification was revealed to have poor specificity. An alternative assay was developed and post hoc analysis indicates the two arms were balanced.

Randomisation was achieved using a method of dynamic minimisation with a probability factor of 0.8, designed to balance the pre-defined characteristics in both study treatment arms.

Allocation was performed as follows (Source: Eli Lilly submission, Section 4.3, Table 8, pp 37-38):

> *Each site accessed the interactive voice response system (IVRS) or interactive web response system (IWRS) reachable 24 hours a day, to randomise the patient. After entering the patient's information, the IVRS assigned the patient to a treatment arm based on the dynamic randomisation algorithm. A unique identification number was assigned to each patient.*

Following randomisation and drug administration, tumour response was assessed every 6 weeks according to the Response Evaluation Criteria in Solid Tumour (version 1.1) and survival was assessed every 2 months until study completion. All patients were followed for a minimum of 30 days after the last dose of olaratumab and thereafter every 4 to 6 weeks until all olaratumab-related toxicities resolved, stabilised, returned to baseline, or were deemed irreversible (Eli Lilly submission, Section 4.3, Table 8, p 38).

The timing of planned safety analysis is more unclear (Source: Eli Lilly submission, Section 4.3, p 41):

> *Blood samples were collected for pharmacokinetic and immunogenic analyses. Safety was assessed for all patients who received at least one dose of study treatment. Adverse events and clinical laboratory toxicity were graded according to the National Cancer Institute Common Terminology Criteria for Adverse Events (version 4.0). Cardiac function was monitored by echocardiography or multi-gated acquisition (MUGA) scanning before treatment start and before treatment at cycles 5 and 7.*

Drug administration and data collection protocols are outlined in Table 6. By design, there is awareness of the treatment allocated for both the patient and primary care givers from an open-labelled study. Awareness of treatment allocation will have introduced the potential for bias within the study, particularly with reporting of adverse events. However, based on the treatments administered within the study, an open-label study design was the most appropriate study design to be utilised.

Table 6: Treatment protocol

Treatment	Administration	Data collection
Phase 1b		
OlaDox	Olaratumab (15 mg/kg) intravenously on day 1 and day 8 plus doxorubicin (75 mg/m^2) on day 1 of each 21-day cycle for up to eight cycles	NR
Phase 2		
OlaDox	Olaratumab (15 mg/kg) intravenously on day 1 and day 8 plus doxorubicin (75 mg/m^2) on day 1 of each 21-day cycle for up to eight cycles. Beginning with cycle 9, olaratumab (15 mg/kg) on day 1 and day 8 of each subsequent 21 day cycle until documented progressive disease or discontinuation for any other reason.	Tumour response was assessed every 6 weeks according to the Response Evaluation Criteria in Solid Tumour (version 1.1) Survival was assessed every 2 months until study completion
Dox	Doxorubicin (75 mg/m^2) on day 1 of each 21-day cycle for up to eight cycles. If discontinuing Dox due to disease progression occurring during or after completion of the initial 8 cycles, can receive optional Olaratumab (15 mg/kg) on Days 1 and 8 of each 21 day cycle, until progressive disease, unacceptable toxicity or discontinuation for any other reason.	Tumour response was assessed every 6 weeks according to the Response Evaluation Criteria in Solid Tumour (version 1.1) Survival was assessed every 2 months until study completion

As noted in the company submission, the protocol for doxorubicin monotherapy differs from typical UK practice (Source: Eli Lilly submission, Section 1.5, p 21):

> *…there is uncertainty relating to treatment duration due to generalisability of the trial population to UK clinical practice: up to 8 cycles of olaratumab/doxorubicin were administered in the US-based clinical trial, compared with a maximum of 6 cycles of doxorubicin monotherapy typically used in UK clinical practice*

On the days in which OlaDox was administered, Ola was always administered prior to Dox. According to our clinician (PS) and following the response to clarification questions, we are satisfied that it is possible to differentiate toxicity effects between Ola and Dox. The company's submission details the following protocol regarding discontinuation of therapies (Source: Eli Lilly submission, Section 4.3, p 41):

> *In general, discontinuation of one study agent (Ola or Dox) did not necessitate discontinuation of the other for patients in the OlaDox arm of Phase 2. In the event of alteration or discontinuation of Ola therapy due to an Ola-related toxicity, Dox did not need to be altered, and the planned Dox schedule was maintained. Similarly, Ola therapy was not altered or discontinued for Dox-related toxicity.*
>
> *If treatment with olaratumab was withheld for more than 6 continuous weeks (that is, 2 treatment cycles) due to an olaratumab related toxicity that did not resolve, olaratumab was*

permanently discontinued. Once all study treatment was permanently discontinued, the patient completed the end-of-therapy evaluations and entered long-term follow-up…If a patient in the OlaDox arm of Phase 2 was permanently withdrawn from olaratumab therapy due to a toxicity clearly attributed to olaratumab, the patient could continue to receive doxorubicin for a maximum of 8 cycles as long as all other study criteria were met.

The only concomitant medication was dexrazoxane (750 mg/m^2) which may be administered from cycle five to eight for patients treated with doxorubicin (Source: Eli Lilly submission, Section 4.3, p 39). Dexrazoxane is a cardioprotective agent, which is not generally used in the UK.

4.2.1.3 Study duration

The study duration was defined in the company submission as follows (Source: Eli Lilly submission, Section 4.3, Table 8, p 37):

- First patient was enrolled (assigned to therapy) on 06 October 2010,
- Primary outcome (PFS) data cut-off date was 15 August 2014 and
- Final data cut-off date was: 16 May 2015

The ERG considered the study duration was suitable, enabling adequate assessment of the outcomes following treatment for STS.

4.2.1.4 Blinding

The treatment of STS within OlaDox necessitated an open-labelled design due differing time periods of treatment (OlaDox on day 1 and day 8 of 21 day cycle; Dox on day 1 of 21 day cycle). Efforts to minimise bias were as follows (Source: Eli Lilly submission, Section 4.6, p 56 -57):

Lilly internal study team remained blinded to treatment assignment until the time of interim efficacy analyses.

A blinded independent assessment of disease progression and treatment response was conducted following the final PFS database lock to evaluate any potential systematic bias favouring any one of the treatment arms with respect to PFS assessment.

As previously mentioned, the open-label design creates an opportunity for bias, particularly for reporting of AEs by care providers, who are not blinded.

4.2.1.5 Inclusion/exclusion

Table 7 gives the inclusion/exclusion criteria from the JGDG trial. Those listed in the company's submission agree with those listed in the RCT paper and are appropriate for the NICE Scope.[15]

Table 7: Eligibility criteria

Key inclusion criteria	Key exclusion criteria
• At least 18 years of age and have a life expectancy of more than 3 months. • A histologically or cytologically confirmed malignant STS • Measurable disease as defined by RECIST (Version 1.1) • Normal coagulation, haematologic, hepatic, and renal function. • An ECOG PS score of 0-2 at study entry • A left ventricular ejection fraction (LVEF) ≥50% assessed at baseline	• Histologically or cytologically confirmed Kaposi's sarcoma • Untreated central nervous system metastases • Prior treatment with doxorubicin, daunorubicin, idarubicin, and/or other anthracyclines and anthracenediones (that is, mitoxantrone) or therapy with any agent that targets the PDGF or PDGFR • Prior radiation therapy to the mediastinal/pericardial area • Concurrent treatment with other anti-cancer therapy, including other chemotherapy, immunotherapy, hormonal therapy, radiotherapy, chemo-embolisation, targeted therapy, or an investigational agent or non-approved use of a drug or device within 4 weeks prior to study entry • Unstable angina pectoris, angioplasty, cardiac stenting, or myocardial infarction 6 months prior to study entry

Key: ECOG PS, Eastern Cooperative Oncology Group performance status
Source: (Eli Lilly submission, Section 4.3, Table 9, p 40)

4.2.1.6 Location

The location of investigation sites was reported to be as follows (Source: Eli Lilly submission, Section 4.3, p 40):

> *The multi-centre study, JGDG, was conducted, at 16 clinical sites in the US only. All sites that participated in the study specialised in the treatment of sarcoma.*

Clinical practice at the US sites causes some uncertainty with regard to generalising to the UK population, since (Source: Eli Lilly submission, Section 1.5, p 20):

> *…up to 8 cycles of olaratumab/doxorubicin were administered in the US-based clinical trial, compared with a maximum of 6 cycles of doxorubicin monotherapy typically used in UK clinical practice.*

Therefore, the maximum cumulative dose of Dox administered in the UK is limited to 450 mgm^2, whereas the limit in the US study was 600 mgm^2.

Furthermore, with regard to dexrazoxane, which was administered for cardiac protection in the study at the investigator's discretion, the company state (Source: Eli Lilly submission, Section 2.4, p 24):

Dexrazoxane is not routinely used in UK clinical practice and is licensed only for use in patients with metastatic breast cancer. Section 4.4 of the olaratumab SPC- special warnings and precautions for use states that doxorubicin can cause cardiotoxicity and recommends the use of appropriate cardioprotective measures such as ECHO or MUGA scans and/or the use of cardioprotective agents throughout treatment but does not mandate the use of Dexrazoxane (15). Of note is that the cardio-protective measures recommended in the SPC are the same as those currently undertaken with standard of care doxorubicin. In conclusion, the use of OlaDox in UK clinical practice is not expected to change service provision and management since dexrazoxane is not commonly used in clinical practice and the cardiac-protective measures required for OlaDox are the same as those currently undertaken with doxorubicin.

It should be noted that dexrzoxane was only administered from cycle five onwards.

4.2.1.7 Study endpoint

The study endpoints and definitions are presented in Table 8

Table 8: Study endpoints

End point	Definition
Primary end point	
Progression free survival (PFS)	Time from date of randomisation to the earliest date of documented tumour progression or death from any cause. Patients who died without a reported prior progression were considered to have progressed on the day of their death.
Secondary end points	
Overall survival (OS)	Time from randomisation until of death from any cause.
Objective response rate	The proportion of patients achieving a best overall response of partial response (PR) or complete response (CR), according to RECIST, from randomisation until disease progression/recurrence.
Change in tumour size	The maximum reduction from baseline per patient in the sum of target lesions.
Safety	Adverse events were summarised by MedDRA™ System Organ Class and preferred term, classified from verbatim terms. The incidence and percentage of patients with at least one occurrence of a preferred term were included, according to the most severe NCI-CTCAE Version 4.0 grade reported.

End point	Definition
Pharmacokinetics and immunogenicity	Evaluation of the association between tumour PDGFRα expression and clinical outcomes, including PFS, ORR, etc...; and to explore potentially relevant biomarkers of olaratumab

Key: CR, Complete response; ORR, objective response rate; OS, overall survival; PFS, progression free survival; PR, partial response
Source: Eli Lilly submission, Section 4.3, Table 8, p 37.

These endpoints agree with the paper [16] and the protocol for the trial. The ERG considers them reasonable for a study investigating STS.

4.2.1.8 Statistical methods

4.2.1.8.1 Analysis population

The different populations reported within Eli Lilly's submission for their analyses, along with their definitions are presented in Table 9.

Table 9: Analysis population

Analysis Population	Definition
Intent-to-treat population (ITT)	The full study data set from the JGDG study containing data on 133 patients, including patients receiving all lines of treatment.(Eli Lilly submission, p11) The ITT population was used for the analysis of the primary and secondary efficacy endpoints. Subjects in the ITT population were analysed as randomised.
HRQoL evaluable population	HRQoL data were not collected during the JGDG trial.
Safety population	All randomised subjects who had received at least one dose of study treatment. The incidence and percentage of patients with at least one occurrence of a preferred term were included in the analysis. (Eli Lilly submission, p47)

The ITT and safety populations are defined appropriately. There was no evaluable HRQoL population.

4.2.1.8.2 Determination of sample size

Eli Lilly report in their submission that the method for determination of sample size was as follows (Source: Eli Lilly submission, Section 4.4., Table 11, p. 45):

The phase 2 planned sample size was 130 patients (110 PFS events), which assumed a 50% improvement in median PFS (HR 0.67) for the olaratumab plus doxorubicin group, a statistical power of 80%, and a two-sided significance level of 0.20. A planned interim analysis of the primary endpoint was done with a nominal α spend of 0.0001, resulting in a final nominal adjusted α level of 0.1999 (two-sided).

The company have assigned a significance level of 0.20, rather than the standard 0.05, as the JGDG study is a phase 2 study designed to provide preliminary evidence.

4.2.1.8.3 Primary and secondary efficacy analysis

The company report the following for their main efficacy analysis (Source: Eli Lilly submission, Section 4.4, p. 46):

> *Statistical analyses were performed on the ITT population for the following efficacy parameters: PFS, OS, ORR, and change in tumour size. The phase 2 primary endpoint was PFS with analysis powered for a two-sided α (significance) level of 0.2 and statistical power of 80%. A planned interim analysis of the primary endpoint was done with a nominal α spend of 0.0001, resulting in a final nominal adjusted α level of 0.1999 (two-sided). The primary analysis of PFS and the secondary analysis of OS were based on the Mantel's log rank test and the descriptive statistics of Kaplan-Meier. This was reported together with a 95% confidence interval (CI) and included sensitivity and subgroup analyses. The PFS and OS analysis included all randomly assigned patients.*

As mentioned above, the company have assigned a significance level of 0.20, rather than the standard 0.05, as the JGDG study is a phase 2 study designed to provide preliminary evidence.

Further analyses were performed as follows (Source: Eli Lilly submission, Section 4.4, p. 46):

> *Additional analyses were done with Cox proportional hazards models to estimate HRs. Stratified analyses were performed for the primary and all secondary survival analyses. Stratification was planned so that the analyses would account for the 4 IVRS factors. However, as there were very limited numbers of patients either with performance status 2 or PDGFRα negative, the stratified analyses that were conducted used only the other 2 IVRS variables (number of lines of prior therapy and histologic subtype). In addition, since very few patients entered the trial with synovial sarcoma, the histology variable was defined for the stratified analysis as LMS versus non-leiomyosarcoma (non-LMS).*
>
> *The ORR in each treatment group was compared using the Fisher's exact test, and exact confidence bounds (95% CI) were determined. Duration of response was estimated with the Kaplan-Meier method; a 95% CI was provided for the median duration of response.*

The analysis stratification factors were:

- Number of lines of previous treatment (0 vs. 1 or more)

- Histological tumour type (Leiomyosarcoma vs. Other)

With regard to stratification, overstratification can lead to loss of information, but unstratified analyses are not appropriate when there is heterogeneity between strata. Given the variables used for stratification are considered prognostic indicators, this suggests that the stratified analyses may be more appropriate. However, the company have reported both.

Four sensitivity analyses were performed for PFS which involved altering dates of progression and changing censoring criteria (patients were not censored if death or progression occurred after two or more missed visits).

Two sensitivity analyses were undertaken for OS, where survival time was censored from the earliest start date of any post-study anti-cancer therapy, and according to 5 specific regimens.

Safety analyses were performed on the population who had received at least one dose of study treatment. The incidence and percentage of patients with at least one occurrence of a MedDRA™ preferred term were included and summarised.

Subgroup analyses were pre-defined in the protocol and performed on PFS and OS by two randomisation stratification factors; disease histology (leiomyosarcoma and non-leiomyosarcoma) and lines of prior systemic chemotherapy regimens (no prior lines of therapy for advanced disease versus 1 or more prior lines of therapy). (Source: Eli Lilly submission, Section 4.4, p. 48)

OS was analysed within each of several subgroups defined by the potential prognostic factors (see Table 12) and stratification factors (N.B. - there were very limited numbers of patients in the other two randomisation stratification factors; i.e. patients with ECOG PS 2 or PDGFRα negative tumours).

Overall, the ERG agrees the statistical analyses were appropriate.

4.2.2 Results

4.2.2.1 Population distribution

In total, 133 people were randomised. Of these, 64 subjects received OlaDox, and 65 people received Dox. The number of participants evaluable for each of the different population (ITT, safety, evaluable), are presented in Table 10. No HRQoL data were collected.

Table 10: Population distribution for analysis

Analysis population	OlaDox (n=66)	Dox (n=67)

ITT	66	67
Safety[a]	64	65
Evaluable	64	65

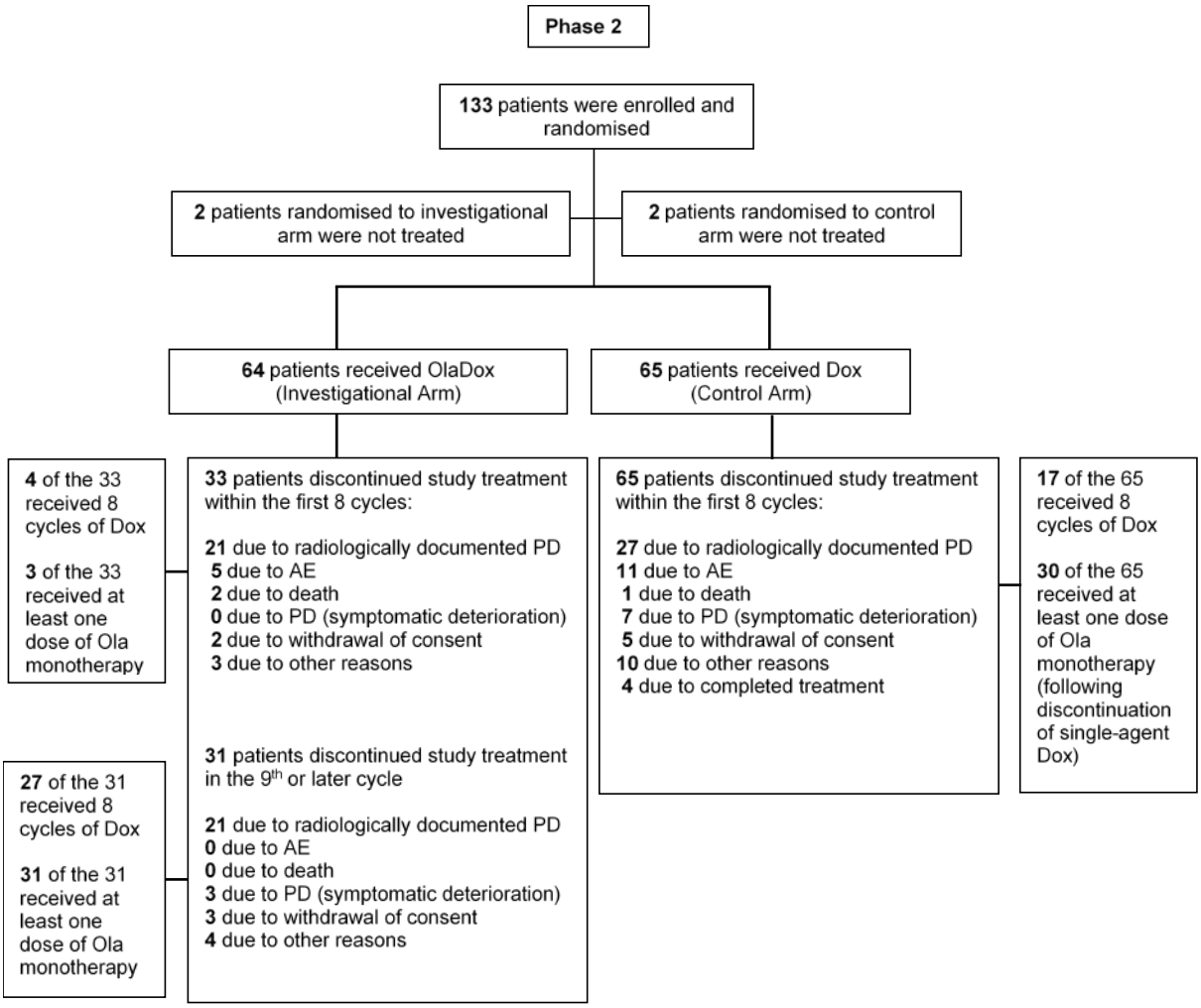
Key: Dox, doxorubicin; OlaDox, olaratumab and doxorubicin; HRQoL, health related quality of life; ITT, intent-to-treat
Notes: a, The company submission also includes 15 participants in the phase 1b trial, to give an OlaDox safety population of 79. Source Eli Lilly submission, Section 4.12, p. 99
Source: Eli Lilly submission, Section 4.5, p. 52

4.2.2.2 Participant flow

The phase 1b trial is mentioned, although this is not an RCT and the participants did not take part in phase 2 (JGDG trial). Fifteen patients were enrolled and treated. The most common reason for discontinuation of study therapy was radiologically documented progression of disease (80%). (Source: Eli Lilly submission, Section 4.5, p. 50)

The participant flow for phase 2 is displayed in Figure 4.

Figure 4: Phase 2 CONSORT diagram



Key: AE, adverse event; Dox, doxorubicin; Ola, olaratumab; PD, progressive disease.
Source: Eli Lilly submission, Section 4.5, p. 51

The ERG feel a frequency table of displaying the number of cycles of Ola and Dox per patient accompanying the CONSORT diagram would provide clarity, particularly since patients in the OlaDox arm, who discontinued doxorubicin early based on dose modification rules, may have received olaratumab monotherapy before completion of cycle eight. Eli Lilly have confirmed that discontinuation of 'study treatment' refers to discontinuation of both olaratumab and doxorubicin in the OlaDox arm. However, if, for example, a participant discontinued Dox after 5 cycles due to Dox toxicity, they may have further cycles of Ola monotherapy, therefore increasing the total number of cycles.

Therefore, we understand that:

For the OlaDox arm, out of the 33 participants who discontinued study treatment ≤ Cycle 8:

- 4 received 8 cycles of OlaDox
- 29 received ≤7 cycles of OlaDox and of these 29, 3 had at least one dose of Ola monotherapy.

For the OlaDox arm, out of the 31 who discontinued study treatment ≥ Cycle 9:

- 27 received 8 cycles of OlaDox and then received ≥1 cycle of Ola monotherapy
- 4 received ≤7 cycles OlaDox, but had at least 9 cycles of treatment in total (including Ola monotherapy).

OlaDox arm overall:

- 34 participants (3+31) received Ola monotherapy, but not all received 8 cycles of Dox.
- 31 participants (4+27) received 8 cycles of OlaDox.

For the Dox arm:

- 17/65 had 8 cycles of Dox.
- 48/65 had ≤7 cycles of Dox.
- 30/65 had at ≥1 cycle of Ola monotherapy.

Finally, 31/64 participants in the OlaDox arm received 8 cycles of Dox compared with 17/65 in the Dox only arm.

4.2.2.3 Baseline characteristics and demographics

Baseline characteristics of the ITT population are summarised in Table A1 (Appendix 1). The demographic characteristics are generally well balanced between those randomised to the

OlaDox and Dox groups. There was a slight imbalance in sex with 60.6% female in the OlaDox arm and 50.7% in the Dox arm. However, the ERG agrees that this is acceptable due to the small sample size. The OlaDox and Dox treatment groups were comparable for baseline disease characteristics.

4.2.2.4 Clinical effectiveness results

4.2.2.4.1 Primary efficacy analysis – progression free survival

The Kaplan-Meier plot of time to PFS is presented in Figure 5 for investigator assessed and Figure 6 for independent assessment, with a summary of both presented in Table 11 (taken from the company's submission).

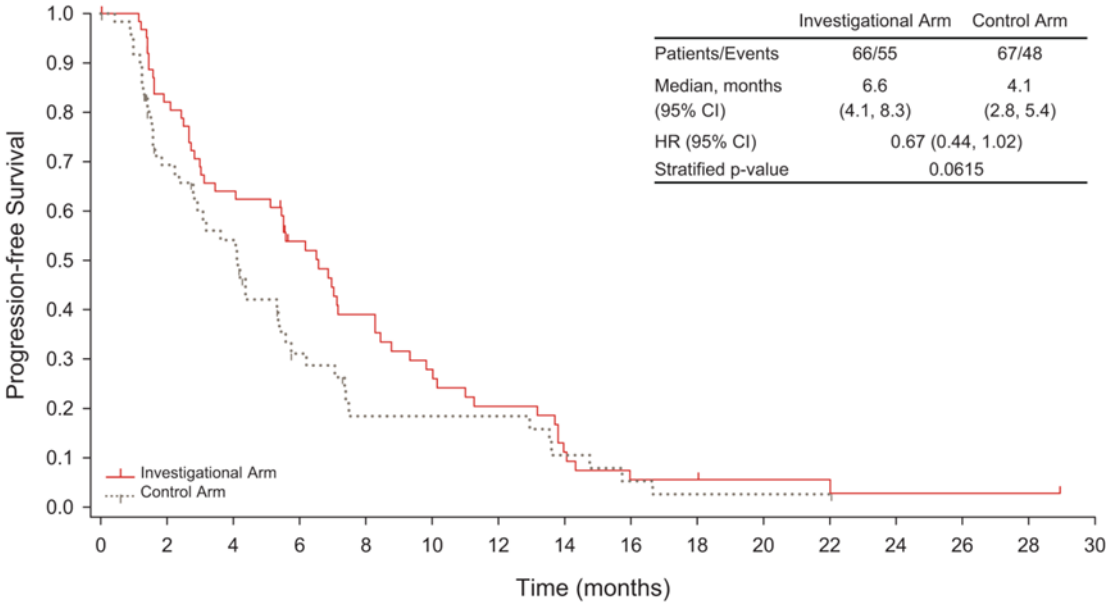
For analysis stratified for number of lines of previous therapy and histological subtype, the study achieved the significance level for PFS, however, this was fixed at 0.1999 rather than the more conventional 0.05 used for phase 3 analysis (Source, Eli Lilly submission, Section 4.7, p. 58):

> *The study met the protocol-defined final significance level for PFS (2-sided alpha = 0.1999), with an investigator-assessed improvement in median PFS of 2.5 months in the OlaDox arm over Dox alone (6.6 months [95% CI: 4.1, 8.3] versus 4.1 months [95% CI: 2.8, 5.4], respectively; HR = 0.672; p = 0.0615).*

Of the 103 events total events, in the blinded, independent assessment, 22 events were not considered PD (Table 11) and 10 were lost or corrupted radiographic images. However, although this increased the median PFS of 3.8 months in the OlaDox arm over Dox, this had a minimal effect on the hazard ratio (HR, 0.670; p, 0.1208).

The PFS analysis was performed with and without stratification for number of lines of previous therapy. Stratification minimises the potential for bias by restricting comparisons to more homogeneous groups. The unstratified results for OlaDox vs Dox following one or more lines of therapy indicate a statistically significant difference (6.5 months [95% CI: 2.8, 8.3] versus 5.4 months [95% CI: 1.6, 5.6], respectively; HR = 0.41; p = 0.0273). However, this is of less relevance to the decision problem since olaratumab is expected to be used in a first-line setting in the UK.

Figure 5: Kaplan-Meier plot of PFS (investigator assessed) for ITT population



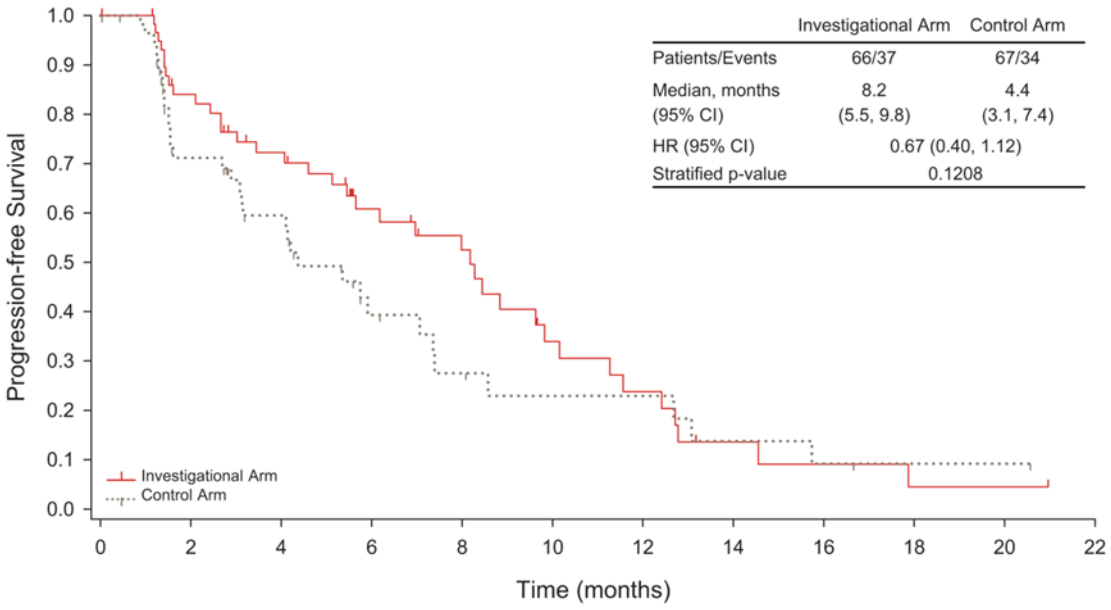
	Investigational Arm	Control Arm
Patients/Events	66/55	67/48
Median, months (95% CI)	6.6 (4.1, 8.3)	4.1 (2.8, 5.4)
HR (95% CI)	0.67 (0.44, 1.02)	
Stratified p-value	0.0615	

Number at Risk

Investigational Arm	66	50	39	29	21	15	11	6	3	3	2	2	1	1	1	0
Control Arm	67	38	28	13	7	7	7	4	2	1	1	1	0	0	0	0

Key: CI, confidence interval; HR, hazard ratio; ITT, intent-to-treat; PFS, progression-free survival.
Source: Eli Lilly submission, Section 4.7, p 61

Figure 6: Kaplan-Meier plot of PFS (independent assessment) for ITT population



	Investigational Arm	Control Arm
Patients/Events	66/37	67/34
Median, months (95% CI)	8.2 (5.5, 9.8)	4.4 (3.1, 7.4)
HR (95% CI)	0.67 (0.40, 1.12)	
Stratified p-value	0.1208	

Number at Risk

Investigational Arm	66	44	34	23	18	10	7	3	2	1	1	0
Control Arm	67	34	24	11	7	5	5	3	2	1	1	0

Key: CI, confidence interval; HR, hazard ratio; ITT, intent-to-treat; PFS, progression-free survival.
Source: Eli Lilly submission, Section 4.7, p 61

Table 11: Summary of PFS for ITT population

	Investigator Assessment		Blinded Independent Review	
	OlaDox N = 66	**Dox N = 67**	**OlaDox N = 66**	**Dox N = 67**
Number of Events, n (%)	55 (83.3)	48 (71.6)	37 (56.1)	34 (50.7)
Number Censored, n (%)	11 (16.7)	19 (28.4)		
No Baseline Tumour Assessments	1 (1.5)	2 (3.0)	7 (10.6)	10 (14.9)
No Post-Baseline Tumour Assessments	2 (3.0)	2 (3.0)		
Death or Progression After Two or More Missed Visits	1 (1.5)	3 (4.5)	2 (3.0)	5 (7.5)
Start of New Anti-cancer Therapy	5 (7.6)	5 (7.5)	18 (27.3)	6 (9.0)
No Documented Progression	2 (3.0)	6 (9.0)	2 (3.0)	11 (16.4)
Withdrew Consent	0	1 (1.5)	0	1 (1.5)
Median[a] (months)	**6.6**	**4.1**	**8.2**	**4.4**
95% CI[a]	(4.1, 8.3)	(2.8, 5.4)	(5.5, 9.8)	(3.1, 7.4)
Q25 - Q75[a]	2.7 – 10.2	1.6 – 7.4	3.0 – 11.6	1.5 – 8.6
3 months PFS Rate[a] (%)	69.0	59.9	76.4	66.7
95% CI[a]	(55.7, 78.9)	(45.9, 71.4)	(62.8, 85.6)	(51.8, 77.9)
Stratified Log-rank p-value[b,d]	0.0615		0.1208	
Stratified Hazard Ratio[c,d]	0.672		0.670	
95% CI[c]	(0.442, 1.021)		(0.401, 1.117)	
Unstratified Log-rank p-value[b,d]	0.1112		0.2157	
Hazard Ratio[c,d]	0.730		0.743	
95% CI[c]	(0.494,1.079)		(0.464, 1.190)	

Key: CI, confidence interval; ITT, intent-to-treat; N, number of randomised patients; n, number of patients in category; PFS, progression-free survival; Q, quartile. Data cut-off date: 15 August 2014.

Notes: a, Estimated by the Kaplan-Meier method; b, Derived from a two–sided test; c, Hazard ratio is expressed as olaratumab + doxorubicin/doxorubicin and estimated from Cox model; d, Between olaratumab + doxorubicin arm and doxorubicin alone arm.

Source: Eli Lilly submission, Section 4.7, Table 17, p 60

The following four sensitivity analyses were performed, with analysis number 4 described as *ad hoc*:

1. If new anti-cancer treatment started before progression, the patient was considered to have disease progression at the date of the new cancer treatment; if death or progression occurred after 2 or more missed visits, the date of death or progression was used; and if lost to follow-up without progression, the patient was considered to have disease progression at the date of the last adequate assessment.
2. Used the actual reported date of progression or death regardless of missing assessments, treatment discontinuation or new anti-cancer treatment.

3. Added clinical progression (symptomatic deteriorations) as progressive events to the primary analysis.
4. Censoring rules were the same as the primary analysis but patients were not censored if death or progression occurred after 2 or more missed visits.

Analyses number 1 and 3 produced a statistically significant result in favour of OlaDox (HR, 0.623; p, 0.0135 and HR 0.631; p, 0.028, respectively). However, all the sensitivity estimates are similar to the primary analysis (HR, 0.672; p, 0.0615).

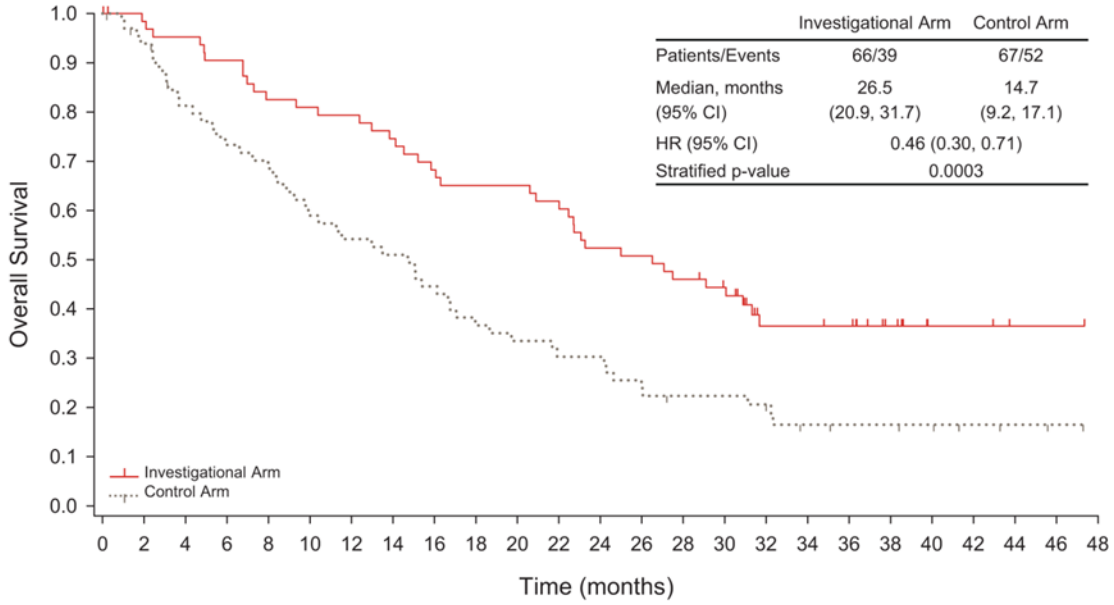
4.2.2.4.2 Secondary efficacy analysis

Overall survival

Analysis was planned when 91 deaths (71%) had occurred; 39 (59.1%) in the OlaDox group and 52 (77.6%) in the Dox group. The median OS was 26.5 months (95% CI: 20.9, 31.7) in the OlaDox arm and 14.7 months (95% CI: 9.2, 17.1) in the Dox arm giving a median OS increase of 11.8 months. The stratified HR gives a statistically significant result of 0.463 (95%CI 0.301 to 0.710).

The Kaplan-Meier plot of time to death from any cause is presented in Figure 7 and a summary of OS is presented in Table 12 (both taken from the submission).

Figure 7: Kaplan-Meier plot of OS for ITT population



Number at Risk

Investigational Arm	66	62	60	57	52	51	50	47	43	41	41	39	33	32	29	26	16	16	15	8	3	3	1	1	0
Control Arm	67	61	51	46	43	37	34	32	28	23	21	19	19	15	13	13	10	7	6	6	5	3	2	1	0

Key: CI, confidence interval; HR, hazard ratio; ITT, intent-to-treat; OS, overall survival
Source: Eli Lilly submission, Section 4.7, p 65

Table 12: Summary of OS for ITT population

	OlaDox N = 66		Dox N = 67
Number of Deaths, n (%)	39 (59.1)		52 (77.6)
Number Censored, n (%)	27 (40.9)		15 (22.4)
Alive, n (%)	25 (37.9)		12 (17.9)
Lost to follow-up, n (%)	0		1 (1.5)
Withdrawal of Consent, n (%)	2 (3.0)		2 (3.0)
Median Survival (months)	**26.5**		**14.7**
95% CI[a]	(20.9, 31.7)		(9.2, 17.1)
Q25 - Q75[a]	13.8 – NE		5.5 – 26.0
3 months Survival Rate[a] (%)	95.2		87.6
95% CI[a]	(86.0, 98.4)		(76.8, 93.6)
6 months Survival Rate[a] (%)	90.5		73.3
95% CI[a]	(80.0, 95.6)		(60.6, 82.5)
Stratified Log-rank p-value[b,d]		0.0003	
Stratified Hazard Ratio[c,d]		0.463	
95% CI[c]		(0.301, 0.710)	
Unstratified Log-rank p-value[b,d]		0.0017	
Hazard Ratio[c,d]		0.517	
95% CI[c]		(0.341, 0.786)	

Key: CI, confidence interval; ITT, intent-to-treat; N, number of randomised patients; n, number of patients in category; NE, not evaluable; Q, quartile.
Notes: a, Estimated by the Kaplan-Meier method; b, Derived from a two–sided test; c, Hazard ratio is expressed as olaratumab + doxorubicin/doxorubicin alone and estimated from Cox model; d, Between olaratumab + doxorubicin arm and doxorubicin alone arm.
Source: Eli Lilly submission, Section 4.7, p 64

The OS analysis was performed with and without stratification. As mentioned for PFS, stratification minimises the potential for bias by restricting comparisons to more homogeneous groups. The pre-specified stratification factors were previous lines of therapy and histological subtype.

Further analysis was performed on the first-line population (65% of participants) which also displayed improved OS in the OlaDox arm (Source: Eli Lilly submission, Section 4.7, p 65):

The median OS was 29.1 months (95% CI: 16.3, NE) in the OlaDox arm and 14.7 months (95% CI: 8.0, 18.7) in the Dox arm (stratified HR = 0.47 [95% CI: 0.27, 0.81]; p = 0.0051).

Sensitivity analysis considering the impact of post-study systemic anti-cancer therapy, including Ola, suggests the main analysis is robust, with regard to improved OS benefit in the OlaDox arm (Table 13).

Table 13: Sensitivity analysis of OS

Overall Survival	Stratified HR (95% CI)[b,c]	p-Value[a,c]
Main Analysis	0.463 (0.301, 0.710)	0.0003
Sensitivity Analysis 1- based on censoring at the date of starting new anti-cancer treatment	0.425 (0.193, 0.933)	0.0284
Sensitivity Analysis 2 -based on censoring at the date of starting selected post-study anti-cancer therapies (pazopanib, eribulin, gemcitabine + docetaxel, doxorubicin, and trabectedin).	0.353 (0.192,0.647)	0.0005
Analysis on Dox participants receiving Ola monotherapy vs. Dox participants not receiving Ola monotherapy	1.013	0.9660

Key: CI, confidence interval; HR, hazard ratio.
Notes: a, Derived from a two–sided test; b, Hazard ratio is expressed as olaratumab + doxorubicin/doxorubicin alone and estimated from Cox model; c Between olaratumab + doxorubicin arm and doxorubicin alone arm.
Source: Eli Lilly submission, Section 4.7, p 67

Table 13 also shows that there was no difference in OS between patients in the Dox arm receiving Ola monotherapy and those who did not.

Post hoc sensitivity analysis was performed on the number of cycles of Dox therapy received. Exposure data are summarised in Table 14.

Table 14: Exposure to OlaDox, Dox or Ola monotherapy by study arm

	OlaDox Arm N = 64	Ola monotherapy N = 34	Dox Arm: Ola monotherapy N = 30	OlaDox Arm N = 64	Dox Arm N = 65
	Duration of Ola Treatment (wks)			Duration of Dox Treatment (wks)	
Mean (SD)	31.4 (26.71)	25.6 (25.16)	17.6 (31.25)	17.6 (7.72)	13.6 (8.21)
Median	26.1	15.1	7.0	21.3	12.3
Range	3.0 – 128.0	4.0 – 104.0	3.0 – 134.0	3.0 – 29.0	3.0 – 25.4

Source: Eli Lilly submission, Section 4.1, p 100

It is clear from Table 14 that the OlaDox arm received a greater exposure to doxorubicin than the Dox arm (median 21.3 weeks vs. 12.3 weeks). This is explained by the company as follows (Source: Eli Lilly submission, Section 4.7, p 69):

> *Per protocol, patients were to be treated with Dox either in combination with Ola or as single agent, up to 8 cycles, or until PD, whichever came first. Disease progression was the main cause for Dox treatment discontinuation, on both arms (63.6% in the OlaDox arm, 40.3% in the Dox arm). These data support that the observed difference in number of cycles of Dox administered between the 2 study arms is likely attributable to the earlier time of progression for patients in the Dox arm.*

Post hoc sensitivity analyses were performed to establish the effect of the number of Dox cycles on OS (Table 15).

Table 15. Sensitivity Analyses of OS by Exposure to Doxorubicin (ITT)

	OlaDox N = 66	Dox N = 67
Excluding patients discontinuing study treatment within 8 cycles due to AE or symptomatic PD		
Patients	61	49
OS Events	35	36
Median, months	26.8	16.1
Unstratified OS HR	0.55	
Unstratified log rank p-value	0.012	
Excluding patients completing <4 cycles doxorubicin		
Patients	49	38
OS Events	24	28
Median, months	31.7	17.1
Unstratified OS HR	0.47	
Unstratified log rank p-value	0.005	
Excluding patients completing <5 cycles doxorubicin		
Patients	41	31
OS Events	20	22
Median, months	31.7	17.5
Unstratified OS HR	0.51	
Unstratified log rank p-value	0.027	
Excluding patients completing <6 cycles doxorubicin		
Patients	39	28
OS Events	18	19
Median, months	31.7	18.7
Unstratified OS HR	0.51	
Unstratified log rank p-value	0.038	

Source: Eli Lilly submission, Section 4.7, p 70

Eli Lilly suggest that if the benefit in OS was due to the increased exposure to Dox in the OlaDox arm, then the event rate between groups would become more balanced (i.e. HR closer to one) as the participants receiving fewer cycles of Dox were removed.

Patients Electing to Receive Olaratumab monotherapy after Progression on the Dox arm

Following disease progression occurring during or after completion of doxorubicin single agent treatment, 30/65 patients on the Dox arm received olaratumab monotherapy. For PFS, the company state that (Source: Eli Lilly submission, Section 4.7, p 70):

> *PFS values for Dox arm patients electing to receive olaratumab monotherapy were numerically lower when compared to Dox arm patients that did not elect to receive olaratumab monotherapy*

With regard to OS, there was no statistically significant difference whether the patients had subsequently received Ola monotherapy or not (HR 1.013; p value 0.9660). The median OS with olaratumab monotherapy was (13.5 months; 95% CI: 8.4, 21.7) versus those not receiving olaratumab (15.1 months; 95% CI: 5.9, 18.7)

Objective response rate

Eli Lilly report PR, CR and ORR following both investigator and independent assessment (Table 16). The results were numerically favourable for OlaDox, however, the result was not statistically significant.

Table 16: Response to Treatment (Investigator & Independent Assessments) (ITT)

	Investigator Assessment			Blinded Independent Assessment		
	OlaDoxN=66)	Dox (N=67)	P-Value	OlaDox(N=66)	Dox (N=67)	P-Value
Best Overall Response, n %						
Complete response (CR)	2 (3.0)	1 (1.5)		3 (4.5)	1 (1.5)	
Partial response (PR)	10 (15.2)	7 (10.4)		9 (13.6)	4 (6.0)	
Stable disease (SD)	39 (59.1)	34 (50.7)		37 (56.1)	36 (53.7)	
Progressive disease (PD)	11 (16.7)	15 (22.4)		11 (16.7)	15 (22.4)	
Not evaluable (NE)	4 (6.1)	10 (14.9)		6 (9.1)	11 (16.4)	
Objective response rate (ORR) (CR+PR), n %						
No.of patients(%)	12 (18.2)	8 (11.9)	0.3421[b,c]	12 (18.2)	5 (7.5)	0.0740 [b,c]
95% CI[a]	9.8, 29.6	5.3, 22.2	0.3214[b,d]	9.8, 29.6	2.5, 16.6	0.0679[b,d]
Disease control rate (CR+PR+SD), n %						
No. of patients (%)	51 (77.3)	42 (62.7)		49 (74.2)	41 (61.2)	
95% CI[a]	65.3, 86.7	50.0, 74.2		62.0, 84.2	48.5, 72.9	

Key: CI, confidence interval; ITT, intent-to-treat; N, number of randomised patients; n, number of patients in category.

Notes: a, Estimated using binomial distribution; b, Between olaratumab + doxorubicin arm and doxorubicin alone arm; c, Derived from two-sided Fisher's exact test, d, Derived from 2-sided Cochran-Mantel-Haenszel test adjusted by the stratification factor.

Source: Source: Eli Lilly submission, Section 4.7, p 72

4.2.2.4.3 Subgroup analysis

Pre-planned subgroup analysis was performed for PFS and OS as follows:

Stratification factors

- PDGFRα (exploratory assay) (Positive vs. Negative)
- Number of lines of previous treatment (0 vs. 1 or more)
- Histological tumour type (Leiomyosarcoma vs. Other)
- ECOG PS (0-1 vs. 2)

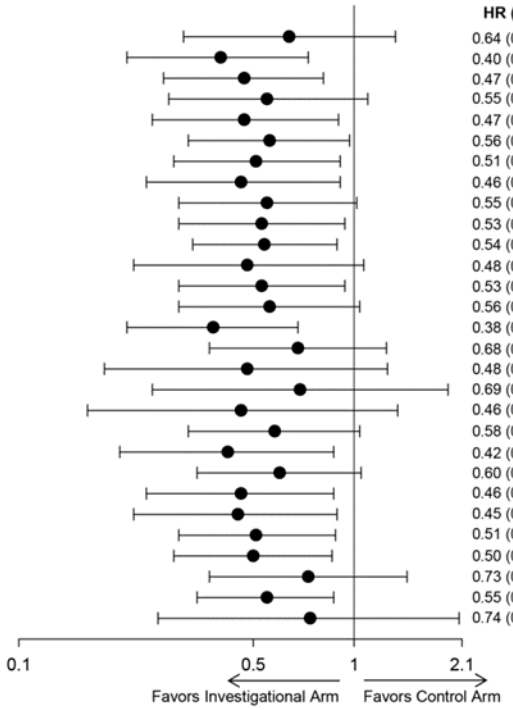
Other baseline characteristics

- Sex (Male vs. Female)
- Age (years)(18 <65 vs. ≥65)
- Duration of disease (months) (≤ 9 vs. >9)
- Platelets at baseline ($\times 10^9$/L) (≤300 vs. >300)
- White blood cell count at baseline ($\times 10^9$/L) (≤10 vs. >10)
- Primary tumour present (Y vs. N)

The forest plot of the subgroup analyses for OS is displayed in Figure 8.

Figure 8: Forest plot of OS subgroup hazard ratios (with 95% CI) (ITT Population)

Category	Subgroup	Investigational Arm # Events	Investigational Arm N	Control Arm # Events	Control Arm N	HR (95% CI)
PDGFRα (Assay 2)	Positive	14	18	17	19	0.64 (0.31, 1.33)
	Negative	19	37	26	37	0.40 (0.21, 0.73)
Number of lines of previous treatment	0	21	40	36	47	0.47 (0.27, 0.81)
	1 or more	18	26	16	20	0.55 (0.28, 1.10)
Histological tumor type	Leiomyosarcoma	16	24	24	27	0.47 (0.25, 0.90)
	Other	23	42	28	40	0.56 (0.32, 0.97)
ECOG PS	0	20	36	30	38	0.51 (0.29, 0.91)
	1	16	26	19	26	0.46 (0.24, 0.91)
Sex	Male	16	26	28	33	0.55 (0.30, 1.02)
	Female	23	40	24	34	0.53 (0.30, 0.94)
Age (yrs)	18-<65	30	48	33	43	0.54 (0.33, 0.89)
	≥65	9	18	19	24	0.48 (0.22, 1.07)
Weight (kg)	<81.4	19	29	31	37	0.53 (0.30, 0.94)
	≥81.4	20	37	21	30	0.56 (0.30, 1.04)
Duration of disease (mos)	<14.95	20	33	29	34	0.38 (0.21, 0.68)
	≥14.95	19	33	23	33	0.68 (0.37, 1.25)
Duration of most recent prior systemic treatment (mos)	<4.12	8	11	10	12	0.48 (0.18, 1.26)
	≥4.12	10	15	6	8	0.69 (0.25, 1.91)
Grade	Grade 1-2	5	12	10	15	0.46 (0.16, 1.35)
	Grade 3	21	29	25	29	0.58 (0.32, 1.04)
	Unknown/Not assessed	13	25	17	23	0.42 (0.20, 0.87)
Albumin at baseline (g/L)	<38.0	22	29	30	37	0.60 (0.34, 1.05)
	≥38.0	17	37	22	30	0.46 (0.24, 0.87)
Liver metastases	Yes	17	26	19	22	0.45 (0.22, 0.89)
	No	22	40	33	45	0.51 (0.30, 0.88)
Platelets at baseline (10E9/L)	≤300	26	49	29	41	0.50 (0.29, 0.86)
	>300	13	17	23	26	0.73 (0.37, 1.44)
WBC at baseline (10E9/L)	≤10	34	59	38	52	0.55 (0.34, 0.87)
	>10	5	7	14	15	0.74 (0.26, 2.06)



Key: ECOG, Eastern Cooperative Oncology Group; PDGFRα, platelet-derived growth factor receptor α; PS, performance status.

Notes: Subgroup analyses are based on electronic case report form (eCRF) entries. Duration of disease is the time from date of histology or pathology confirmation of soft-tissue sarcoma to date of informed consent. Critique of trials identified and included in the indirect comparison and/or multiple treatment comparison

Source: Eli Lilly submission, Section 4.8, p 73

Overall, the results are consistent with the Ola arm being favoured for improvement in OS. However, several of the results are not statistically significant and the confidence intervals are very broad. Furthermore, in some cases, the population available for analysis is very small.

A stratified Cox multivariate model of OS was also performed and adjusted for OS specific prognostic factors to give the following results (Source: Eli Lilly submission, Section 4.8, p 73):

After adjusting for these factors, the OS HR for the treatment effect was 0.429 (95% CI: 0.267, 0.690), consistent with the stratified univariate OS HR of 0.463 reported for the primary OS analysis. The OS HR from the multivariate model provides evidence that the OS outcome was not biased by any baseline imbalances in these potential prognostic factors.

Overall results of PFS and OS according to histology are described below. No further details are given for PFS subgroup analysis (Source: Eli Lilly submission, Section 4.8, p 74):

The PFS HRs are similar between LMS (HR = 0.795) and non-LMS (HR = 0.702). Significant improvements in OS were observed with OlaDox vs. Dox for both subtypes, although the magnitude of improvement in the LMS subgroup was greater than the non-LMS (15.1 month improvement; HR = 0.473 [95% CI: 0.248, 0.900]; p = 0.0198; and 7.3 month improvement; HR = 0.556[95% CI: 0.320, 0.967]; p = 0.0348, respectively).

4.2.2.4.4 Adverse events

JGDG included 129 subjects in the safety population, which includes 64 in the OlaDox arm and 65 in the Dox arm. Thirty participants in the Dox arm received at least one dose of olaratumab monotherapy after discontinuation of single agent doxorubicin.

Some imbalance was evident in reported AEs (Source: Eli Lilly submission, Section 4.5, p. 52):

More patients in the Dox arm than in the OlaDox arm discontinued study treatment for reasons of AEs (16.4% vs. 7.6%). The imbalance may be a reflection of the different definitions of treatment discontinuation between study arms, that is, in the OlaDox arm, discontinuation of one agent in the absence of discontinuation of the other was not captured as treatment discontinuation.

The increased AE's in the OlaDox arm may also be a reflection of the extent of Dox exposure (Source: Eli Lilly submission, Section 4.12, p100):

The median number of cycles of doxorubicin received was greater in the OlaDox arm (median: 21.3 weeks or approximately 7.1 cycles) compared to the Dox arm (median: 12.3 weeks or approximately 4.1 cycles). Consequently, the cumulative dose of doxorubicin received was higher in the OlaDox arm (mean: 416.4 mg/m2; median: 487.6 mg/m2) compared to the Dox arm (mean: 328.9 mg/m2; median: 299.6 mg/m2).

the difference in exposures between arms reflects the longer PFS for patients on the OlaDox arm, which permitted them to receive more cycles of doxorubicin (median, 7.0 cycles) compared to patients on the Dox arm (median, 4.0 cycles).

A summary for adverse events (AEs) is presented in Table 17, which reports the incidences of AEs for > 10 % of people in any treatment arm. Tables A3 and A5 (Appendix 1) provide more details on AEs for both treatment arms. Note that these analyses do not include events occurring during the period of olaratumab monotherapy after discontinuation of doxorubicin in the Dox arm, unless otherwise specified.

Table 17: Summary of Treatment-Emergent Adverse Events

	OlaDox N = 64 n (%)	Dox N = 65 n (%)
Any AE	63 (98.4)	64 (98.5)
Related to any Study Drug	63 (98.4)	63 (96.9)
Related to Olaratumab	56 (87.5)	NA
Related to Doxorubicin	62 (96.9)	63 (96.9)
Any Serious Adverse Event	27 (42.2)	25 (38.5)
Related to any Study Drug	14 (21.9)	17 (26.2)
Related to Olaratumab	10 (15.6)	NA
Related to Doxorubicin	12 (18.8)	17 (26.2)
Any Grade ≥3 AE	51 (79.7)	45 (69.2)
Related to any Study Drug	43 (67.2)	36 (55.4)
Related to Olaratumab	29 (45.3)	NA
Related to Doxorubicin	40 (62.5)	36 (55.4)
Any AE Leading to Discontinuation of any Study Drug	8 (12.5)	12 (18.5)
Any AE Leading to Discontinuation of Olaratumab Only	1 (1.6)	NA
Any AE Leading to Discontinuation of Doxorubicin Only	3 (4.7)	12 (18.5)
Any AE Leading to Discontinuation of both Olaratumab and Doxorubicin	4 (6.3)	0
Any AE with Outcome of Death within 30 Days of Last Dose	0	5 (7.7)[a]
Related to any Study Drug	0	2 (3.1)
Related to Olaratumab	0	NA
Related to Doxorubicin	0	2 (3.1)

Key: AE, adverse event; N, number of treated patients; NA, not applicable.
Notes: a, Deaths are counted for both the doxorubicin treatment and during the olaratumab monotherapy stage. There were 4 deaths that occurred within 30 days of last dose of doxorubicin. There was 1 death that occurred after the patient received olaratumab monotherapy; Adverse event with missing or unknown relationship to study drug is counted as 'related'.
Source: Eli Lilly submission, Section 4.12, p 102

Table 18: Grade 3 to 4 Treatment emergent adverse events occurring in ≥10% of patients in any treatment group

	OlaDox (n=64)		Dox (n=65)	
	Grade 3	**Grade ≥4**	**Grade 3**	**Grade ≥4**
Patients with any adverse event[a,b]	24 (38%)	27 (42%)	25 (38%)	20 (31%)
Neutropenia[c,d]	12 (19%)	22 (34%)	5 (8%)	16 (25%)
Anaemia[e]	8 (13%)	0	6 (9%)	0
Leucopoenia[d, f]	14 (22%)	9 (14%)	5 (8%)	6 (9%)
Febrile neutropenia[g]	7 (11%)	1 (1.6%)	9 (14%)	0
Treatment-related adverse event	18 (28%)	25 (39%)	19 (29%)	17 (26%)
Adverse event leading to discontinuation of treatment	1 (2%)	3 (5%)	3 (5%)	5 (8%)
Serious adverse event				
Any event	20 (31%)	7 (11%)	14 (22%)	8 (12%)
Treatment-related event	8 (13%)	6 (9%)	11 (17%)	5 (8%)

Notes: a, Adverse events and clinical laboratory toxicity were graded according to the National Cancer Institute Common Terminology Criteria for Adverse Events (version 4.0); b, The adverse events listed here were reported in at least 15% of patients in the olaratumab plus doxorubicin group, except as noted in footnote; c, Consolidated term comprising the following preferred terms: neutropenia and neutrophil count decreased; d, Some patients reported both neutropenia and leukopenia; e, Consolidated term comprising the following preferred terms: anaemia and haemoglobin decreased; f, Consolidated term comprising the following preferred terms: leukopenia and white blood cell count decreased; g, These events are included here because they were considered clinically important.

Source: Eli Lilly submission, Section 4.12, p 103

The percentages of patients who experienced ≥1 TEAE and ≥1 treatment-emergent SAE were generally similar between the two arms, however, Table 18 shows an increase of 13% Grade 4 TEAE in the OlaDox arm and in Table 17, any Grade 3 AE was 79.7% in OlaDox compared with 69.2% in the Dox arm. In contrast, the percentage of patients with ≥1 TEAE leading to discontinuation of any study drug was lower for the OlaDox arm compared with the Dox arm (12.5% vs. 18.5%). Furthermore, there were no TEAEs resulting in death in the OlaDox arm, whereas there were 5 deaths in the Dox arm (which includes one patient receiving Ola monotherapy.

The company describe the most common AEs leading to discontinuation of treatment as follows (Source: Eli Lilly submission: Eli Lilly submission, Section 4.12, p 102):

The most common adverse event leading to patient discontinuation of doxorubicin was ejection-fraction decrease in 3 (5%) of 64 patients in the OlaDox arm and 4 (6%) of 64 patients in the Dox arm. The most common adverse event leading to discontinuation of olaratumab was infusion-related reaction in two (3%) of 64 patients.

For OlaDox, the most common Grade 3 to 4 TEAEs was neutropenia at 53%, as compared to 33% for Dox. Febrile neutropenia was, however, similar between arms (OlaDox 13% vs. Dox 14%).

Death due to disease progression was reported for 38 patients in the OlaDox arm and 44 patients in the Dox arm. One death in both the OlaDox and Dox arm had an unknown cause and a further six deaths in the Dox arm attributed to aspirational pneumonia, respiratory failure, sepsis, septic shock, and small bowel obstruction.

With regard to AEs as a result of the cardiotoxicity of doxorubicin (Source: Eli Lilly submission, Section 4.12, p 105):

> *The prevalence of cardiac dysfunction … was 23% (15 patients) with OlaDox and 17% (11 patients) with Dox. Excluding the patients with peripheral oedema (none reported other adverse events to suggest cardiac dysfunction), the total prevalence of cardiac dysfunction was 8% (five patients) with OlaDox and 6% (four patients) with Dox.*

The ERG agrees with the company that (Source: Eli Lilly submission, Section 4.12, p 105):

> *…the safety results of study JGDG demonstrate that OlaDox has an acceptable, monitorable and generally manageable tolerability profile despite a significant difference in median cumulative doxorubicin exposure between the 2 treatment arms.*

4.2.3 Interpretation

Key efficacy findings from the RCT reported from the submission were as follows:

Progression free survival

At the significance level of 0.1999, OlaDox was superior to Dox monotherapy for progression free survival. The investigator-assessed improvement in median PFS was 2.5 months in the OlaDox arm over Dox alone (6.6 months [95% CI: 4.1, 8.3] versus 4.1 months [95% CI: 2.8, 5.4], respectively; HR = 0.672; p = 0.0615).

Secondary endpoints

In the ITT population and first-line setting, OlaDox improved OS benefit: the median OS was 29.1 months (95% CI: 16.3, NE) in the OlaDox arm and 14.7 months (95% CI: 8.0, 18.7) in the Dox arm.

For ORR, the results were numerically favourable for OlaDox as opposed to Dox montherapy, however, the results were not statistically significant.

No HRQoL data were available.

Adverse events

The percentages of patients who experienced ≥1 TEAE and ≥1 treatment-emergent SAE were generally similar between the two arms, however, there was a higher number of participants experiencing Grade 3 AE and Grade 4 TEAE in the OlaDox arm. This imbalance may be a reflection of the higher Dox exposure in the OlaDOx arm. The percentage of patients with ≥1 TEAE leading to discontinuation of any study drug was lower for the OlaDox and, there were no TEAEs resulting in death in the OlaDox arm, whereas there were 5 deaths in the Dox arm.

4.2.3.1 Strengths and limitations

Strengths

- Multicentre, appropriately randomised design of the RCT JGDG
- The population recruited to study JGDG was representative of the typical UK patient population

Limitations

- The JGDG trial is a Phase 2 study intended to provide preliminary data, therefore the population is small and the significance level set at 0.1999 (rather than the more conventional 0.05 for larger trials). The increased significance level reduces the region of acceptance i.e., there is more likelihood of rejecting the null hypothesis (that there is no significant difference between arms).
- The open-label design introduces the risk of bias
- The maximum number of cycles of doxorubicin in the JGDG is eight, whereas standard UK clinical practice is 6.
- The use of subsequent therapies following treatment assignment

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

Due to the lack of any other direct comparisons between OlaDox and the comparator treatments (Dox only and four regimens using a combination of Dox and Ifo at different doses), no direct pairwise meta-analyses were possible. Hence, the company used a mixed treatment comparison (MTC) method.

The company performed a systematic literature review to identify studies that investigated drug treatments for advanced STS that was not amenable to alternative treatments (see Section 4.1.1, p 40 for details on searches and inclusion criteria).

From all sources included in the searches, 248 publications were eligible based on the inclusion/exclusion criteria, of which 179 were single arm trials and 69 were RCTs. Of these RCTs, 28 linked to OlaDox via common comparators included within the trial, and 6 included interventions being considered in the decision problem. Four of these were phase 3 trials (Santoro, Le Cesne, Judson, Seddon)[19-22] and two were phase 2 trials (Tap, Maurel).[16, 23] Table 19 presents a summary of the six studies included in the NMA. Further details on participant characteristics are in Appendix 4. Judson et al (2014)[21] have the youngest population (median age, IfoDox 47 years; Dox 48 years) in comparison to Tap et al (2016)[16] with the oldest population (median age, OlaDox 58.5 years; Dox 58 years). In contrast to the other studies who had a higher proportion of females to males, Maurel et al (2009) [23] had a higher percentage of males, (IfoDox 55%; Dox 61%). The other characteristics, although poorly reported, appear generally comparable.

Additional points noted by the company include (Eli Lilly submission, section 4.10):

- *It appears that all the studies were open label, although three of them do not state this explicitly.*
- *Two studies, patients in the control arm were able to receive the investigational drug after disease progression. In Tap (2016), patients randomised to Dox could receive Ola monotherapy after disease progression, and in Judson (2014) (39), patients randomised to IfoDox could receive Ifo monotherapy after progression.*
- *The four phase 3 studies included more than twice the number of patients per arm than the phase 2 studies (128-258 patients vs. 64-67 patients) and had longer follow-up times (up to 59 months vs. 12 months).*
- *All the studies included patients with various types of STS. Doxorubicin was the control intervention in all studies, except for Le Cesne (2000), which had IfoDox in the control arm.*
- *The company found most included studies were poorly reported, particularly with regard to randomisation and treatment allocation*

Table 19: Characteristics of the trials used to carry out the NMA

Trial Name/ Study Design	Median age (range)	Comparison	Population Analysed (No. of Patients)	PFS						OS						ORR (%)
				Median (months)	95% CI		HR	95% CI		Median (months)	95% CI		HR	95% CI		
					Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper	
I5B-IE-JGDG Tap (2016) [16]	58	OlaDox	ITT population (n = 66); 2 years minimum follow-up	6.6	4.1	8.3	0.73[a] 0.672[b]	0.494[a] 0.442[b]	1.079[a] 1.021[b]	26.5	20.9	31.7	0.517[a] 0.463[b]	0.341 0.301	0.786 0.710	18.2
		Dox	ITT population (n = 67); 2 years minimum follow-up	4.1	2.8	5.4	-	-	-	14.7	9.2	17.1	-	-	-	11.9
Maurel et al. (2009) [23]	49	IfoDox	ITT (n = 64); follow-up = 12.3 months	5.5	NR	NR	1.3	0.71	1.48	NR	NR	NR	0.71	0.45	1.13	24.1
		Dox	ITT (n = 67); follow-up = 12.3 months	6.0	NR	NR	-	-	-	NR	NR	NR	-	-	-	23.4
EORTC 62851 Santoro et al. (1995) c [19]	51	IfoDox	ITT (n = 258); follow-up = 55.2 months	11.0	NR	NR	NR	NR	NR	12.7	NR	NR	NR	NR	NR	NR
		Dox	ITT (n = 263); follow-up = 55.2 months	10.6	NR	NR	-	-	-	12.0	NR	NR	-	-	-	NR
EORTC 62903 Le Cesne et al. (2000) [20]	50	IfoDox (high dose, 6 cycles)	ITT (n = 145); follow-up = NR	6.7	NR	NR	NR	NR	NR	12.65	NR	NR	NR	NR	NR	21.4
		IfoDox (low dose, 1 cycles)	ITT (n = 149); follow-up = NR	4.4	NR	NR	-	-	-	12.88	NR	NR	-	-	-	20.8
EORTC 62012 Judson et al. (2014)		IfoDox	ITT (n = 227); follow-up = 59 months	7.4	6.6	8.3	0.74	0.6	0.9	14.3	12.5	16.5	0.83	0.67	1.03	28.4
	48	Dox	ITT (n = 228); follow-up = 56 months	4.6	2.9	5.6	-	-	-	12.8	10.5	14.3	-	-	-	14.4

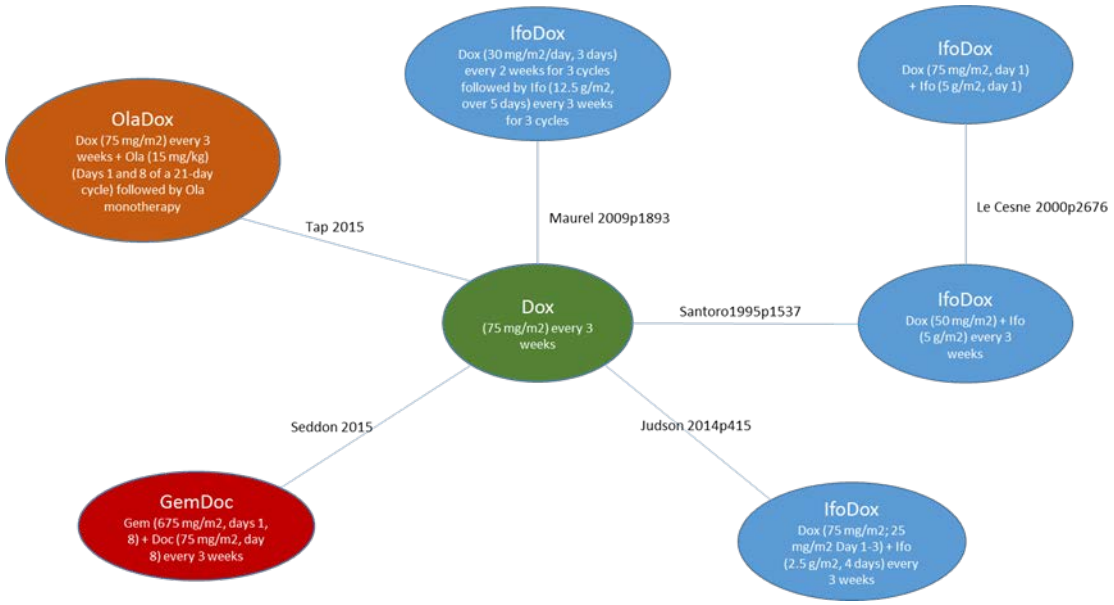
[21]																
GeDDiS Seddon et al. (2015)[22]	55	GemDoc	ITT (n = 128); follow-up = 19 months	5.5	NR	NR	1.28	0.98	1.67	14.5	NR	NR	1.07	0.77	1.49	NR
		Dox	ITT (n = 129); follow-up = 19 months	5.4	NR	NR				16.4	NR	NR				NR[d]

Key: CI, confidence interval; Dox, doxorubicin; EORTC, European Organisation for Research and Treatment of Cancer; GemDoc, gemcitabine + docetaxel; HR, hazard ratio; IfoDox, ifosfamide + doxorubicin; ITT, intention-to-treat; Lilly, Eli Lilly and Company; NR, not reported; OlaDox, olaratumab + doxorubicin; ORR, objective response rate; OS, overall survival; PFS, progression-free survival; STS, soft tissue sarcoma.

Notes: a, unstratified; b, stratified c, Trial EORTC 6281 has an additional arm evaluating the treatment cyclophosphamide + vincristine + Dox + dacarbazine, which was not included in the meta-analysis; d, complete and partial response were not reported separately from stable disease by Seddon et al., 2015. Data for complete and partial response and stable disease combined were used in the NMA for Seddon et al., 2015, in a model that accounted for different patterns of available response data approach.

A network diagram is presented (Figure 9) showing that OlaDox is linked to Dox monotherapy via the Tap study[16], Dox monotherapy is linked to four additional interventions: three different regimens of IfoDox, and GemDoc (not considered within the decision problem). One of these IfoDox regimens is linked to a further IfoDox regimen that is not directly tested against Dox monotherapy; thus, there are five IfoDox regimens in total. Of note, the Le Cesne study[20] includes two IfoDox regimens, one of which also includes a recombinant human granulocyte-macrophage colony-stimulating factor (rhGM-CSF), sargramostim, as part of the regimen; this is not mentioned in the company report. Thus, the six trials generate a network of seven nodes (treatments), one of which (GemDoc) is not considered within the decision problem.

Figure 9: Evidence for the network meta-analysis.



Key: Doc, docetaxel; Dox, doxorubicin; Gem, gemcitabine; GemDoc, gemcitabine + docetaxel; Ifo, ifosfamide; IfoDox, ifosfamide + doxorubicin; Ola, olaratumab; OlaDox, olaratumab + doxorubicin.
Notes: This figure cites study JGDG (OlaDox vs Dox) as Tap 2015 based on the CTOS conference abstract, which preceeded the Lancet paper (Tap 2016)
Source: Eli Lilly submission, Section 4.10, p80.

Of the remaining interventions, four are compared directly with Dox monotherapy. The company states that expert clinical opinion supports the inclusion of each IfoDox regimen as a separate treatment node, rather than collapsing the IfoDox regimens into one overall node. However, the company also states that clinical opinion supports the view that the different IfoDox regimens can be considered to have equal efficacy, possibly implying that the clinical differences lie in the toxicity profiles. Therefore, it may have been reasonable to perform a sensitivity analysis that collapsed (where feasible) the IfoDox regimens into one treatment

node. This approach would reflect the approach taken in the economic modelling, whereby one of five the IfoDox regimens, that used in the Judson trial[21], which consisted of 75mg/m2 Dox and 10g/m2 Ifo, was used as a generic representation of an IfoDox regimen that would be suitable for comparison with Dox monotherapy and OlaDox. An advantage of this approach would be to reduce the complexity of the network.

There are no trials in the network with more than two arms. The company also presented a wider network including 17 interventions that are linked to Dox monotherapy and therefore can be linked to OlaDox, but 11 of these are not included within the decision problem.

As shown in Table 20, the doses used for IfoDox varied across studies, therefore they were used as separate interventions, rather than pooled.

Table 20: Intervention and comparator doses used in trials for NMA

Author	Intervention	Comparator
Judson et al. 2014	IfoDox: doxorubicin (75 mg/m^2) + ifosfamide (10g/m^2) every 3 weeks up to 6 cycles	Dox: doxorubicin (75 mg/m^2) every 3 weeks up to 6 cycles
Le Cesne et al 2000	IfoDox: doxorubicin (75 mg/m^2) + ifosfamide (5 g/m^2) every 3 weeks up to 6 cycles	IfoDox: doxorubicin (50 mg/m^2) + ifosfamide (5 g/m2) every 3 weeks up to 10 cycles
Maurel et al. 2009	IfoDox: doxorubicin (90 mg/m^2) every 2 weeks for 3 cycles followed by ifosfamide (12.5 g/m^2) every 3 weeks up to 3 cycles	Dox: doxorubicin (75 mg/m^2) every 3 weeks up to 6 cycles
Santoro et al 1995	IfoDox: doxorubicin (50 mg/m^2) + ifosfamide (5 g/m^2) every 3 weeks up to 8 cycles	Dox: doxorubicin (75 mg/m^2) every 3 weeks up to 8 cycles
Seddon	GemDoc: gemcitabine (675 mg/m^2) + docetaxel (75 mg/m^2) every 3 weeks for up to 6 cycles	Dox: doxorubicin (75 mg/m^2) every 3 weeks up to 6 cycles
Tap et al. 2016	OlaDox: doxorubicin (75 mg/m^2) + olaratumab (15 mg/kg) every 3 weeks up to 8 cycles	Dox:doxorubicin (75 mg/m^2) every 3 weeks up to 8 cycles

Endpoints considered within the meta-analysis were PFS, OS, ORR, and discontinuation of treatment due to drug toxicity. A further potential endpoint, quality of life, was not included in

the meta-analysis, as the relevant data were rarely reported. Only PFS and OS are relevant for the economic modelling.

Of the six included studies, two allowed patients in the control arm (Dox monotherapy) to receive the investigational drug after disease progression; these were the Tap study[16], where the patients were allowed to receive Ola monotherapy, and the Judson study[21] where patients were allowed to receive Ifo monotherapy. The four phase 3 studies had more patients and longer follow-up times compared with the phase 2 studies. Such differential follow-up time may lead to bias when using models that assume proportional hazards, in the event that the proportional hazards assumption is not valid. A notable difference between the Tap study and the other trials was that the Tap study included patients who had previously received other forms of therapy (except anthracyclines), whereas the other studies included patients who had received no prior chemotherapy.[16] The company reports that all studies included the Kaplan–Meier plot for OS and five out of six (excepting Santoro[19]) included the Kaplan–Meier plot for PFS. To promote comparability among studies, PFS was assessed by investigator (rather than independent observer) in the Tap study[16], as two other studies also used investigator assessment (Judson and Maurel)[21, 23], while in one study the method of PFS assessment was not reported (Seddon)[22]. The company does not mention the method of PFS assessment for the Santoro study.[19] The company states that only trials that clearly reported the endpoint of interest were included in the meta-analysis for that endpoint.

4.3.1 Quality assessment of the included trials

There are some concerns regarding the quality of the included trials. For example, for three trials (Maurel, Le Cesne, Seddon)[20, 22, 23], there was not sufficient clarity to determine whether randomisation had been carried out appropriately, and for five of the six studies (all except Tap)[16] it was not clear whether treatment allocation was adequately concealed. Hence, the risk of bias cannot be excluded for these studies.

The outcomes available for each outcome across studies are set out (Eli Lilly submission, Table 28, p89). For OS, HRs were available for all six studies; however, for two studies (Santoro, Le Cesne)[19, 20], the HR was derived from a method involving digitisation of the KM plot, allowing the patient level data to be simulated, thus facilitating the derivation of HRs and associated standard errors (SEs). The method used for this process is described by Guyot et al. (212)[24] Based on additional information provided by the company, the authors confirmed the use of the software package (DigitizeIt) used to digitise the KM plots. However, no information was provided regarding the quality of the published KM plots and the perceived quality of the extracted data. Guyot et al. (2012) claim that their methodology was `excellent' in terms of accuracy and reproducibility with regard to median survival and

probability of survival, but results were less good for hazard ratios, especially where the numbers at risk and numbers of events were not reported.[24] The HRs were not directly used for the MTC analysis, as individual participant data were required for the fractional polynomial model, but by calculating HRs from the reconstructed data, and comparing with reported HRs where available, the validity of the method for reconstructing data can be evaluated.

In the Santoro trial, the KM plot reported OS starting from 100% survival, the numbers of participants at risk at the start of each year in each group, and the total numbers of observed events, are reported. As the company's submission correctly reports, the KM plot for PFS does not start from 100% survival and hence cannot be used to reconstruct patient level data and hence the HR and SE. Le Cesne et al. (2000) reports KM plots for both OS and PFS with the overall numbers at risk by year, and total number of events. However, the KM plot for PFS was not used to derive the reconstructed dataset and HR due to the issue with the Santoro PFS KM plot not being suitable for derivation of the HR as mentioned above. Evidence for lack of proportional hazards was observed for OS in five out of the six included studies that had an available KM plot (four of these five required data reconstruction), and for all four of the studies that had an available KM plot for PFS, three of which required data reconstruction.

Guyot et al. (2012) report that the method to reconstruct the HRs may be less accurate in the presence of non-proportional hazards, and also state that their purpose of reconstructing the KM data is not specifically to generate an approximation of the HR. Therefore, the lack of proportional hazards in the majority of studies that required data reconstruction is a caveat to the validity of such data. Guyot et al. (2012) also reported the loss of ability to reconstruct valid data when number of patients at risk and/or number of observed events are not reported. No HRs were reported for the Santoro and Le Cesne studies; HR were derived from reconstructed data for OS but not for PFS, as discussed above. The Judson paper also provides KM plots for OS and PFS, with numbers at risk at 5-monthly intervals and overall numbers of events. Maurel et al (2009) presents KM plots and HRs for both OS and PFS, but does not provide numbers at risk by arm at specified time points, or the total numbers of events. Therefore, the use of the KM plots to reconstruct individual data for the Maurel study may be of limited validity. The Seddon study is reported only in abstract form, and although the HRs are reported for both OS and PFS , we have been unable to identify a publication that reports the relevant KM plots from which the individual patient data could have been reconstructed. The lead author confirmed the full manuscript would soon be published, however, not be in time for this appraisal.

From additional information provided by the company, no requests were made for individual patient data from the original studies. All data reconstruction was performed using information (such as KM plots and numbers of patients at risk and number of observed events) available in the public domain. However, based on additional information supplied by the company, the HR was reported for four studies for both OS and PFS (Judson, Maurel, Seddon and Tap), and using reconstructed data for these studies (including Tap, where IPD were available), the HRs and confidence interval boundaries were similar for all studies.

The company report states that NMAs were performed for OS, PFS, ORR, and discontinuation due to adverse events; however, only OS and PFS were relevant for the economic modelling. As the NICE Decision Support Unit methodology for NMA of survival analysis is still being developed, the company considered two methods for meta-analysis of survival analysis that have been recently proposed. The first method (Woods et al.), is a Bayesian NMA that allows different outcome metrics such as HRs, median survival and survival counts to be combined. This method assumes proportional hazards. The second method (Jansen) uses reconstructed patient level data with a model including Bayesian fractional polynomials and does not assume proportional hazards

Due to concerns regarding the proportional hazards assumption from the visual inspection of KM plots, the company sought to formally test the proportional hazards assumption using the method of Grambsch and Therneau.[25] Evidence for lack of proportional hazards was found for the Judson et al. study, comparing Dox with IfoDox, for both OS and PFS, and, with borderline significance, for the Tap study, comparing Dox with OlaDox, again for both OS and PFS. Therefore, the company correctly decided to use fractional polynomial methods that do not assume proportional hazards. The company set out the KM plots and log-log plots for OS and PFS for all studies included in the NMA, as well as p-values for the formal test of proportional hazards. We agree that the proportional hazards assumption is not upheld for the two studies mentioned above and therefore the method using fractional polynomials is more appropriate. The company used the DIC to determine whether a first- or second-order polynomial was a better fit to the data, and to select the relevant parameters. Based on the provided DIC results from the Bayesian models, the selected OS model was a second-order model with P1 set to 0 and P2 set to 1; the selected PFS model was a second-order model with both P1 and P2 set to -2. The DIC is an appropriate metric for model selection and, based on the provided data, the appropriate models have been selected.

The company provides the code for statistical models, but not the full input to the models, including data, and initial values; hence, the models cannot be replicated.

The ORR was modelling using a Bayesian ordinal probit model; discontinuation due to adverse events was modelled using Bayesian binomial models. These models followed the NICE DSU Technical Support Document guidelines. No clear definition of the objective response rate (ORR) is provided. The company states that all available data for best response to treatment, as reported in the studies, were used in the NMA. This comprises data on disease progression, stable disease, partial response, and complete response. The company does not report which outcomes are reported for each trial, and the data derived from each trial that is then inputted into the Bayesian models are not reported. This outcome is modelled using an ordinal probit model based on those of the NICE DSU (Technical support document 2); the code for the models are reported in the Appendix, but not the full data inputted into WinBUGS, therefore the models cannot be replicated.

The safety outcome is toxicity leading to discontinuation of treatment, which is modelled using a binomial distribution. Again, the company does not report which trials are included in this model, nor the associated data derived from the studies, and again, the code for the model is reported but not the full data inputted to WinBUGS, so the models cannot be replicated.

4.3.2 Implementation of network meta-analysis in WinBUGS

The technical implementation (number of iterations, burn-in, number of chains) of the models in WinBUGS is described for all models with the exception of the fractional polynomial models (additional data from the company has confirmed the implementation details for these models). Additional data from the company has also confirmed the methods used to check for convergence (namely, iteration plots, Gelman–Rubin diagnostics, and plots of the posterior distributions), and confirmed that no difficulties were observed with regard to model convergence.

Due to the small size of the network, the company's submission reported patient and demographic characteristics for informal comparison, rather than including these covariates in the NMA. There is some variation across studies with regard to covariates, for example, the JGDG trial [16] has the highest median age (58 years), with median ages ranges from 48 to 58. The proportion of male patients ranges from 39% to 58%. There is some systematic missing data (data not recorded for the study as a whole) across the studies. The included studies were published across a timespan ranging from 1995 to 2016; as the company discusses, there may have been changes over time that would influence the results, for example changes in additional care that patients would receive as well as their allocated treatment within the trial. The earliest published trial is Santoro et al. (1995), followed by Le Cesne et al. (2000). As the Le Cesne study connects to the network only via Santoro, a

sensitivity analysis excluding these studies may have been helpful (these studies are included for OS but not for PFS due to lack of useable data). Overall, we agree with the argument that it would be counterproductive to include covariates in the NMA, as the resulting increased model complexity would be unlikely to increase the value of the model results. However, it would be helpful to set out the results (such as HRs) from both the published sources where date were reported, and the data reconstruction for all studies, to allow comparison of the results and facilitate discovery of any general patterns in the outcomes. (This data was supplied by the company as additional information.)

Due to the lack of closed loops within the NMA, no loop inconsistency was possible. Also, as there were no direct comparisons that were represented by more than one trial, statistical heterogeneity within a direct comparison could not occur. Therefore, all NMA models used a fixed effect approach, and non-informative priors were used throughout. Although ideally, both fixed effect and random effects models would be used and compared, with a range of priors, possibly including informative priors based on expert clinical opinion, the decision of the company was to use only fixed effect models with non-informative priors. This approach is appropriate to the nature of the network and dataset, with a small number of studies, spread across a high number of treatment nodes. Informative priors may be difficult to elicit in this clinical situation, and may lead to the possibility of influencing the results of the model inappropriately.

Clinical heterogeneity was explored by setting out study level patient and trial characteristics to assess whether the trials were comparable and whether differences could bias the NMA results. Although the authors mention the use of meta-regression to investigate possible sources of heterogeneity, in actuality a meta-regression approach was not used, due to the small number of studies within the NMA. This is a reasonable decision, and the exploration of clinical heterogeneity by means of descriptively comparing the study level characteristics of patients and trials is an acceptable method to investigate sources of heterogeneity.

4.3.3 Results of the network meta-analysis

The principal summary measure for the fractional polynomials NMA was the mean difference in months for each comparator versus OlaDox. The economic model included only the parameters (scale and two shape parameters) for the survival functions to inform the survivor proportion at each cycle. The median survival times for OlaDox and a selected IfoDox regimen[26] were also reported and compared to those from the relevant trial.

For both OS and PFS the median survival time for OlaDox derived from the NMA was close to that observed in the Tap study. The median survival times derived from the NMA for IfoDox were slightly different from those reported by Judson, which could be due to the fact

that the NMA combined results for four other IfoDox regimens as well as for the Judson regimen, but could also have been due to other issues with the model, such as the validity of reconstructed data for the IfoDox trials, or the fit of the fractional polynomial model.

The results of the NMA for OS and PFS are set out. These results refer to the fixed effect second order fractional polynomial model (Eli Lilly submission, Figure 16, Appendix 8), with values for P1 and P2 as described in Appendix 10 for OS and PFS (this model was confirmed to be correct by additional information supplied by the company). OlaDox was found to have a statistically significant increase in OS in comparison with Dox only and with three out of the four IfoDox regimens; IfoDox (10 g/m^2, 75 mg/m^2), IfoDox (5 g/m^2, 50 mg/m^2), IfoDox (5 g/m^2, 75 mg/m^2). There were no statistically significant differences between OlaDox and IfoDox (12.5 g/m^2, 90 mg/m^2). With regard to PFS, there was an increase in mean survival comparing OlaDox with Dox only, and with one of the two IfoDox regimens with data available for PFS (IfoDox 12.5 g/m^2, 90 mg/m^2), that was weakly statistically significant (p-value 0.04 in both cases). (The mean difference is provided with a range which is assumed to be the 95% credible interval, although this is not stated.)

Based on cumulative rankograms (supplied with additional information), OlaDox had the highest probability of being the `best' treatment with regard to OS in three analyses: all patients, unstratified HR: 82.4%; all patients, stratified HR: 90.0%; first line patients only: 85.3%. With regard to PFS, there was an increase in mean survival comparing OlaDox with Dox only and with one of the two IfoDOx regimens with data available for PFS that was weakly statistically significant (p-value 0.04 in both cases). There was no statistically significant difference in mean survival comparing OlaDox with the second IfoDox regimen. In four out of six comparisons (using PFS assessed in different ways and different patient subgroups), OlaDox had the highest probability of being the `best' treatment, although the highest probabilities were associated with the stratified HR models: with investigator assessment of outcome the probability was 65.0, and with blinded independent review of outcome, the probability was 63.1%. The only other treatment regimen to have the highest probability of being the `best' treatment for any analysis was Dox (75mg/m2) + Ifo (10g/m2).

The results of meta-analyses comparing OlaDox against each other treatment in the decision problem individually in a paired comparison using HRs as the outcome measure are presented in Appendix 11 of the company submission. These results are also from the fixed effect second order fractional polynomials NMA model (Eli Lilly submission, Figure 16, Appendix 8), as stated in the additional information. In the text it is stated that these results are not included in the economic model, whereas in the footnotes of the company submission, Table 17, Appendix 11, it is stated that these values (presumably the HRs) were used in the economic model, so this issue remains unclear (although based on the

description of the economic model it appears that only the survival function parameters were used to inform the economic model). The main analyses included data from patients who received OlaDox as a first line treatment only, and produced HRs for overall survival that were significantly lower for OlaDox compared with Dox monotherapy and for one of the four IfoDox treatments. For two of the IfoDox treatments, there was a borderline significant reduction in the HR for OS (p-values 0.06) and for the fourth IfoDox treatment there was no significant difference in HR. Sensitivity analyses using any line of OlaDox produced comparable results; two sets of results were presented, one for a stratified ITT analysis, the other for an unstratified ITT analysis. It is not stated what stratification was used in the stratified analysis. For the comparison between OlaDox against Dox monotherapy, and three of the four IfoDox treatments, the strength of evidence for improved OS with OlaDox was stronger in the stratified analysis than in the unstratified analysis. Result for equivalent analyses using PFS found no statistically significant improvement in PFS when comparing OlaDox against Dox monotherapy and the two IfoDox regimens with available data that could be used in the NMA model. Only one result produced a borderline significant result, which was the stratified ITT analysis comparing OlaDox against Dox monotherapy. Some further results from the main analysis of first line treatment only are presented, stating that for OS, OlaDox had the highest probability of being the best treatment (85.2%) with an associated SUCRA score of 0.97. However, the equivalent values for the other five treatments in the decision problem are not provided. For PFS, the probability of OlaDox being the best treatment was 43.5%, with one of the IfoDox treatments having a probability of being the best treatment of 52.8%. Again, the values for the other two treatments in the decision problem with available data for PFS are not provided.

The predicted best response rates from the NMA are presented for three categories of response: (i) stable disease, partial response, or complete response; (ii) partial response or complete response; (iii) complete response only. The predicted best response rates are set out for all six interventions included in the decision problem. It is not clear which model was used, but it is assumed that the results relate to the model in Figure 7. The intervention with the predicted best response rate across all three categories was Dox (75mg/m^2) + Ifo (10g/m^2). The predicted best response rates for OlaDox were the third highest of the six interventions. Cumulative rankograms for all treatments were provided as additional information; these indicated that the Dox (75mg/m^2) + Ifo (10g/m^2) intervention had the highest probability (0.77) of being the `best' intervention, whereas OlaDox was the third best out of six treatments within the decision problem, with a probability of 0.069. When looking at ORR in terms of difference in probit score (assumed to be derived from the same model) no

statistically significant differences were observed comparing OlaDox with any of the five alternative treatments in the decision problem.

Discontinuation due to adverse events was analysed using odds ratios, again, the specific model is not stated but it is assumed to be the model in Figure 11. The OR for discontinuation due to adverse events was below 1 for all comparisons of OlaDox versus any of the other five treatments in the decision model, and was statistically significantly lower for three of the five comparisons, with weak evidence for a lower OR in one of the comparisons.

4.3.4 Overall critique

The NMA includes data from six studies published over a timespan of approximately 20 years, with some variation in patient demographic characteristics, inclusion criteria (notably the JGDG trial that evaluated OlaDox included patients who had previously received other forms of chemotherapy, unlike the other included studies), as well as the inevitable changes over time in additional treatments. Hence, there are some caveats in the validity of combining such studies.

The six studies include seven treatments, six of which are included in the decision problem. The network is connected through a common reference intervention, Dox monotherapy, included in all except one of the trials. The final trial compares two different IfoDox regimens, one of which is connected to the network via Dox monotherapy, through a connection with one of the same IfoDox regimens used in a different trial. Each pairwise comparison in the network has only one trial providing data.

With regard to the survival analysis modelling, the company correctly points out the lack of proportional hazards in some of the trials, and thus opts for an approach to survival analysis that does not require the proportional hazards assumption to be valid, specifically the fractional polynomials method.[27] However, this method requires individual patient data, which was only available for the JGDG trial. Hence, the authors used the method described by Guyot [24] to reconstruct individual patient data using digitised KM plots and numbers at risk/numbers of events for the remaining studies. When reporting the HRs and confidence intervals that were reported in the publications (where available) alongside those derived from the reconstructed data, the two sets of values were in general very close. However, no sensitivity analyses, for example using different software to digitise the KM plots, or using different methods to simulate the individual patient data, were reported. Also, the quality of the KM plot images is not discussed, nor the perceived quality of the reconstructed data. We were unable to locate the published KM plot for one of the studies (Seddon)[22] where reconstructed data were required. For one of the studies, the numbers of patients at risk and

numbers of observed events were not reported, which raises concerns about the validity of the reconstructed data, despite similarity of the published and reconstructed HRs. Therefore, the validity of using the reconstructed data to perform the NMAs to inform the economic model is called into question.

The code for the NMA models for the different outcome measures is provided, but not the full model input including the data and initial values, therefore the results presented in the submission cannot be replicated. The use of fixed effect models is appropriate due to the nature of the network. Although non-informative priors were used throughout all models, no sensitivity analyses appear to have been performed, possibly using different non-informative priors. The issue of prior distributions that are intended to be vague having an undue influence on the model may be of less concern for a fixed effect model than for random effects; however, in this example, due to the small number of datasets that are thinly spread across the treatment nodes, it may be the case that the prior distributions may influence the results. Therefore, a sensitivity analysis using alternative non-informative priors would add to the robustness of the results.

4.4 Conclusions of the clinical effectiveness section

The company submission identified a single RCT trial (JGDG) that matched the decision problem. Information from this study was reported in detail. The RCT was a phase 2, open-label trial with a small population, however, it was well conducted.

The primary efficacy endpoint for the RCT was an ITT comparison of progression free survival for patients randomised to OlaDox versus patients randomised to Dox. An improvement was demonstrated which reached the planned statistical significance level of 0.1999. Statistical significance was also reached for overall survival, indicating OlaDox to be superior to Dox monotherapy.

With regard to the network meta-analysis, the main analyses included data from patients who received OlaDox as a first line treatment only, and produced HRs for overall survival that were significantly lower for OlaDox compared with Dox monotherapy and for one of the four IfoDox treatments. For two of the IfoDox treatments, there was a borderline significant reduction in the HR for OS and for the fourth IfoDox treatment there was no significant difference in HR. In general, the ERG consider the NMA to be appropriate given the challenges of not having individual patient data.

5. Cost-effectiveness

5.1 ERG comment on companies review of cost-effectiveness evidence

5.1.1 Searches

The company conducted a systematic literature review (SLR) of the evidence of economic analyses, cost and resource use and utility studies associated with adults with advanced or metastatic STS. Two searches were performed using the same search strategy but different dates; the first one conducted in January 2015, which included published studies from January 1999 to January 2015; and an updated search conducted in September 2016, which included published studies from October 2014 to September 2016. The searches were limited to 2004 onwards and the decision to limit in this way does not appear to be explained in the text.

The following electronic databases were searched: MEDLINE, MEDLINE In-Process, Embase, EconLit, The Cochrane Library, the Health Economics Evaluations Database and Biosciences Information Services (BIOSIS).

In addition to searching the published literature, targeted desktop research was performed to identify relevant HTA documents from NICE, the Scottish Medicines Consortium, the Canadian Agency for Drugs and Technologies in Health, and the International Network of Agencies for Health Technology Assessment.

A hand search of bibliographic lists of SLRs and HTA documents was also undertaken to identify further studies of interest.

Separate database searches were performed for economic evaluations and cost, resource-use and utility studies.

5.1.2 Inclusion/exclusion criteria

Table 21 presents the inclusion and exclusion criteria that were used for screening the studies. The separate database retrievals for economic evaluations and cost, resource-use, and utility studies were screened using all criteria to ensure that all relevant studies were included (e.g., any utility studies that appeared only in the economic evaluations search retrieval would be included). The search included all pharmacological therapies used in the treatment of STS in any line of therapy since the comparators for the OlaDox had not been finalised at the time the SLR was conducted.

Table 21: Inclusion and Exclusion criteria for the economic SLR

Criteria	Included	Excluded
Population	• Adult patients with advanced or metastatic STS not amenable for surgery or radiotherapy[a] • Adult patients with – Liposarcoma – Fibrosarcoma – Histiocytoma – Leiomyosarcoma – Rhabdomyosarcoma – Synovial sarcoma – Angiosarcoma – Malignant peripheral nerve sheath tumours – Ewing sarcomas and primitive neurectodermal tumours – Fibromatosis	• Children • Gastrointestinal stromal tumours • Kaposi's sarcoma
Interventions	• All pharmacological treatments for advanced or metastatic STS in any therapy line	• Chemotherapy followed by haematopoietic stem-cell transplantation
Study type	• Economic evaluations – Cost-effectiveness analyses – Cost-benefit analyses – Cost-utility analyses – Cost-minimisation analyses – Cost analyses • Prospective, retrospective, cross-sectional, or other studies reporting costs or resource utilisation (e.g., observational studies, clinical trials, database analyses, surveys, and Delphi panels) • Utility studies (including studies where utility weights were mapped from other instruments, e.g., disease-specific patient-reported outcome measures) • Systematic reviews of economic analyses or cost or resource-use studies[b]	• Commentaries and letters (publication type) • Consensus reports • Non-systematic reviews • Articles reporting cost estimates that are not based on data (e.g., commentaries making general reference to cost burden)

Key: a, Patients with recurrent disease, distant metastases, or locally advanced STS were included; b, Systematic reviews were included at level 1 screening, used for identification of primary studies, and then excluded at level 2 screening.

5.1.3 Results

5.1.3.1 Description of economic evaluations identified in the SLR

The cost-effectiveness systematic review identified 19 publications that reported cost-effectiveness and/or cost-utility analyses in STS. The only publication pertaining to interventions in this submission was a publication by Guest et al. (2013) comparing the cost-effectiveness of IfoDox to trabectedin.[28] The information from this publication was not utilised in the company's cost-effectiveness analysis because it reported a study which was not conducted from a UK perspective. The ERG, however, believes that this study is relevant to the decision problem since it was conducted in Western European countries, and that the publication contains valuable information which could be used in the company's analysis.

5.1.3.2 Description of identified studies reporting health related quality of life

Since HRQoL data was not reported in JGDG study, Lilly conducted a systematic literature review to identify published health-state utility estimates. The company identified the only prior NICE TA in this area, trabectedin TA185.(2010)[1] Three publications, Reichardt et al. (2012),[2] Amdahl et al. (2014)[3] and Delea et al. (2014),[4] fulfilled the requirements of the NICE reference case since quality of life was measured directly from STS patients using the EQ-5D or EORTC-QLQ-C30 mapped to EQ-5D, with valuation based on a UK population.

Amdahl et al. (2014)[3] and Delea et al. (2014),[4] both pertained to the cost-effectiveness of pazopanib in UK and Canada respectively. Post-progression utilities in these studies were obtained by combining treatment-specific estimates of the mean utility decrement post-progression vs. pre-progression from the PALETTE trial (reflecting utility immediately post-progression) with an estimate of utility in the terminal stage of the disease based on Shingler et al. (2013).[29]

The study by Reichardt et al. (2013)[2] reports health state utility values (HSUVs) by line of chemotherapy and health state, progression-free and progressed (Eli Lilly Submission, Table 55, p173). HSUVs from this study were, therefore, considered most appropriate for inclusion in the model, although the study had some limitations. The values were based on a small sample (the number of assessments in each health state ranged from 12 to 35) and represented a mixed population of patients with STS (n=94) or bone sarcoma (n=20), although the majority of assessments were for STS patients. In addition, the study's requirement that *patients provide a response* may have excluded patients with early disease progression. The company acknowledged that this may have resulted in higher utility values than would be expected for all patients with disease progression.

Eli Lilly did not conduct any separate searches for adverse event literature. The company argued that the incidence of adverse events for the interventions of interest are expected to be best characterised in the clinical trials, which were identified in the systematic review of clinical evidence. Resource use, costs and health utility estimates relevant to mSTS patients experiencing adverse events were identified as part of the systematic review of economic evidence (including economic evaluations, resource use, cost and utility estimates relevant to mSTS).

Utility decrements for grade 1, 2, 3 and 4 adverse events were included in the base-case analysis (Appendix 6). Utility decrements were applied for the duration of the event for each treatment group, as recorded in the JGDG study (dates of commencement and resolution of events recorded on the case report form or the length of hospital stay, whichever was the longer), validated by clinicians at the UK advisory board.

For grade 3 and 4 adverse events not reported in publications from systematic literature review of cost-effectiveness studies proxy values from NSCLC populations were used.

5.1.4 Conclusions

Although no conclusions were provided by the submitted review, there was only one publication pertaining to interventions in this submission. However, although this publication was conducted in Western European countries, it was not included in the company's cost-effectiveness analysis because it was not conducted from a UK perspective.

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

5.2.1 NICE reference case checklist

Table 22: Critical appraisal checklist based on NICE Reference Case[17]

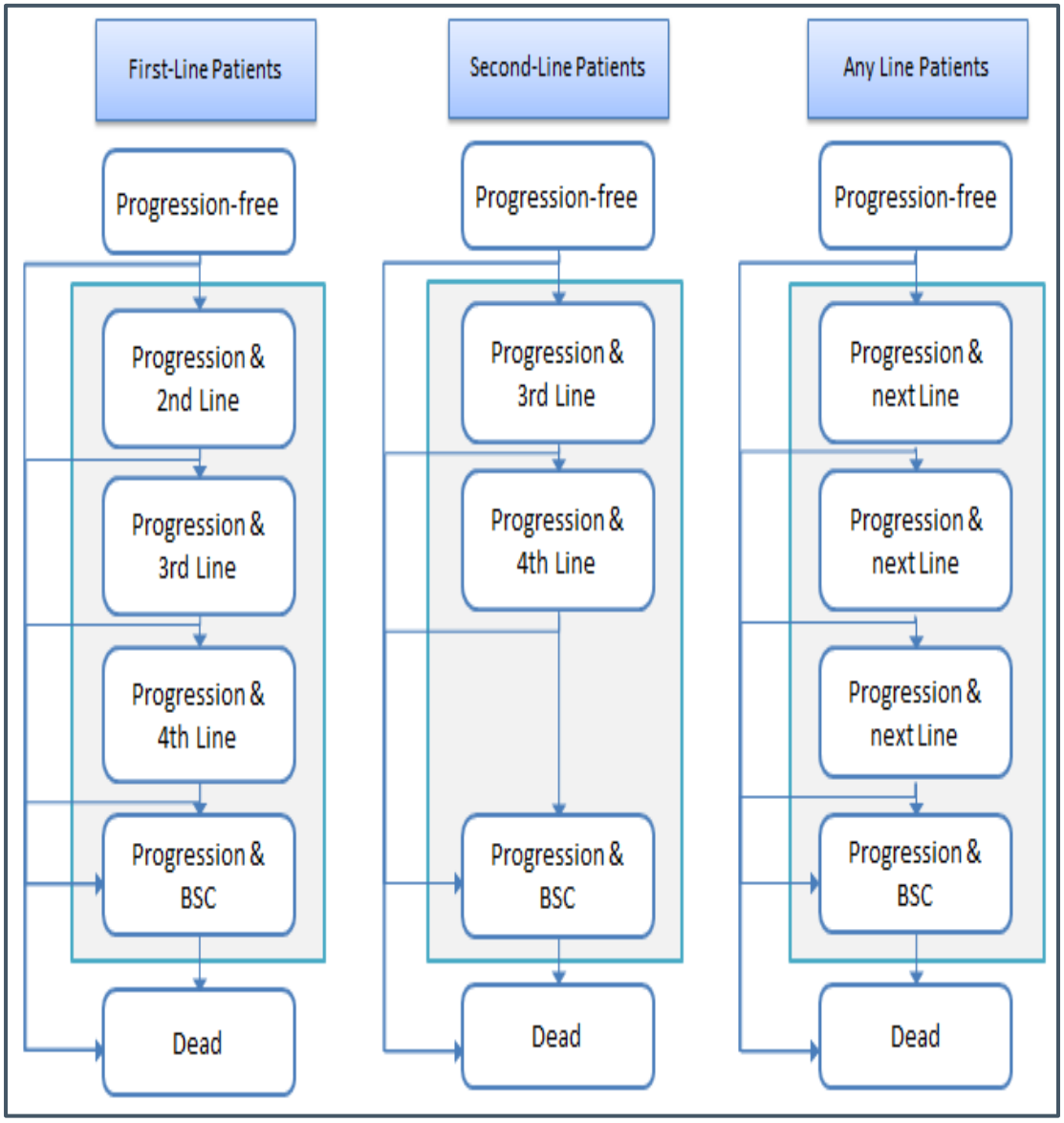
NICE reference case requirement		Critical appraisal	Reviewer comment
Defining the decision problem	The scope developed by NICE	No	The company considered Ola as a 1st-line therapy in their economic analysis, which is not in line with the final scope. This was based on the place of OlaDox in the UK STS treatment pathway, as recommended by clinical experts. The evidence base for OlaDox vs. IfoDox analysis included 6 studies; in 5 of those studies, the patient population were 1st-line patients, while only in one study, JGDG, 65% of patients were 1st-line.
Comparator(s)	As listed in the scope developed by NICE	Yes	Eli Lilly included all comparators listed in the NICE Scope.
Perspective on outcomes	All direct health effects, whether for patients or, when relevant, carers	Yes	
Perspective on costs	NHS and PSS	Yes	
Type of economic evaluation	Cost–utility analysis with fully incremental analysis	No	Fully incremental analysis is not applicable as the patients receiving Dox and IfoDox in clinical practice are somewhat different.
Time horizon	Long enough to reflect all important differences in costs or outcomes between the technologies being compared	Yes	
Synthesis of evidence on health effects	Based on systematic review	Yes	OlaDox vs. Dox comparison was based on a single study, JGDG trial;[16] comparison of OlaDox vs. IfoDox was based on a NMA of 6 clinical studies.
Measuring and valuing health effects	Health effects should be expressed in QALYs. The EQ-5D is the preferred measure of health-related quality of life in adults.	Yes	

Source of data for measurement of HRQL	Reported directly by patients and/or carers	Yes	Data on patients' quality of life was not collected in JGDG study.[16]
Source of preference data for valuation of changes in HRQL	Representative sample of the UK population	Yes	
Equity considerations	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Yes	
Evidence on resource use and costs	Costs should relate to NHS and PSS resources and should be valued using the prices relevant to the NHS and PSS	Yes	
Discounting	The same annual rate for both costs and health effects (currently 3.5%)	Yes	
Equity weighting	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	Yes	

5.2.2 Model structure

The model presented by Eli Lilly is a cohort-based partitioned survival model. It is informed by the JDGD study[16], a systematic review of country-specific resource use, costs and utilities, the company's own observational study of resource cost and use, and multiple oncologists' and external consultants' advice on STS and model implementation. The structure of the model described in the submissions is presented in Figure 10.

Figure 10: Company's model structure



Source: Eli Lilly model submission, "Overview" worksheet.

The company's evidence is submitted with the intention that Olaratumab would be used a *first-line* treatment, hence the first column of Figure 10 is most relevant to the model. There are 3 health states in the model; progression free survival, post-progression survival and death, with those experiencing disease progression having up to 3 further lines of therapy available and best supportive care (BSC).

In the base case analysis, patients enter the model upon commencement of receiving first line treatment, OlaDox or the comparator, Dox/IfoDox. The patients can then remain progression free, during which time they continue their first-line therapy until the completion of treatment, come off it for another reason (toxicity, physician/patient decision) or their disease progresses, whichever is soonest. Whilst Dox has a capped number of

administrations, Ola administrations were taken until disease progression in the baseline study, which the model mirrors. Those who have progressed are then placed in one of the post-progression survival (PPS) lines of treatments. It should be noted that strictly speaking the subsequent lines of treatment are not modelled as independent states within PPS, but this model informed cost-choices of subsequent therapies.

Disease pathways are as follows:

- Patients in PFS can continue in PFS, enter PPS at any one of the subsequent lines of treatment, or they can die.
- Patients in PFS are placed in one of subsequent lines of treatment within PFS. They can then remain in this line, move through the subsequent lines of treatment (not necessarily linearly) or die. Subsequent treatments are dynamic in that they may differ across cohorts; i.e. patients receiving Dox as a first-line would not receive it in further lines of treatment, whereas non-Dox first-liners might.

The proportions of patients in each state at time t is calculated through the use of PFS and OS data and hazard is a function of t. PFS survival in the base-case is directly estimated from JGDG data using a Kaplan-Meier fit, whereas OS survival data are calculated from JGDG data and external data to provide extrapolation beyond the trial timeline. PPS survival is then the difference between PFS and OS.

The cycle length is one week, which does not directly correspond to a treatment cycle. The treatment cycle proposed is a 21-day cycle in which Dox is administered once (day 1) for all arms and Ola is administered on days 1 and 8 for the OlaDox arm. Table 23 summarises the company's structural assumptions.

Table 23: Methodological and structural assumptions for OlaDox vs Dox and IfoDox

Assumption	Justification
Structural assumptions	
Patients in the JGDG study are broadly representative of the STS population expected to receive OlaDox in UK clinical practice in terms of their underlying risk of progression and death, and response to treatment.	As described in Section 4.1.3 (external validity) of this submission, in the context of a small trial, the JGDG study population was generally representative of the broader STS population in the UK. There were similarities between the JGDG study and UK clinical practice in terms of the use of the SOC Dox as the control arm, the Dox dose (75mg/m^2), patient eligibility criteria (PS 0-2) and the

	requirement for cardiac monitoring in patients on Dox. The mean age (range) of patients) in the trial (57.5 years, 22 – 86) was broadly reflective of patients in UK practice and similar to that reported in the GeDDiS study (Median age 55 years) (Seddon et al).[22] The proportion of female patients was higher in JGDG (56%), similar to the GeDDiS trial which also reported a higher proportion of female patients (61%). In the UK there are approximately equal proportions of males and females, however the ratio of males to females by age band varies considerably (Figure 17 Section 4.13) More than 25 different histological subtypes were represented in Study JGDG (Table 13), the most frequent being LMS (36% in OlaDox arm and 40% in Dox arm), undifferentiated pleomorphic sarcoma (15% in OlaDox arm and 21% in Dox arm), and liposarcoma (12% in OlaDox arm and 22% in Dox arm). This is broadly reflective of the breakdown of STS cases by subtype in the UK population in the sense that LMS and liposarcoma are amongst the top three specified cases of STS.[8] However the proportion of LMS patients in the trial (36% in the OlaDox arm) is higher than that observed in the UK (18%).[8] Furthermore, the relatively high proportion of LMS patients in the trial meant that many other subtypes were present in lower numbers than would be expected in clinical practice. Considering that JGDG was a small trial, it would seem unrealistic to expect the trial to more closely reflect the STS breakdown (> 50 subtypes) in the UK population. Finally, there is no difference between the standard of care between LMS and non-LMS in terms of first-line treatment.
In the model patients with metastatic STS receive up to 4 lines of systemic therapy.	The model allows for up to four lines of therapy in total, based on several sources of evidence. A study by Kantar Health (2013) in which 78 physicians

	(who treated a total of 2,839 western European STS patients monthly in March 2013) were interviewed reported that 44.8% of patients receiving first-line therapy went on to receive second-line therapy, 24.4% of the second-line patients received third-line therapy, and 11.6% of third-line patients received fourth-line therapy. In the JGDG study, 9% of patients in the Dox arm and 13.6% in the OlaDox arm received three subsequent regimens after their study treatment. Based on this information and on clinical opinion, the model allows for up to four lines of therapy.
No half-cycle correction is used in the model	As the cycle length (1 week) is short in relation to the overall model time horizon (lifetime of the patient cohort from the initiation of treatment, i.e., 25 years), no half-cycle correction is applied (Nemeth and Vinczicczki, 2013[30]

Source: Eli Lilly submissions, table 70, p. 195.

5.2.3 Population

The model population represents adults with advanced soft tissue sarcomas not amenable to curative treatment with surgery or radiotherapy, who have not been previously treated with doxorubicin, and where patients have had no prior line of systemic treatment (excludes adjuvant/neoadjuvant chemotherapy).

Patients are 58 years old at the start of the time horizon, based on the mean age of UK cancer patients (Cancer Research UK)[8], have a body surface area (BSA) of 1.91m^2 (Health Survey England[31]), and weigh 77.3kgs (GeDDis trial)[22]. 56% of patients are female (JGDG study).[16] Patient characteristics are varied as part of sensitivity and scenario analyses. No subgroups were to specifically be examined, nor were any subgroups found to have statistically different effects with respect to treatment (Eli Lilly submission, pp. 217-218).

5.2.4 Interventions and comparators

The model proposed by Eli Lilly estimates the cost-effectiveness of OlaDox compared to either Dox monotherapy or IfoDox. The details of the interventions and comparators (dosage, route and duration of administration) are listed below in Table 24.

Table 24: Dosage, route and duration of administration for OlaDox and comparators

Regimen	Drug	Planned dosage per treatment cycle	Duration of administration	Route	Source
OlaDox	Ola	15 mg/kg on days 1 and 8	21-day treatment cycles until disease progression	IV infusion (60 minutes)	Tap et al. (2016)[16]
	Dox	75 mg/m² on day 1	Up to eight 21-day treatment cycles or disease progression	IV infusion (15-60 minutes)	Tap et al. (2016)[16]
	Dex[a]	750 mg/m² on day 1	Treatment cycle 5-8	IV infusion within 30 minutes prior to every Dox infusion	Tap et al. (2016)[16]
Dox	Dox	75 mg/m² on day 1	Up to eight 21-day treatment cycles or disease progression	IV infusion (15-60 minutes)	Tap et al. (2016)[16]
	Dex[a]	750 mg/m² on day 1	Treatment cycle 5-8	IV infusion within 30 minutes prior to every Dox infusion	Tap et al. (2016)[16]
IfoDox[b]	Ifo	3 g/m^2 on days 1, 2, and 3	Up to six 21-day treatment cycles or disease progression	CIV infusion (5 days)	Erkisi et al. (1996)[32]
	Dox	60 mg/m^2 on day 1		IV infusion	
	Mesna	0.5 g/m^2 IV day 1 before Ifo 1.5 g/m^2 with Ifo 1 g/m^2 orally 2 hours and 6 hours after Ifo	Each cycle	IV infusion and oral	Judson et al. (2014)[26]
	Filgra-stim[c]	5 µg/kg for 7 days	Each cycle	Subcutaneous	Judson et al. (2014) [26] Maurel et al. (2009)[23]

Key: CIV, continuous intravenous; Dex, dexrazoxane; Dox, doxorubicin; Ifo, ifosfamide; IfoDox, ifosfamide + doxorubicin; IV, intravenous; OlaDox, olaratumab + doxorubicin; STS, soft tissue sarcoma;

Notes: All regimens are discontinued in the event of unacceptable toxicity. Information from the five main European Union countries suggests that dosing for Dox monotherapy is consistently about 75 mg/m^2 every 3 weeks. Dox dosage for combination therapy with Ifo is 60 mg/m^2 for France and Germany (no information is available for the other countries). The number of treatment cycles for Dox-based therapies is determined by limits on the maximum cumulative dose of Dox due to the cardiotoxicity risk derived from anthracyclines.

a, In Study JGDG, Dex (given at a 10:1 ratio with Dox, i.e., 750 mg/m²) could be administered on day 1 of treatment cycles 5 to 8 in both treatment arms at the investigator's discretion to reduce potential Dox-related cardiotoxicity. In the model, the user can define the proportion of patients receiving Dex; the base case value is based on data from Study JGDG.

b, The IfoDox dosing schedule varies by country. Based on clinical opinion, in the UK the usual schedule is Dox 60 mg/m^2 on day 1 and Ifo 9 g/m^2 spread over 3 days (UK Advisory Board Meeting; 12 April 2016). A similar regimen was investigated by Erkisi et al. (1996): Ifo 1.8 g/m2 per day on days 1-5 (to equal 9 g/m2 total) and Dox 60 mg/m^2. However, this does not connect to OlaDox in the evidence network via a common comparator. Clinicians at an advisory board conducted in the UK indicated that the efficacy of the different IfoDox schedules included in the

evidence network could not be combined or considered equivalent. The advisors agreed that for the UK the efficacy of IfoDox should be based on the trial reported by Judson et al. (2014) and that the costs in the economic model should be based on UK practice (Dox 60 mg/m^2 and Ifo 9 g/m^2).
c, In the Judson trial, pegfilgrastim dosed 6 mg subcutaneously, day 5. Pegfilgrastim is rarely used in the UK; therefore, it has been replaced with filgrastim in the economic model for costing purposes.

Source: Eli Lilly submission, table 58, p.180.

As the company notes, dexrazoxane (Dex) is not expected to be used in UK practice, but the base case includes its usage from the JGDG trial. It is important to note that Ola monotherapy would be carried out until disease progression (or patient decision to discontinue/unacceptable toxicity) if up to 8 treatment cycles of Dox were completed in the OlaDox arm. In the JGDG trial, some Dox monotherapy patients took Ola monotherapy after completing the Dox treatment plan, and after disease progression. The company's base case, however, does not incorporate either comparator arm taking Ola monotherapy. The company's assumptions and reasoning behind this is discussed in section 5.3.5.1.3, p 134.

5.2.5 Perspective, time horizon and discounting

In the model, the perspective on costs was the NHS and personal social services perspectives, and the perspective on health effects was the direct health effects on patients, in accordance with the NICE reference case.

The baseline model time horizon was 25 years, which is justified as a lifetime horizon based on the OS data from the JGDG study extrapolated to beyond the censoring. The lifetime horizon is varied in sensitivity analyses.

Costs and utilities are discounted at a rate of 3.5 per cent, in line with the NICE reference case.

5.2.6 Treatment effectiveness and extrapolation

Treatment effectiveness was estimated from the JGDG trial and from post-hoc analyses conducted on the data collected.

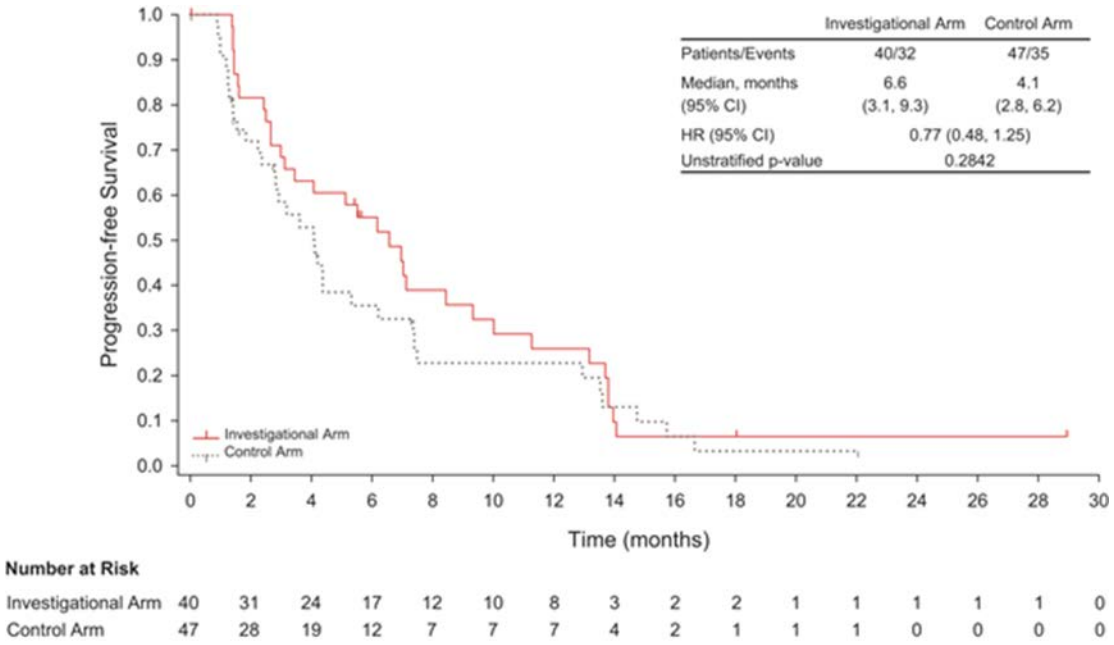
The economic model used the following clinical endpoints:

- Overall survival (OS), the time from entering the model until death from any cause.
- Progression free survival (PFS), the time from entering the model until disease progression
- Post progression survival (PPS), the time from disease progression until death.

PFS was modelled using a Kaplan-Meier approach. The justification was that parametric fittings to PFS data achieved neither good fits, nor replicated the convergence of the treatments arms. Given that the KM curves come close to 0 within the trial duration, the

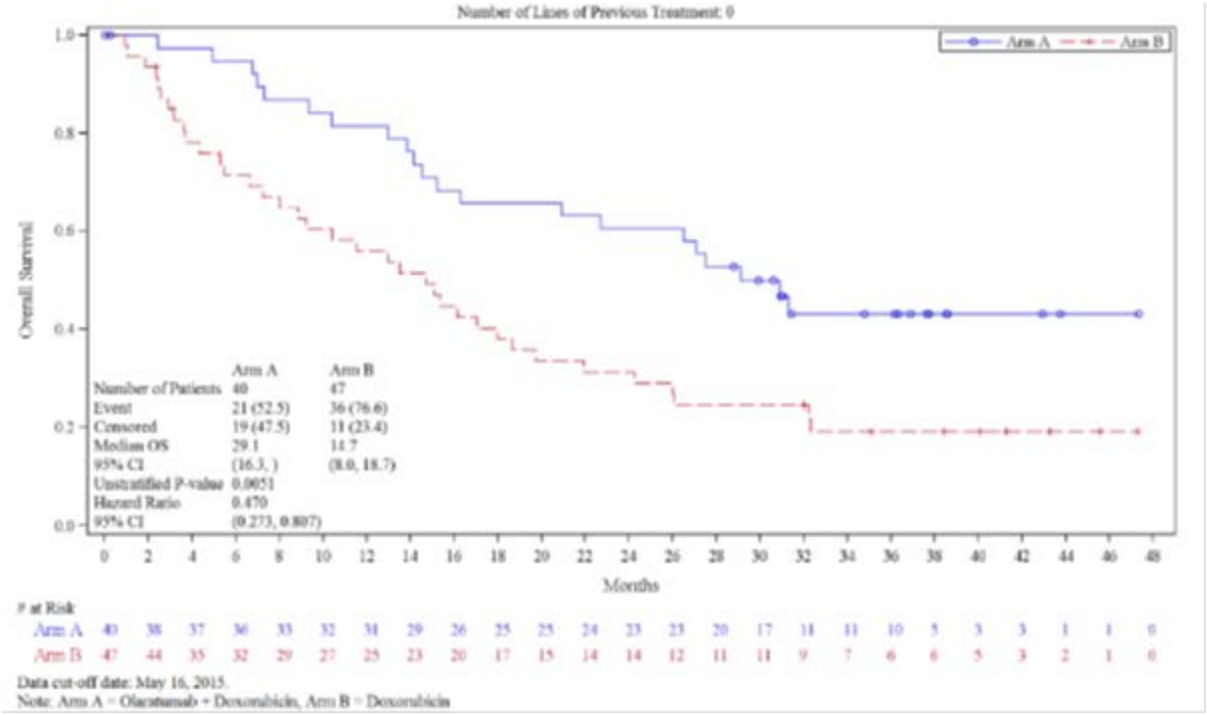
company also judged that extrapolation beyond the trial was not necessary. Investigating parametric approaches was explored in the sensitivity analysis.

Figure 11: PFS Kaplan Meier curve for JGDG (ITT, investigator assessment). 1st-line



Source: Eli Lilly submission, Figure 21 p. 138.

Figure 12: OS Kaplan Meier plot for first-line JGDG population



Source: Eli Lilly submission, figure 25, p. 144

The JGDG trial exhibited significant censoring with respect to overall survival (see Figure 12).[16] Several parametric models were fitted to overall survival data:

- Exponential;
- Log-logistic;
- Weibull;
- Gamma;
- Log-normal;
- Gompertz.

Log-cumulative hazard plots were approximately parallel, and the company argued that this justified assuming proportional hazards for OS. Given proportional hazards, the company modelled OS from data from the JGDG trial with treatment as a covariate (their "arms together" approach). A stepwise process was essentially used to determine other covariates included in the fittings. Survival models were then fitted to JGDG data in accordance with NICE DSU guidance (Source: Eli Lilly submission, p.147).

Table 25: Covariates explored for inclusion in the OS parametric models

Treatment
Age (continuous or categorical split at 65, 70, or 75)
Sex
Tumour type (leiomyosarcoma vs. other)
Line of treatment (first vs. subsequent)
ECOG PS (0 vs. 1 or 2)
PDGFRα expression (positive vs. negative)
Interaction (treatment × tumour type)
Interaction (treatment × line of therapy)

Note: Randomisation in Study JGDG was stratified by PDGFRα expression (positive vs. negative), number of previous lines of treatment (0 vs. > 0 lines of prior systemic therapy), histological tumour type (leiomyosarcoma vs. synovial sarcoma vs. other subtype), and ECOG PS (0-1 vs. 2). Treatment and treatment interaction terms were included in functions fitted to both treatment arms together only. Bolded covariates were kept in the model.
Source: Eli Lilly submission, table 44, p.143

The choice of KM curves over parametric functions was based on the following considerations:

- For both, ITT and the 1st-line subgroup, Log-cumulative hazard plots of PFS for the treatment and comparator arms converged (Eli Lilly submission, pp. 138-139, Fig. 23 and Fig. 24), suggesting that the proportional hazards (PH) assumption is not valid.

In the test of proportionality of hazards between treatment arms, the p-value was 0.13. So, the “arms together” approach could not be used.

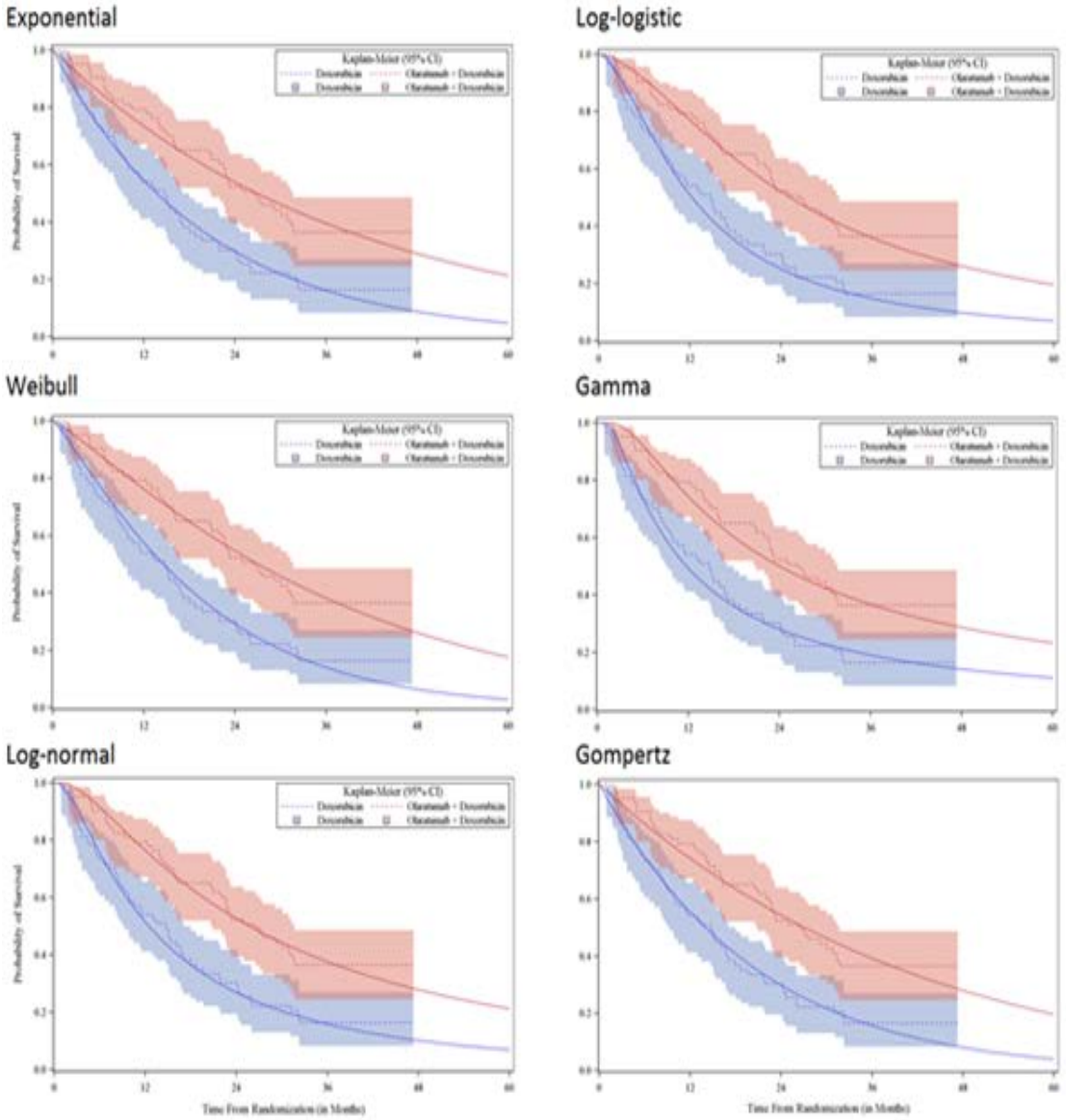
- Parametric survival models with covariates were fitted to PFS data from the individual treatment arms of the JGDG trial; the covariates were the line of treatment, histology LMS, PDGFRα expression, ECOG PS 1 as well as age and sex (Eli Lilly submission, Appendix 25). However, they did not replicate the convergence of the PFS KM curves and did not achieve a good fit to the PFS data.
- Since the KM curves for PFS have essentially reached zero (Figure 5 and Figure 6, p 61 and 61), extrapolation was not necessary.

In the base-case analysis, the company utilised KM curves for the investigator assessment of PFS. The impact of KM estimates obtained in the blinded independent review and the effect of using individual arm parametric functions on the cost-effectiveness results were explored in sensitivity analyses (Eli Lilly submission, Appendices 24 and 25).

The company selected the gamma distribution for their base case analysis extrapolation of OS. This was selected based on the respective Akaike information criteria (AIC), Bayesian information criteria (BIC) and visual fit to existing KM data and clinical plausibility, i.e. the fit to existing external longer term data for a similar patient population (primarily Van Glabbeke et al. (1999)[5]). Figure 13 shows their models. The company then dropped the exponential and log-logistic models. This is because the log-logistic function had similar AIC/BIC to log-normal, but a worse visual fit to the K-M, and the exponential was dropped as it is nested within the Weibull and AIC/BIC were similar.

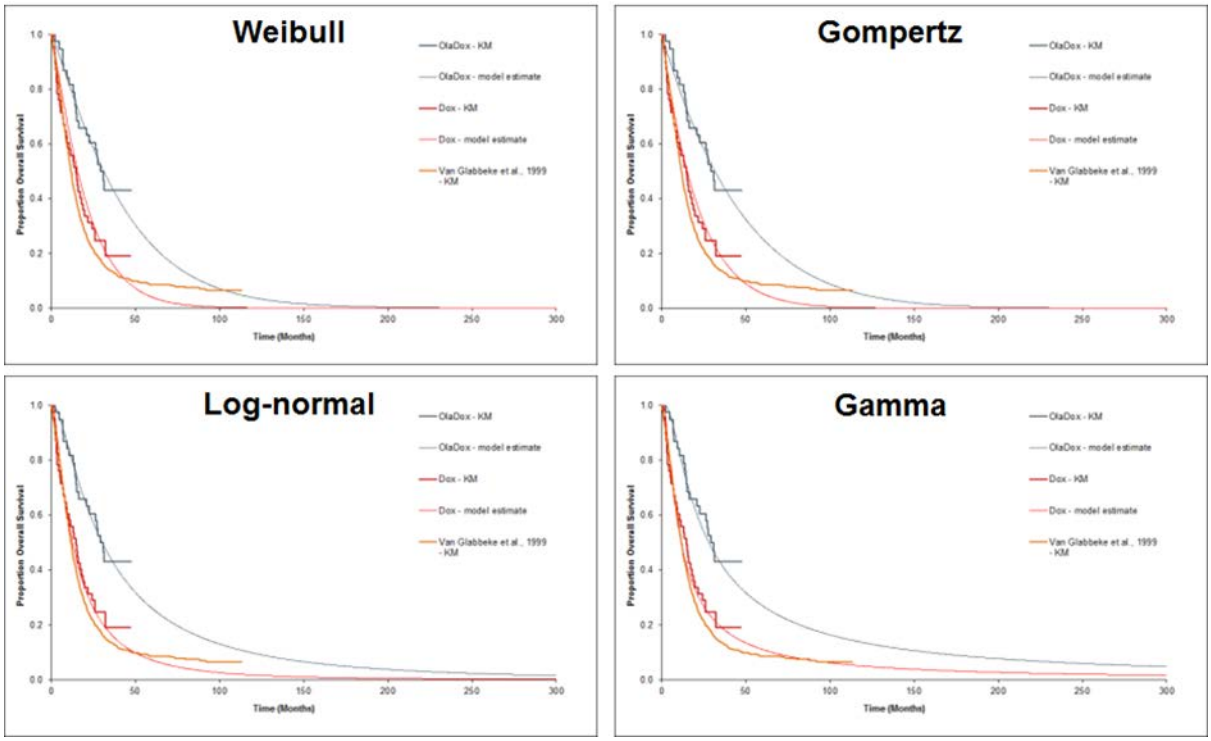
The company also assumed a hazard ratio of one for the Dox and OlaDox arms after 32 months, citing the uncertainty of the long plateaus of the JGDG K-M curves making the treatment effect past 32 months hard to ascertain. 32 months was chosen over 47 (the end of the follow-up period) because it was the last mortality event observed in the OlaDox arm. The impact of 32 months vs 47 months is investigated in the sensitivity analysis. To account for higher mortality rates among the STS population as opposed to the general population, the company applies a hazard ratio of 5.19 at the end of the Van Glabbeke follow-up period to their own survival extrapolation.[5] The Van Glabbeke study stopped follow up after 9 years, so the company applies the HR that they report to general GP mortality rates at this point to extrapolate OS past the JGDG and Van Glabbeke follow-up periods for both arms. They vary this assumption in their scenario analyses; the effect on the ICERs is small for OlaDox, but substantial for IfoDox.

Figure 13: OS ITT Kaplan-Meier data and arms together functions



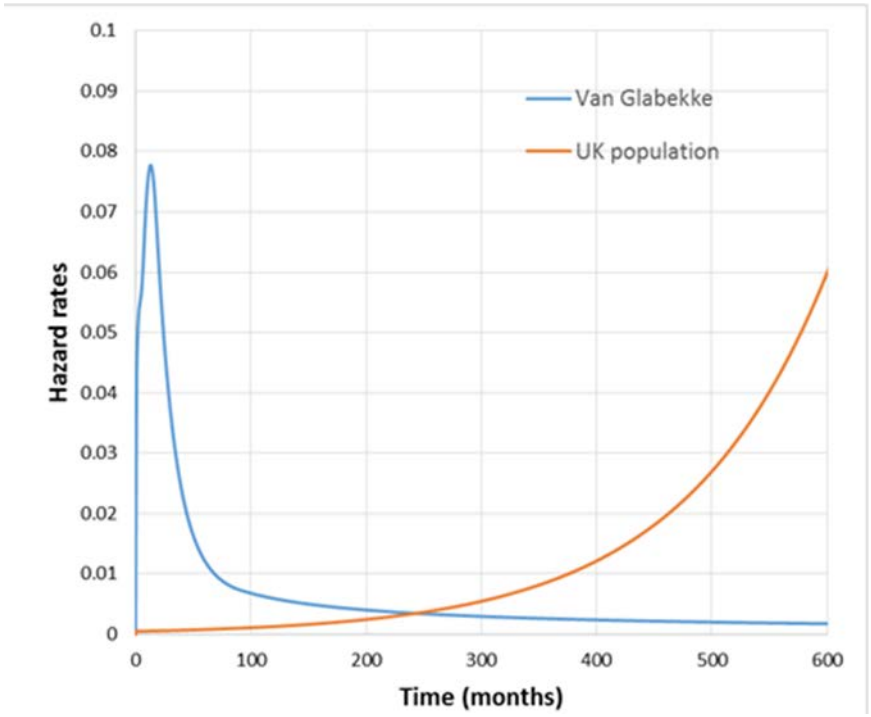
Source: Eli Lilly submission, figure 30, p. 146.

Figure 14: Comparison of first-line OS parametric survival models "arms together" functions with Van Glabbeke study results



Source: Eli Lilly submission, figure 32, p. 149.

Figure 15: Predicted hazard rates for the Van Glabbeke population compared with the United Kingdom general population



Note: this figure represents the hazard function from Van Glabbeke et al. (), extrapolated beyond the study observational period of 114 months.
Source: Lilly's submission, Fig. 34, p.151

The JGDG trial allowed both Dox and OlaDox patients to undergo Ola monotherapy, with some Dox patients taking Ola after disease progression. The company undertook a number of treatment switching style analyses to see if this biased survival estimates, and conclude there is no evidence for this (Source: Eli Lilly submission, p. 157).

The company used a network meta-analysis and Bayesian fractional polynomials to construct an IfoDox arm to compare with OlaDox, as this was not part of the JGDG trial.

The NMA included interventions relevant to the UK that were OlaDox, Dox and 4 different IfoDox regimes. Five endpoints of interests were considered for the meta-analysis: PFS, OS and ORR, QoL, and discontinuation due to AEs or toxicity. Only OS and PFS are relevant for the economic model. A total of 6 studies were included in the final NMA, which are shown in Figure 9, p 77. In order to obtain OS and PFS data, the company digitised KM curves and used these data to reconstruct patient-level data. Fractional polynomial models of OS were used to extrapolate up to 47 months (longest follow-up period amongst the studies) to compare mean survival. The results from this NMA were then used in the Bayesian fractional polynomial model to construct the indirect comparison arm, OlaDox vs. IfoDox. See Section 4.3, p 73 for further details.

Treatment effect

The company stated (p. 150):

> *Recognising the uncertainty associated with the long plateaus at the end of the JGDG KM curves, the base-case analysis assumes no treatment effect (ie applies a HR of 1 to the Dox OS curve) from 32 months onwards. The time of 32 months was selected rather than 47 months (end of JGDG trial follow up) from which to assume no treatment effect because this is the time at which the last mortality event was observed in the OlaDox arm. This is a conservative approach because although there were no mortality events after 32 months, several patients were not censored until later, with 10 and 3 first-line OlaDox patients still at risk at 36 and 42 months, respectively.*

The effect of alternative assumptions was explored in two sensitivity analyses, one assuming the same HR after 32 months as in the trial, and another one where HR was tapered over the period of 12 months after the end of observational period; an additional sensitivity analysis was performed which included KM data up to 47 months.

5.2.7 Health related quality of life

Health effects were measured in quality adjusted life-years (QALYs) in accordance with the NICE reference case. As the JGDG trial did not collect any health-related quality of life (HRQoL) data, the company conducted a systematic literature review to identify published

health-state utility estimates. Of the 13 studies uncovered in the SLR (see section 5.1.3.2 (p. 84).for a detailed critique of the company's SLR), 3 studies were judged to provide consistent utility estimates. Reichardt (2012)[33] was judged to provide the best estimates for the base case. Table 26 shows the estimated HSUVs – note that the base case assumes utilities are independent of response or disease stability, which are examined as part of the sensitivity analyses. All estimates were either directly measured using the EQ-5D or by mapping the EORTC-QLQ-C30 to the EQ-5D. The base case HSUVs from Reichardt were directly reported from metastatic STS patients using the EQ-5D. The patient population differs somewhat in age from the modelled population, with a mean age of 49.5 (at diagnosis) versus 58 years old in the model. Eighty-three per cent of the population had mSTS, the remaining had metastatic bone sarcomas. The company notes some potential for bias from the Reichardt study because they excluded patients who progressed before the response assessment (Eli Lilly submission, p 170, Table 54).

Table 26: Health State utility estimates used in the economic model in the base case and sensitivity analysis

Parameter	Mean (n, SD)	Source	Justification
Utility weights for health states: First-line, base case analysis			
Progression-free	0.720 (17, 0.31)	Reichardt et al. (2012) [2]	The publication meets the requirements of the NICE reference case (see accompanying text)
Progressed	0.56 (28, 0.27)	Reichardt et al. (2012)[2]	The publication meets the requirements of the NICE reference case (see accompanying text)
First-line, sensitivity analysis			
Progression-free with response	0.792 (100, 0.169)	Shingler et al. (2013) [29]	Not reported in any of the studies meeting the requirements of the NICE reference case (Reichardt, Amdahl or Delea). These were consistent with the other values in the model. Used as SA to assess impact of response on QoL.
Progression-free with stable disease	0.72 (17, 0.31)	Reichardt et al. (2012) [2]	Meets requirements of the NICE reference case (see accompanying text)

Source: Eli Lilly submission, p.173

The model accounts for utility decrements arising from adverse events (AEs). The disutility values were obtained from a range of existing studies, with AEs missing from these publications being proxied using NSCLC populations (Source: Eli Lilly submission, Tables 56-57, pp.174-175). For AEs of grade 1 or 2, the company used an "ad hoc literature search" (Source: Eli Lilly submission, p. 175) to identify utility decrements.

The total utility decrements were applied in the following way (Source: Eli Lilly submission, p.174):

Utility decrements were applied for the duration of the event for each treatment group, as recorded in the JGDG study (dates of commencement and resolution of events recorded on the case report form or the length of hospital stay, whichever was the longer), validated by clinicians at the UK advisory board.

5.2.8 Resources and costs

5.2.8.1 Lilly Observational study and systematic literature review

The company conducted an SLR to identify resource usage and costs specific to the UK. 2 studies were judged to provide estimates that could be used the cost-effectiveness model. The company considered, however, that limited data were uncovered and therefore conducted their own "observational study" (Source: Eli Lilly submission pp.176-177).

Twenty one UK-based oncologists completed web-based medical record abstractions, from which patient characteristics and resource use and costs were derived.

5.2.8.2 Drug acquisition

Drug utilisation was based off the JGDG study, and varied according to UK clinical practice in sensitivity analyses.

The total acquisition cost of each drug treatment was directly taken from the JGDG trial; as all patients had discontinued their randomised treatment before the study's cut-off the company decided it was unnecessary to extrapolate beyond the follow up. The treatment costs are then based on (Source: Eli Lilly submission, p.178):

The mean dose per administration and the mean number of administrations for each drug based on data from the JDGD study.

Costs were then directly calculated using UK patient characteristics, a mean weight of 77.3kg (GeDDiS trial)[22] and a BSA of 1.91m^2 (Health Survey for England 2013).[31] An average dosage based on either mg/kg or mg/m^2 was then worked out for the respective drugs to determine total costs.

- For Ola, this was done by assuming a log-normal distribution around the mean weight of patients, and then using the JGDG mean dosage to create a weighted mean per administration cost based on the population proportion that would receive the relevant combination of Ola vials (190mg, 500mg). Complete vial wastage and the availability of both vial sizes assumed in the base case.

- For Dox, this was done by assuming a log-normal distribution around the mean BSA, and then creating a weighted per administration cost of Dox based on the proportion of patients that would receive the relevant combination of vial sizes (10mg, 50mg, 200mg). Compete wastage is also assumed in the base case.
- For Dex, this was done by assuming a log-normal distribution around the mean BSA, and then creating a weighted per administration cost of Dex based on the dose and number of 500mg vials required. Complete wastage is assumed in the base case.

The acquisition costs of IfoDox were calculated by multiplying the planned dosage by the number of administrations per cycle by the number of cycles reported in Judson et al. 2014[26]. Table 27, Table 28 and Table 29 show the dosages, mean number of administrations and costs for the base case analysis.

Table 27: Drug acquisition costs for OlaDox, Dox and IfoDox

Drug			**Lilly's base case**		**PenTAG's base case**
	Brand / formulation	Pack size (mg)	Pack price	Reference	Pack price
Ola	Lartruvo	190-mg vial	[redacted]	[redacted]	[redacted]
Ola	Lartruvo	500-mg vial	£1,000	Lilly (list price)	£1,000
Dox (2mg/ml)	Generic	10mg vial	£1.65	eMIT (12 year period to end June 2015)	£1.53 (same source)
Dox (2mg/ml)	Generic	50mg vial	£4.16		£4.04
Dox (2mg/ml)	Generic	200mg vial	£16.89		£20.30
Dex	Cardioxane	500mg	£156.57	BNF 70 (Sept 2015 - March 2016)	£156.57[1]
Ifo, (1mg/ml)	Generic	1000mg	£91.32		£66.08[1]
Ifo, (1mg/ml)	Generic	2000mg	£179.88		£130.04[1]
Mesna (1g/10ml)		1000mg vial	£29.41		£9.77[1]
Mesna (400mg/4ml)		400mg vial	£13.41		£3.95[1]
G-CSF (0.12mg/0.2ml)	Filgrastim Nivestim	0.12mg	£36		£36[1]
G-CSF (0.3mg/0.5 ml)	Filgrastim Zarzio	0.3mg	£58		£50.15[1] Zarzio®(Sandoz)
G-CSF (0.48mg /0.5 ml)	Filgrastim Zarzio	0.48	£93		£79.90[1] Zarzio®(Sandoz)

Notes: [1] BNF (February 2017: https://www.evidence.nhs.uk/formulary/bnf/current/8-malignant-disease-and-immunosuppression/81-cytotoxic-drugs/treatment-for-cytotoxic-induced-side-effects/urothelial-toxicity/mesna/mesna)

Table 28: Mean dose for OlaDox, Dox and IfoDox for base case analysis (first-line)

Analysis	Mean Dose	Measurement of Uncertainty (Distribution)	Reference
First-line analysis OlaDox vs Dox			
OlaDox			
Ola (mg/kg)	[redacted]	SE = 0.034 (normal)	Lilly data on file 4[34]
Dox (mg/m^2)	74.5	SE = 0.491 (normal)	Lilly data on file 4[34]
Dex (mg/m^2)	730.6	SE = 6.860 (normal)	Lilly data on file 5 2016[34]
Dox			
Dox (mg/m^2)	74.5	SE = 0.348 (normal)	Lilly data on file 4, 2016[34]
Dex (mg/m^2)	724.4	SE = 8.826 (normal)	Lilly data on file 5, 2016[34]
First-line analysis OlaDox vs IfoDox			
Ifo (mg/m^2)	3000	SE = 300 [a] (normal)	Clinical opinion
Dox (mg/m^2)	60	SE = 6 [a] (normal)	Clinical opinion
Mesna (mg/m^2)	4000	SE = 400 [a] (normal)	Judson et al 2014
Filgrastim (mg/Kg)	0.005	SE = 0.00[a] (normal)	Maurel et al 2009

Key: Dex, dexrazoxane; Dox, doxorubicin; IfoDox, ifosfamide + doxorubicin; Ola, olaratumab; OlaDox, olaratumab + doxorubicin; SE, standard error.
Notes: a, Assumed to be 10% of reference value

Table 29: Mean number of administrations for OlaDox, Dox and IfoDox for base case analysis (first-line line)

Analysis	Mean Number of Administrations	Measurement of Uncertainty (Distribution)	Reference
First-line therapy analysis			
OlaDox			
Ola	[redacted]	SE = 3.025 (normal)	Lilly data on file 6, 2016[34]
Dox	5.6	SE = 0.410 (normal)	Lilly data on file 6, 2016[34]
Dex	3.5	SE = 0.208 (normal)	Lilly data on file 7, 2016[34]
Dox			
Dox	4.2	SE = 0.382 (normal)	Lilly data on file 6, 2016[34]
Dex	3.2	SE = 0.271 (normal)	Lilly data on file 6, 2016[34]
IfoDox[a]			
Ifo	13.3	SE = 1.326[a] (normal)	Judson et al 2012[21]
Dox	4.4	SE = 0.442[a] (normal)	Judson et al 2012[21]
Mesna	13.3	SE = 1.326[b] (normal)	Judson et al 2012[21]
Filgrastim	30.9		Maurel et al 2009[23]

Notes: a, Mean number of administrations based on planned number of administrations per cycle times the mean number of IfoDox cycles estimated from Judson et al 2012[21](3 x 4.419 for IfoDox and 3 x 4.096 for Dox).
Source: Eli Lilly submission, tables 58-61, pp. 180-183.

5.2.8.3 Drug administration

Administration costs were determined based on advice from the UK STS advisory board (Source: Eli Lilly submission, p. 179). The HRG costs used by the company are shown below in Table 30. Ola administrations are to be taken with a H1 antagonist as pre-medication. The manufacturer judged that this additional cost would be directly incorporated into the delivery HRG, with no separate cost in the model (Source: Eli Lilly submission, p. 179). We note some errors with the reported HRG codes which we discuss in section 5.3.7, p145.

Table 30: Administration costs for OlaDox and comparators

Regimen	Cost per administration	Delivery codes
Ola + Dox, day 1	£329.32	SB13Z (Daycase); NHSRC [35])
Ola + Dox + Dex, day 1	£329.32	SB13Z (Daycase); NHSRC [35])
Ola, day 1	£185.53	SB12Z (outpatient); NHSRC[35])
Ola, day 8	£204.47	SB15Z (outpatient); NHSRC[35])
Dox, day 1	£185.53	SB12Z (outpatient); NHSRC[35])
Dox + Dex, Day 1	£185.53	SB12Z (outpatient); NHSRC[35])
IfoDox, per cycle	£1,781.86[a]	DZ17V (EI); NHSRC[35])

Notes: a, Calculated as National average Unit cost of £1518.3 divided by average length of stay 2.56 days = £594 /day x 3 days.
Source: Eli Lilly submission, table 62 p.181

5.2.8.4 Subsequent lines of therapy costs

Patients whose disease progresses may receive the subsequent lines of therapy. The company costed these subsequent lines of treatment based on the proportion of patients who received them according to their observational study and the types of therapy received (Source: Eli Lilly submission, p. 130). The cost of the therapy is then combined with each therapy's incidence to calculate total costs. In the model's base case, a total cost of drugs is worked out for subsequent lines of therapy from the Eli Lilly observational study and then divided by mean OS in the base case to get a per cycle cost.

The company assumes as a base case that both OlaDox and the comparators incur the same subsequent therapy costs, calculated from the observational study. This assumption is varied in the scenario analyses.

5.2.8.5 Monitoring costs

Cardiac monitoring frequency and timing are based on the ESMO guidelines. The costs are taken from the NHS reference costs, and a weighted average cost is then calculated taking data from the Lilly observational study to identify the proportional of patients undergoing ECHO or MUGA scans.

Table 31: Frequency and costs of cardiac monitoring tests

Assessment	Cost	Uncertainty (Distribution)	Source
Cardiac monitoring			
Percentage of patients receiving cardiac monitoring	100%	SE = 10%[a] (normal, truncated at 1)	Assumption
Cost per test for (£)			
MUGA	£192.12	SE[b] (normal)	Weighted average RN22Z; NHSRC [35])
Echocardiography	£152.80	SE = 0.070[c] (normal)	EY51Z; NHSRC [35])

Key: a, Assumed to be 10% of the mean value; b, The (SE) is calculated for each service code used for the weighted average (NMDA, NMOP, NMOTH); c, The SE was derived from the interquartile range.
Source: Eli Lilly submission, Table 63, p.185

Follow up visits and imaging costs were calculated in accordance with UK specific clinical practice advice and the company's own observational data (Source: Eli Lilly submission, p. 185):

As per clinical opinion, patients in the UK are likely to be followed-up every three months for the first five years, with a physical examination, CT scan / PET / MRI conducted at each visit. In years six and seven, visits occur every six months and annually thereafter. It is assumed that all patients incur outpatient visit and physical examination costs at each visit. However, only a certain proportion of patients undergo CT scan / PET / MRI at each visit.

NHS reference costs are used for unit costs. Table 32 shows the estimates of these costs.

Table 32: Unit costs and resource use for regular follow-up visits and imaging

Variable	Value	Measurement of uncertainty (Distribution)	Reference
Frequency of follow-up visits (number of months between each visit)			
0-5 years	3	SE = 0.3[a] (normal)	Assumption based on clinical opinion, 2015)
5-7 years	6	SE = 0.6[a] (normal)	
After 7 years	12	SE = 1.2[a] (normal)	
Unit Costs (£)			
Outpatient visit and physical examination	£146.72	SE[b] (normal)	Weighted average WF01A; NHSRC [35])
Computerised tomography scan	£120.92	SE[c] (normal)	Weighted average RD24Z; NHSRC [35])
Positron emission tomography	£517.00	SE[d] (normal)	Weighted average RN07A; NHSRC [35])
Magnetic resonance imaging	£124.53	SE[b] (normal)	Weighted average RD26Z; NHSRC [35])
Resource use for each regular follow-up visit			
Outpatient visit and physical examination	100%	Fixed	Assumption
Computerised tomography scan	92%	n/N = 183/199 (beta)	UK observational study, Lilly data on file, 2016[34]
Positron emission tomography	9%	n/N = 18/199 (beta)	UK observational study, Lilly data on file, 2016[34]
Magnetic resonance imaging	14%	n/N = 27/199 (beta)	UK observational study, Lilly data on file , 2016[34]

Key: a Assumed to be 10% of the mean value; b, The uncertainty (SE) is calculated for each service code used for the weighted average (370 and 800); c The uncertainty (SE) is calculated for each service code used for the weighted average (IMAGDA, IMAGOP, IMAGOTH); d, The (SE) is calculated for each service code used for the weighted average (NMDA, NMOP, NMOTH)

Source: Eli Lilly submission, table 64 pp. 185-186.

5.2.8.6 Adverse event costs

Adverse event costs are calculated by combining the proportion of events likely to require hospitalisation based on data from the JGDG trial with estimates of costs per event. These costs include outpatient visit costs.

In the base case the costs of hospitalisation are based on NHS reference costs, which reflect the length of stay specific to UK practice (but not specific to STS patients). The

manufacturer does not use length of stays recorded in the US JGDG trial, as they argue this does not reflect UK practice. Grade 1/2 AEs were not included in base case. The AEs were independent of the Markov modelling (not cycle dependent) and occur in the first year of the model. The full list of AEs and their utility decrements/costs is presented in Appendix 6.

5.2.8.7 Health-state unit costs and resource use

Costs that are not associated with systematic therapies are shown in Table 33. The two non-death health states are costed by applying an average weekly health state cost to cycles of the model. As the JGDG trial took place in the US, the company's observational study was used to identify these costs for the UK in the base case.

Table 33: Health state costs in the economic model

Health state costs	Value	Measurement of Uncertainty (Distribution)	Reference
Source: Lilly observational study first-line analysis			
Progression-free	£131	N/A	Lilly data on file 10, 2016[34]
Progressed (excluding radiotherapy/surgery costs)	£35	N/A	Lilly data on file 10, 2016[34]

Source: Eli Lilly submission, table 65, p. 186

5.2.8.8 Post-progression treatment costs

The company argues that the PFS costs for OlaDox patients compared to standard treatments are uncertain. The base case assumes that total PPS costs are independent of survival, despite Ola significantly extending OS. Their logic is that better outcomes for OlaDox may reduce treatment requirements, or alternatively prolonged survival may result in more lines of subsequent treatments. This is allowed to vary in SA.

Table 34: Post-progression treatment costs in the model

Total Cost of Active Therapy After First Progression	Value	Measurement of Uncertainty (Distribution)	Reference
First-line of therapy analysis (OlaDox, Dox, IfoDox)(base case analysis)			
Total drug costs (£)	£6,082	N/A[a]	Lilly Observational study
Total administration costs (£)	£1,615	N/A[a]	Lilly data on file 11 (2016[34])
Total AE costs (£)	£278	N/A[a]	

Key: a, The uncertainty is included in the total cost for all patients in the observational study. This cost is then multiplied by the probability of receiving subsequent active therapy according to the observational study.

Source: Eli Lilly submission, table 68, p. 190.

5.2.9 Base-case cost-effectiveness results

The results from the base case analysis with the two comparator treatments are shown below in Table 35 and Table 36. OlaDox compared to Dox monotherapy is estimated to cost

[redacted] per QALY (ICER), and OlaDox compared to the IfoDox combination is estimated to cost [redacted] per QALY.

Table 35: Base-case results for the first-line population, OlaDox vs. Dox at list price

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£) incremental (QALYs)
Ola/Dox	[redacted]	3.62	2.115	[redacted]	1.56	0.892	[redacted]
Dox		2.06	1.222				

Source: Eli Lilly submission, table 71, p. 199

Table 36: Base-case results for the first-line population, OlaDox vs. IfoDox at list price

Technologies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£) incremental (QALYs)
Ola/Dox	[redacted]	3.66	2.177	[redacted]	1.3369	0.752	[redacted]
IfoDox		2.34	1.425				

Source: Eli Lilly submission, Table 72, p. 199

Clinical outcomes from the model are shown in Table 37 and Table 38.

Table 37: Clinical outcomes from the model, first-line population, OlaDox vs Dox (years per patient)

Outcome	Mean from clinical trial			Model result (extrapolated to lifetime horizon)		
	OlaDox	Dox	Incremental	OlaDox	Dox	Incremental
Progression free survival (yrs)	_	_	_	0.66	0.52	0.14
Post-progression survival (yrs)	_	_	_	3.52	1.80	1.72
Overall survival (yrs)	_	_	_	4.18	2.32	1.86
Restricted mean*(yrs)	2.51	1.57	0.94	2.39	1.49	0.90

Notes: Clinical outcomes from the model, first-line population, OlaDox vs IfoDox (years per patient); *Restricted to allow comparison of the model mean OS estimates with the KM based data which are only available for 47 months

Source: Eli Lilly submission, Table 73, p.200.

Table 38: Clinical outcomes from the model, first-line population, OlaDox vs IfoDox (years per patient)

Outcome	Model result		
	OlaDox	IfoDox	Incremental
Progression free survival	0.91	0.80	0.11
Post-progression survival	3.44	1.87	1.57
Overall survival	4.35	2.67	1.68

Source: Eli Lilly submission, table 74, p.200.

Disaggregated results are shown below in Table 39 and Table 40.

Table 39: Summary of QALY gain by health state, first-line population, OlaDox vs Dox

Health state	QALY intervention (OlaDox)	QALY comparator (Dox)	Increment	Absolute increment	% absolute increment
Progression-free	0.46	0.36	0.10	0.10	11%
Progressed	1.66	0.86	0.80	0.80	89%
AE	-0.02	-0.01	-0.01	0.01	1%
Total	2.11	1.22	0.89	0.89	100%

Source: Eli Lilly submission, table 75, p.200

Table 40: Summary of QALY gain by health state, first-line population, OlaDox vs IfoDox

Health state	QALY intervention (OlaDox)	QALY comparator (IfoDox)	Increment	Absolute increment	% absolute increment
Progression-free	0.63	0.56	0.07	0.07	9%
Progressed	1.55	0.86	0.68	0.68	91%
AE	-0.02	-0.01	-0.01	0.01	1%
Total	2.18	1.43	0.75	0.75	100%

Source: Eli Lilly submission, table 76, p.201

Table 41: Summary of costs by health state, first-line population, OlaDox vs Dox at list price

Health state	Cost intervention OlaDox	Cost comparator Dox	Increment	Absolute increment	% absolute increment
Progression-free	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Progressed	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
AEs					
Total					

Source: Eli Lilly submission, table 77, p.201

Table 42: Summary of costs by health state, first-line population, OlaDox vs IfoDox at list price

Health state	Cost intervention OlaDox	Cost comparator IfoDox	Increment	Absolute increment	% absolute increment
Progression-free	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Progressed	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
AEs					
Total					

Source: Eli Lilly submission, table 78, p. 202.

Table 43: Summary of predicted resource use by category of costs, first-line population, OlaDox vs Dox at list price

Category	OlaDox	Dox	Absolute Increment	% absolute increment
Cost of study treatment	[redacted]	[redacted]	[redacted]	[redacted]
Total study drug cost	[redacted]	[redacted]	[redacted]	[redacted]
Ola	[redacted]	[redacted]	[redacted]	[redacted]
Dox	[redacted]	[redacted]	[redacted]	[redacted]
Supportive drugs	[redacted]	[redacted]	[redacted]	[redacted]
Administration	[redacted]	[redacted]	[redacted]	[redacted]
Cardiac monitoring costs during and after treatment	[redacted]	[redacted]	[redacted]	[redacted]
Adverse Events costs	[redacted]	[redacted]	[redacted]	[redacted]
Pre-progression general disease management	[redacted]	[redacted]	[redacted]	[redacted]
Regular follow-up visits and imaging	[redacted]	[redacted]	[redacted]	[redacted]
Other direct costs	[redacted]	[redacted]	[redacted]	[redacted]
Post-progression treatment costs	[redacted]	[redacted]	[redacted]	[redacted]
Drug	[redacted]	[redacted]	[redacted]	[redacted]
Administration	[redacted]	[redacted]	[redacted]	[redacted]
Adverse Events	[redacted]	[redacted]	[redacted]	[redacted]
Post-progression general disease management	[redacted]	[redacted]	[redacted]	[redacted]
Regular follow-up visits and imaging	[redacted]	[redacted]	[redacted]	[redacted]
Other direct costs	[redacted]	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	[redacted]	[redacted]

Table 44: Summary of predicted resource use by category of costs, first-line population, OlaDox vs IfoDox at list price

Category	OlaDox	IfoDox	Absolute Increment	% absolute increment
Cost of study treatment	[redacted]	[redacted]	[redacted]	[redacted]
Total study drug cost	[redacted]	[redacted]	[redacted]	[redacted]
Ola / Ifo	[redacted]	[redacted]	[redacted]	[redacted]
Dox	[redacted]	[redacted]	[redacted]	[redacted]
Supportive drugs	[redacted]	[redacted]	[redacted]	[redacted]
Administration	[redacted]	[redacted]	[redacted]	[redacted]
Cardiac monitoring costs during and after treatment	[redacted]	[redacted]	[redacted]	[redacted]
Adverse Events costs	[redacted]	[redacted]	[redacted]	[redacted]
Pre-progression general disease management	[redacted]	[redacted]	[redacted]	[redacted]
Regular follow-up visits and imaging	[redacted]	[redacted]	[redacted]	[redacted]
Other direct costs	[redacted]	[redacted]	[redacted]	[redacted]
Post-progression treatment	[redacted]	[redacted]	[redacted]	[redacted]
Drug	[redacted]	[redacted]	[redacted]	[redacted]
Administration	[redacted]	[redacted]	[redacted]	[redacted]
Adverse Events	[redacted]	[redacted]	[redacted]	[redacted]
Post-progression general disease management	[redacted]	[redacted]	[redacted]	[redacted]
Regular follow-up visits and imaging	[redacted]	[redacted]	[redacted]	[redacted]
Other direct costs	[redacted]	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	[redacted]	[redacted]

Source: Eli Lilly submission, table 80, p.203.

5.2.10 Sensitivity analyses

Eli Lilly performed deterministic and probabilistic sensitivity analyses, as well scenario analyses specific to the UK in order to quantify the effect of parameters' uncertainty in the model and the implications for the ICERs.

5.2.10.1 Probabilistic sensitivity analysis

Probabilistic sensitivity analysis (PSA) is a method of allowing all model parameters which are uncertain to vary simultaneously (for example, the exact HSUV for each state may be uncertain, but the list price of the drug is set by the company and is certain). Parameters were given suitable probability distributions and simultaneously randomly sampled for a total of 1000 simulations by the manufacturer and ICERs were recorded.

The base case PSA (i.e. without varying model assumptions such as the use Dex) results are shown below in Table 45 and Table 46. ICERs from the PSA were [redacted] for OlaDox versus Dox and [redacted] for OlaDox versus IfoDox. Results from the Monte Carlo simulations were also plotted in the (incremental cost, incremental benefit (QALY)) space shown in [redacted]. The proportion of simulations which fall below the "willingness-to-pay" threshold (dotted line) gives the probability of the treatment being cost-effective. OlaDox had a [redacted] probability of being cost-effect against Dox monotherapy and a [redacted] chance of being cost-effect against IfoDox at list price.

Table 45: Base case PSA results for first-line population, OlaDox vs Dox at list price

Technologies	Total costs	Total LYG	Total QALYs	Incremental costs	Incremental LYG	Incremental QALYs	Incremental cost/LYG	ICER incremental (95% CI)
OlaDox	[redacted]							
Dox	[redacted]							
Probability of OlaDox being cost-effective at WTP threshold of £50,000								

Note: Numbers may not compute due to rounding. ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years

Table 46: Base case PSA results for first-line population, OlaDox vs IfoDox at list price

Technologies	Total costs	Total LYG	Total QALYs	Incremental costs	Incremental LYG	Incremental QALYs	Incremental cost/LYG	ICER incremental (95% CI)
OlaDox	[redacted]							
Dox	[redacted]							
Probability of OlaDox being cost-effective at WTP threshold of £50,000								

Note: Numbers may not compute due to rounding. ICER, incremental cost-effectiveness ratio; LYG, life years gained; QALYs, quality-adjusted life years.

Source: Eli Lilly submission, tables 81-82, pp.204-205.

16 [redacted]

[redacted]

[redacted]

17 [redacted]

[redacted]

[redacted]

Cost-effectiveness acceptability curves (CEACs) show the probability of the treatment being cost-effective whilst varying the “willingness to pay”. This is done by changing the gradient of the dotted line from the origin and calculating the proportion of runs below the line. CEACs are shown below in [redacted].

[redacted] 18

[redacted]

[redacted] 19

[redacted]

5.2.10.2 Deterministic sensitivity analyses

Deterministic sensitivity analyses are sensitivity tests that work by changing parameters or structural assumptions by pre-defined amounts, and then investigating the effect these changes have on the base case ICER. The deterministic analyses carried out by the manufacturer are predominantly one-way SAs; an individual parameter or model assumption is adjusted at a time, rather than simultaneously changing multiple ones.

The company categorises the deterministic parameters that they vary as:

- Costs (drug utilisation, administration, monitoring, health state)
- Utilities (HSUVs)
- Drug efficacy (PFS and OS parametric fittings, hazard ratios)
- Analysis settings (discount rates, time horizons, populations)

[redacted] illustrates the company's most impactful deterministic sensitivity analyses and their impacts on the baseline ICERs for OlaDox versus Dox and IfoDox respectively.

[redacted] 20 [redacted]

[redacted]

[redacted]

The company conducted sensitivity analyses to explore the effect of different OS parametric survival functions on the outcome of OlaDox vs. Dox comparison. The resulting ICERs were [redacted] for log-normal, [redacted] for Weibull and [redacted] for Gompertz.

The full list of the manufacturer's changes and the impacts on the ICERs is reported in Appendix 8. Most of their changes had relatively small impacts on the ICERs (within 15% of the baseline estimates). The changes which had the greatest impacts were the choice of parametric survival functions in the OlaDox vs Dox arm (not applicable for IfoDox), changing PPS utilities values and changing drug administrations in the IfoDox arm. Using a Weibull or Gompertz survival function had the greatest impact, increasing the ICER to well beyond [redacted]

5.2.10.3 Scenario analyses

The company also presented a UK specific scenario analysis, in which they adjust multiple inputs and structural assumptions to better represent country-specific characteristics. Their adjustments were (Source: Eli Lilly submission, p. 214):

- Dox cycles being capped at 6 instead of 8 and costs adjusted accordingly, costs for Ola unchanged.
- Efficacy of OlaDox assumed to be unchanged with respect to fewer Dox administrations due to lacking evidence on the efficacy of changing Dox cycles. Threshold analyses also carried out to test this plausibility.
- No patients receive Dex.
- Cardiac AEs adjusted downwards to account for fewer Dox cycles.
- Frequency of cardiac monitoring adjusted downwards.

The company's results arising from the scenario analysis are shown in Table 47 to Table 52. ICERs change from [redacted] for OlaDox vs Dox and [redacted] to [redacted] for OlaDox vs IfoDox.

Table 47: UK practice scenario analysis results for first-line analysis, OlaDox vs Dox at list price

Techno-logies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£) incremental (QALYs)
OlaDox	[redacted]	3.62	2.115	[redacted]	[redacted]	0.892	[redacted]
Dox	[redacted]	2.06	1.222				

Table 48: UK practice scenario analysis results for first-line analysis, OlaDox vs IfoDox at list price

Techno-logies	Total costs (£)	Total LYG	Total QALYs	Incremental costs (£)	Incremental LYG	Incremental QALYs	ICER (£) incremental (QALYs)
OlaDox	[redacted]	3.66	2.177	[redacted]	1.33	0.753	[redacted]
IfoDox	[redacted]	2.34	1.425				

Table 49: Summary of costs by health state, UK practice scenario analysis, first-line analysis, OlaDox vs Dox at list price

Health state	Cost intervention OlaDox	Cost comparator Dox	Increment	Absolute increment	% absolute increment
Progression-free	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Progressed	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
AEs	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Table 50: Summary of costs by health state, UK practice scenario analysis, first-line analysis, Oladox vs IfoDox at list price

Health state	Cost intervention OlaDox	Cost comparator IfoDox	Increment	Absolute increment	% absolute increment
Progression-free	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Progressed	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
AEs	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Source: Eli Lilly submission, Table 88, p. 216

Table 51: Summary of predicted resource use by category of costs, UK practice scenario analysis, first-line analysis, OlaDox vs Dox at list price

Category	OlaDox	Dox	Absolute Increment	% absolute increment
Cost of study treatment	[redacted]	[redacted]	[redacted]	[redacted]
Total study drug cost	[redacted]	[redacted]	[redacted]	[redacted]
Ola	[redacted]	[redacted]	[redacted]	[redacted]
Dox	[redacted]	[redacted]	[redacted]	[redacted]
Supportive drugs	[redacted]	[redacted]	[redacted]	[redacted]
Administration	[redacted]	[redacted]	[redacted]	[redacted]
Cardiac monitoring costs during and after treatment	[redacted]	[redacted]	[redacted]	[redacted]
Adverse Events costs	[redacted]	[redacted]	[redacted]	[redacted]
Pre-progression general disease management	[redacted]	[redacted]	[redacted]	[redacted]
Regular follow-up visits and imaging	[redacted]	[redacted]	[redacted]	[redacted]
Other direct costs	[redacted]	[redacted]	[redacted]	[redacted]
Post-progression treatment costs	[redacted]	[redacted]	[redacted]	[redacted]
Drug	[redacted]	[redacted]	[redacted]	[redacted]
Administration	[redacted]	[redacted]	[redacted]	[redacted]
Adverse Events	[redacted]	[redacted]	[redacted]	[redacted]
Post-progression general disease management	[redacted]	[redacted]	[redacted]	[redacted]
Regular follow-up visits and imaging	[redacted]	[redacted]	[redacted]	[redacted]
Other direct costs	[redacted]	[redacted]	[redacted]	[redacted]
Total	[redacted]	[redacted]	[redacted]	[redacted]

Source: Eli Lilly submission, Tables 85-90, pp. 215-217.

Table 52: Summary of predicted resource use by category of costs, UK practice scenario analysis, first-line analysis, OlaDox vs IfoDox at list price

Category	OlaDox	IfoDox	Absolute Increment	% absolute increment
Cost of study treatment	[redacted]	[redacted]	[redacted]	[redacted]
Total study drug cost	[redacted]	[redacted]	[redacted]	[redacted]
Ola				
Dox				
Supportive drugs				
Administration				
Cardiac monitoring costs during and after treatment				
Adverse Events costs				
Pre-progression general disease management				
Regular follow-up visits and imaging				
Other direct costs				
Post-progression treatment costs				
Drug				
Administration				
Adverse Events				
Post-progression general disease management				
Regular follow-up visits and imaging				
Other direct costs				
Total				

Source: Eli Lilly submission, tables 85-90, pp. 215-217.

5.2.11 Model validation and face validity check

The manufacturer presented their validation and validity check as follows (Source: Eli Lilly submission, p. 218):

Face validity: The model was developed with input from Lilly medical and UK clinicians and health economists. The model was also informed by information obtained from UK STS clinicians at an advisory board held in April 2016.

Internal validity: Verification of all input data with original sources and model programming was conducted by agency staff not involved in development of the model. In addition, the model was validated by an independent health economics consultancy.

Cross validation: No cross-validation could be performed because no suitable studies were identified. The only potentially suitable study identified in the economic SLR was by Guest et al. (2013a) which reported a model evaluating Dox and/or Ifo. However, the model time horizon adopted in this publication was not appropriate for a lifetime analysis of the mSTS population.

External validity: Studies were identified by systematic reviews of clinical trials and observational studies. Dependent, external validity was evaluated by comparing the model outcomes with the JGDG study's Kaplan-Meier curves from JGDG and the study by...Validation for PFS is not shown because the base-case analysis used the Kaplan-Meier PFS data directly; therefore, the model reflected the trial results exactly.

5.3 Critique of companies submitted economic evaluation by the ERG

5.3.1 Model structure

We believe that the cohort-based partitioned survival model with three health states (progression-free, progressed and death) chosen by Lilly is appropriate. It adequately represents the patient pathway.

The model used in TA185,[1] of trabectedin vs. BSC for the treatment of advanced mSTS, had 4 states:

- Disease stabilised with trabectedin
- Progressive disease treated with best supportive care following failure of trabectedin
- Progressive disease treated with best supportive care
- Death

In this model, patients in the trabectedin arm start in the stable state while patients treated with BSC are assumed to start in the progressed health state. This assumption made in TA185[1] is not relevant for the problem under consideration in the current appraisal. Therefore, we consider the model structure chosen by Lilly more suitable for the decision problem.

The cycle length of one week seems appropriate given the treatment administration.

5.3.2 Population

The patient population, as defined by the final NICE scope and licensed indication for OlaDox, consists of Dox-naïve adults with locally advanced or metastatic STS who receive OlaDox as any line of therapy.

However, the base-case cost-effectiveness analysis presented in the submission considered OlaDox as a 1st-line treatment. Eli Lilly stated that this was done "*in view of the anticipated place of OlaDox in the UK STS treatment pathway, as agreed with the NICE team during the scoping workshop*" (Source: Eli Lilly submission, p. 129).

5.3.2.1 OlaDox vs Dox (direct comparison)

A proportion of patients (35%) in the pivotal trial, JGDG,[16] on which the direct comparison was based, underwent prior anticancer treatments. In the company's 1st-line base-case analysis, however, they considered the entire ITT dataset using the line of therapy as a covariate. Due to the small sample size (66 patients in the OlaDox arm and 67 in the Dox arm), this approach was deemed to be more robust than restricting the analysis dataset to

patients receiving OlaDox as 1st-line therapy (40 and 47 patients in the OlaDox and Dox arm, respectively). The ERG agrees with the approach taken by Eli Lilly.

A cost-effectiveness analysis for *any line* of therapy, based on the population from JGDG trial, was also reported: this assumption increased the base-case ICER for OlaDox vs. Dox from [redacted] to [redacted] per QALY.

5.3.2.2 OlaDox vs IfoDox (indirect comparison)

All trials in the evidence network for OlaDox vs IfoDox, with the exception of the JGDG, included 1st-line patients only.

Similar to the approach used for OlaDox vs. Dox analysis, due to the relatively small sample size of the JGDG trial, the entire ITT dataset from JGDG was used in the NMA rather than the 1st-line subgroup of patients.

Eli Lilly reported that a scenario analysis for *any line* of therapy was conducted using the ITT population from the JGDG trial (Source: Eli Lilly submission, Appendix 35, p. 235), which resulted in a lower ICER of [redacted] per QALY compared with the base-case ICER of [redacted] per QALY.

5.3.3 Interventions and comparators

The comparator treatments in the company's submission, Dox and IfoDox, are in line with the NICE scope. The frequency of use of these treatments in the UK, provided by our clinical expert and reported in the submission, are shown in Table 53.

Table 53: Most commonly used regimens as 1st-line therapy in UK practice

Comparator	ERG's clinical expert (PS)	Lilly
Dox	85%	47%
IfoDox	5-10%	16%
Gemcitabine and docetaxel	5-10%	NR in the submission (Eli Lilly submission, Fig. 1, p. 30); ~8% in the poster presentation by Mytelka et al. (2016)[36] reporting the results of Eli Lilly observational study

Key: NR, not reported

Dox was reported by both the ERG and Eli Lilly as the 1st-line therapy most commonly used in mSTS patients in the UK and hence the most important comparator in this appraisal.

5.3.3.1 OlaDox vs. Dox (direct comparison)

In the US-based JGDG study,[16] patients were permitted to receive up to 8 cycles of Dox at a dosage of 75mg/m^2 (a cumulative dose of 600mg/m^2) every three weeks. To mitigate the risk of cardiotoxicity due to Dox, patients in JGDG could also receive the cardioprotectant Dex in cycle 5 to 8.

5.3.3.1.1 Dox regimen in the UK

The recommended standard starting dose of single agent Dox per cycle in STS is 75mg/m^2 of body surface area (BSA) every three weeks .[37] Patients in the UK usually receive up to six cycles of Dox (a total dose of 450mg/m^2), due to the risk of cumulative cardiotoxicity. Dex is not routinely used with Dox in the treatment of STS which was confirmed by our clinical experts.

In the base-case analysis, the use of Dox and Dex was modelled as per JGDG; and a UK practice scenario analysis, with the use of Dox as per UK practice (maximum of 6 cycles instead of 8) and the exclusion of Dex, was explored in a sensitivity analysis (see Section 5.2.10.3, p 120). Respective mean doses and the number of administrations assumed in the base-case and sensitivity analyses are shown in Table 54.

Table 54: Cumulative Dox dose administered in study JGDG and a UK practice scenario

Parameter	OlaDox Arm	Dox Arm	Difference
Maximum of eight treatment cycles (as per JGDG)			
Mean dose (mg/m^2)	73.7	74.7	-
Mean number of administrations	5.7	4.4	-
Mean total cumulative dose (mg/m^2)	416	329	87
Maximum of six treatment cycles only (UK scenario)			
Mean dose (mg/m^2)	73.0	73.4	-
Mean number of administrations	4.7	3.9	-
Mean total cumulative dose (mg/m^2)	342	283	59
Difference in mean total cumulative dose between JGDG (eight cycles) and scenario assuming maximum of six cycles	74	46	28

Dox = Doxorubicin; OlaDox = olaratumab + Doxorubicin. Sources: JGDG study, Lilly data on file 3 (JGDG CSR, 2015[38]).

Source: Eli Lilly submission, Table 42, p. 133

We note that some published studies, as recommended by a UK clinical expert, investigating the relationship between Dox dose and outcomes in mSTS were reviewed by the company. No data were identified that could be used to adjust PFS or OS in the economic model if the Dox dose were limited to 6 cycles, and external clinical opinion was sought at an advisory board attended by nine STS clinical experts from the UK. The clinicians were clear that increasing Dox cycles from 6 to 8 would not be expected to have an impact on the efficacy of the regimen. Based on this advice, the six-Dox-cycle maximum scenario was used in the UK practice scenario analysis, and no adjustment was made to the PFS and OS benefits for OlaDox vs. Dox observed in the JGDG trial.[16]

The mean total cumulative dose of Dox in the UK scenario analysis was the mean dose received within the first 6 cycles of treatment (omitting cycles 7 and 8). We agree that this may reasonably represent the amount of drug that would be received in routine clinical practice if the number of Dox cycles was limited to a maximum of 6 cycles. However, using

this assumption along with clinical effectiveness data from JGDG trial may overestimate the cost-effectiveness of OlaDox.

We also note that in the trial, patients in the OlaDox arm received a mean total cumulative dose of Dox of 416 mg/m^2 which does not exceed the threshold of 450 mg/ m^2 indicated in the UK guidelines for the management of soft tissue sarcomas.[37] For consistency, it would be appropriate to model the number of Dox treatment cycles reported in JGDG study. Therefore, we are satisfied with the respective assumption made by Lilly in their base case. However, we believe that the assumption of no use of Dex would be more appropriate for the main analysis since this assumption would reflect UK clinical practice. Of note, the company assumed in their base case that 59% patients in the OlaDox and 45% in the Dox arm received Dex.

Not modelling Dex as per the JGDG trial, however, may result in inconsistency between costs and health effects due to underestimation of the impact of cardiomyopathy associated with doxorubicin administration on patient quality of life.

Also, modelling the use of Dox as per JGDG but no Dex has only a small effect on the base-case results: ICER reduces from [redacted] to [redacted] per QALY for OlaDox vs. Dox, and from [redacted] to [redacted] per QALY when comparing OlaDox vs. IfoDox.

Therefore in our base case, we adopt the company's approach to modelling Dex.

5.3.3.2 OlaDox vs. IfoDox (indirect comparison)

STS treatment guidelines recommend 9 -10mg/m^2 of Ifo alone. However no dosage is specified for the IfoDox combination.[37] The Ifo SPC recommends one of the two following regimens:

a) 8 - 12 g/m^2 Ifo equally fractionated as single daily doses over 3 - 5 days every 2 - 4 weeks.

b) 5 - 6 g/m^2 Ifo (maximum 10 g) given as a 24 hour infusion every 3 – 4 weeks.

Eli Lilly wrote on p. 134 of their submission:

> *The usual UK practice doses for IfoDox combination in the treatment of advanced STS are 60 mg/m^2 of Dox on day 1 and 9 g/m^2 of Ifo spread over days 1, 2 and 3. This regimen was not represented in the evidence network for the NMA. The efficacy of IfoDox in the model was based on the regimen reported in the study by Judson et al (2014), i.e., a dose of 75mg/m^2 of Dox spread over 3 days and 10g/m^2 Ifo spread over 4 days … since this is the closest to UK practice.*

See Table 20, p78 for further details on regimens from studies examined in the NMA.

Our clinical experts advised us that in terms of efficacy these regimens would not be significantly different.

Since the Ifo SPC recommends the concomitant administration of Mesna and G-CSF to prevent urothelial toxicity and myelosuppression respectively, and these drugs are routinely used in UK practice, the use of these drugs in IfoDox arm was also modelled.

The ERG is satisfied with modelling concomitant drugs in the company's analysis since this is in line with UK practice.

5.3.4 Perspective, time horizon and discounting

Eli Lilly state on p. 132 of their submission that the economic evaluation was conducted from an NHS/PSS perspective and that in the base case, costs and outcomes were discounted at 3.5% per annum in line with the NICE reference case.[17]

A number of sensitivity analyses with 0% and 6% discounting for costs and health effects were undertaken and reported in the company's submission (Eli Lilly submission, p. 210).

The time horizon for the base-case analysis was 25 years, i.e., a lifetime horizon (Eli Lilly submission, p. 136) which is in line with the NICE guidance.[17] The company's decision was based on the OS from the JGDG study extrapolated beyond the study follow-up. We note that this assumption has an impact on cost-effectiveness: assuming a shorter time horizon of 15 years, for example, increases the base-case ICERs for the comparisons of OlaDox vs. Dox and IfoDox from [redacted] to [redacted] per QALY and from [redacted] to [redacted] per QALY, respectively.

5.3.5 Treatment effectiveness and extrapolation

5.3.5.1 OlaDox vs. Dox (direct comparison)

The US-based JGDG study was the only source of clinical effectiveness of OlaDox vs. Dox used by Lilly in their cost-effectiveness analysis.[16]

As noted in Section 5.3.2.1, p123, the analysis was based on the ITT dataset with covariates, which included 66 patients in the OlaDox arm and 67 in the Dox arm (of whom 64 and 65 patients respectively, received the allocated study treatment). The company argued that using data for the 1st-line subgroup (of 40 and 47 patients in the OlaDox and Dox arm, respectively) in the base case would increase uncertainty due to the small sample size.

As we stated in Section 5.3.2.1, we agree with the strategy employed by Eli Lilly.

5.3.5.1.1 Progression-free survival

In the base case, the company used the Kaplan-Meier (KM) PFS data (see section 5.2.6, p 12).

The ERG is satisfied with the choice of KM curves over parametric fits for the base case. However, we believe that the independent assessment of PFS would be more appropriate for the base-case analysis.

Using KM curves from independent assessment increases base-case ICER from [redacted] to [redacted] per QALY. Since the change in the ICER is not very significant, we do not pursue this issue further.

5.3.5.1.2 Progressive disease and overall survival

As the choice of OS extrapolation is key for the cost-effectiveness and the end-of-life analyses, it is essential that this assumption is scrutinised closely.

Since neither the 1st-line nor ITT KM curves for overall survival reached zero during the study observation period of JGDG, 47 months (Figure 12, p 99), parametric survival models were fitted to the OS data to enable extrapolation of the OS curves beyond the study follow-up.

We note that due to the small number of patients and events in the 1st-line subgroup of the phase 2 JGDG study (40/21 and 47/36 patients/events in the OlaDox and Dox arms, respectively), parametric survival models were fitted to the ITT dataset with the line of therapy as a covariate, rather than restricting the analysis dataset to the 1st-line patient subset. The ERG considers this appropriate given the small sample size in the pivotal trial.

Log-cumulative hazard plots for both the 1st-line and ITT populations were approximately parallel (see Appendix 4), suggesting that the proportional hazards assumption may be reasonable for OS. Therefore for the base-case analysis, the “arms together” approach was used when fitting parametric models to the data from both arms (with treatment as a covariate) to allow the inclusion of data on more patients than an “individual arms” approach.

We agree with the “arms together” approach employed by the company in the analysis of overall survival.

Table 25, p 100, shows the covariates explored for the parametric models of OS; those shown in bolded text were significant in most models and were included in the final survival models.

Importantly, the counterintuitive results of the regression analysis, in that the known risk factor for mortality in mSTS patients, age, is found non-significant, may signal a bias in the trial design and/or model specification. According to the sources used by the company to validate extrapolated overall survival curves, age is an important predictor of overall survival in patients with advanced STS (Van Glabbeke, 1999 [5], Karavasilis, 2008[39])

The company stated that selection of the parametric survival functions for inclusion in the company's economic model was based on Akaike information criterion (AIC), Bayesian information criterion (BIC), the visual fit to the KM curves and clinical plausibility, which is in line with NICE DSU guidance.

Extrapolation of overall survival beyond trial follow-up

Due to relatively short follow-up time of the JGDG trial, external data was used to estimate survival for the Dox arm. The company conducted an ad hoc search for literature on long-term survival of patients with advanced STS. Table 55 presents the survival estimates from the identified sources together with those reported in the previous HTA (TA185)[1], predictions from different OS parametric models, and estimates obtained from our clinical expert.

The company explored the clinical plausibility of the 1st-line OS curves for Dox, produced by the different parametric functions, by comparison with OS data from Van Glabbeke et al. (1999).[5] Another study (Karavasilis et al., 2008[39]) was more recent than the one reported by Van Glabbeke, and data were specific to UK patients. However, the Van Glabbeke study included considerably more patients (2,185 vs. 488 in the Karavasilis study).

Table 55: Estimates of longs-term survival in advanced STS patients

Years after diagnosis	Van Glabekke et al. (1999)[a], %	Karavasilis et al. (2008)[b], %	TA185,[1] %	ERG clinical expert's opinion, %	Gutierrez et al. (2007), *median survival*[c], %	OS for Dox arm			
						Gamma (Lilly's base-case)	Log-normal	Weibull	Gompertz
5	8.7	9	5	<10	6	11	7.1	3.9	5.3
9	6.2	5	NR	-	-	6	2.1	0.2	0.3
10	NR	NR	NR	<2.5	1.6	5	1.7	0.08	0.2

Key: a, The results are based on 2,185 patients enrolled in seven trials who received anthracycline-based regimens.[5] The study included patients in the 1st-line and subsequent-line settings; b, All patients (488) were in the 1st-line setting, with most patients on anthracycline-based therapy.[39]; c The results are based on 1348 patients with distant STS from Florida Cancer Data System (1981–2004)[40]

The company stated on p. 149 that (Figure 14 and Figure 15, p 103):

…the gamma survival model was the only one that produced a plausible extrapolation compared to the Van Glabbeke data. The "arms together" gamma parametric model for OS was therefore chosen as the base case for the Dox arm in the economic model.

"Individual arms" functions were explored in the sensitivity analysis in line with NICE guidance.[17]

We agree that based on the visual fit, gamma function provides the best fit to the trial data.

Of note, the 10 year survival predicted by gamma OS is about 5% and 11% in Dox and OlaDox arms, respectively (Table 55). Log-normal OS predicts 1.7% and 4.3% of patients surviving in Dox and OlaDox arms 10 years after diagnosis; predictions of Gompertz OS is 0.2% and 0.3%, respectively; and the respective predictions by Weibull OS are 0.08% and 0.18%. The estimate provided by our clinical expert is <2.5% of patients surviving 10 years after diagnosis.

Importantly, the median age of patients in JGDG trial was 58.5 years. In the Van Glabbeke study,[5] however, patients were substantially younger, with 75.5% of patient ≤ 60 years old. In Karavasilis et al. (2008),[39] patients' median age was 49 years; and it was shown in this study that younger patients were likely to survive longer compared to older patients (Karavasilis et al., 2008, p. 1588).[39]

Since neither AIC nor BIC incorporate the clinical plausibility of any extrapolation, we agree with Lilly's approach of using external evidence to validate the plausibility of extrapolated survival curves. We believe, however, that using the log-term survival statistics reported by Van Glabbeke et al. (1999)[5] to validate extrapolated survival may increase uncertainty in the model predictions, associated with selection bias on the basis of age.[5]

We were advised by our clinical expert (PS) that less than 2.5% of patients with advanced STS would survive 10 years after diagnosis (Table 55), and we believe that for the base-case analysis, using more conservative estimates of long-term survival than those reported in the Van Glabbeke study would be more appropriate.[5]

Another criteria for selection of OS parametric model fits for our base-case analysis are discussed below.

Adjusting mortality for STS patients

Since it was not possible to predict survival in the model beyond 10 years either from the JGDG or the Van Glabbeke data sets, the company compared the hazard of death for

patients from the Van Glabbeke study and patients from the general population of England and Wales of the same age and sex (Figure 15, p 103)

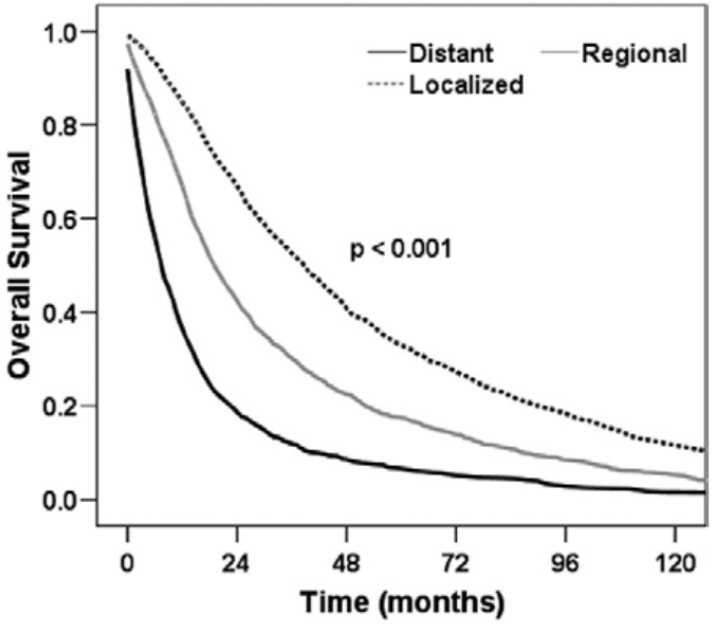
Eli Lilly wrote (Eli Lilly submission, p152):

> *At the end of follow up in the Van Glabbeke study, the hazard for death for patients with STS was 5.19 times higher than that for the general population. Therefore, for the base-case analysis, a HR of 5.19 was applied to the general mortality rates to account for the extra risk of death for cancer patients. Since the HR decreased with time, using 5.19 as an adjustment factor is conservative. For the remainder of model time may overestimate the risk of death in the longer term. If this is the case, this approach would be expected to penalise the OlaDox arm to a greater extent than the Dox arm as there are more patients surviving in the OlaDox arm at the end of the JGDG study. In the base case, the model compares the risk of death from other causes with the risk predicted by the OS function selected and uses the higher of these two risks.*

Of note, the shape of the hazard curve in Van Glabbeke et al. (1999)[5] is unimodal: the hazard of death sharply increases in the first 10 months and then decreases and plateaus starting from about 80 months after diagnosis (Figure 15, p 103). This is characteristic of accelerated failure time models. The Weibull and Gompertz parametric models do not possess such a property - the hazard for these models is monotonic.

We conducted an ad hoc literature search for sources related to long-term survival in the population of interest. We identified a study by Gutierrez et al. (2007),[40] reporting survival of 8,249 patients from Florida Cancer Data System (1981–2004) (Figure 21).

Figure 21: Ten-year survival in STS patients from Gutierrez et al. (2007)



Note: The total number of patients in this study was 8,249; 1,348 patients had distant STS.
Source: Fig. 5, Gutierrez et al. (2007),[40]

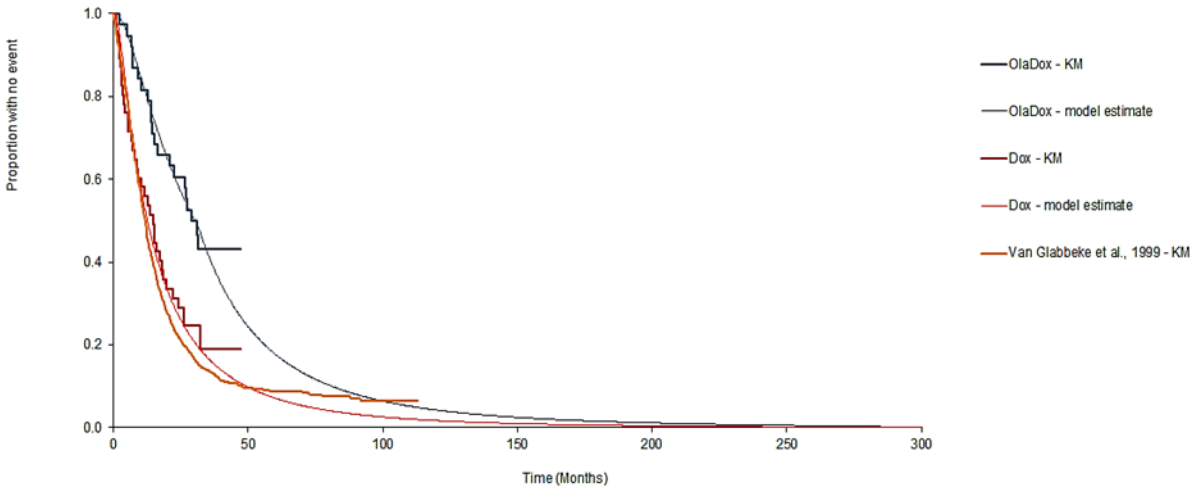
Median 5 and 10 year survival in patients with distant STS, reported by Gutierrez et al. (2007),[40] was 6% and 1.6%, respectively (Table 55, p131.). The median age of patients in this study was 66 years; 49% were males; patient performance status was not reported.

Another study, Maretty-Nielsen et al. (2014)[41], compared 10 years survival outcomes of 1,246 sarcoma patients from Aarhus Sarcoma Registry (148 of whom had metastatic disease) with the survival of general population. The median age of patients was 58 years; 53% were males; performance status of patients was not reported.

As in the Van Glabbeke study, the authors observed non-monotonic hazard of death in the mSTS population from Aarhus Sarcoma Registry. They stated that the majority of deaths in patients with advanced disease occurred within the first 5 years of follow-up. Relative and cancer-specific mortality sharply increased during the first year and then plateaued at around 5 years after diagnosis, which is in line with the results from Van Glabbeke et al. (1999).[5]

Therefore, given the shape of the hazard function observed in mSTS patients, and 5 and 10 year survival estimates provided by our clinical expert (Table 55, p129), we believe that the most appropriate functional form and the most clinically plausible extrapolation of OS is provided by the log-normal function presented in the company's analysis among other candidate models (Figure 14, p 103 and Figure 22).

Figure 22: Extrapolated log-normal functions and Kaplan-Meier curves of OS from JGDG and Van Glabbeke et al. (1999)[5]



Source: Eli Lilly's model

Ten-year model predictions based on the long-normal model are more conservative compared to the relevant predictions by gamma function, and more optimistic compared to those made from Weibull and Gompertz functions (Table 55).

Using the log-normal parametric survival function for OS increases the base-case ICER for OlaDox vs. Dox from [redacted] to [redacted] per QALY. This constitutes item 1 of the PenTAG base case (Table 70, p 157).

Notably, the log-normal model predicts a shorter overall survival for Dox patients, and, therefore, for patients from OlaDox arm (as HR = 1 after the end of the trial follow-up) compared to the company's base case where gamma OS was utilised, which results in higher incremental QALYs for OlaDox vs. Dox, and thus the higher ICER.

Treatment effect

The ERG identified an error in the company's model related to HR settings after the end of observational period. The error has no effect on the company's base-case results where no treatment effect was assumed after 32 months. It only influences a sensitivity analysis exploring the effect of tapering of HR over 12 months - in this analysis, correction of the error increases ICER for OlaDox vs. Dox from [redacted] to [redacted] per QALY.

5.3.5.1.3 Treatment switching

In the trial, upon disease progression, patients in the Dox arm were allowed to receive Ola monotherapy until further disease progression or other discontinuation criteria were met. Among the 65 patients on the Dox arm, 30 later received Ola monotherapy.

Given that our decision problem does not specify the use of Ola monotherapy on progression in the Dox arm, it is necessary to consider whether to attempt to remove the effect of Ola monotherapy on OS for Dox arm.

The company examined a number of naïve and more complex methods proposed by the NICE DSU 16 for the adjustment of treatment switching.[42] Two methods, Rank Preserving Structural Failure Time Model (RPSFTM) and the Iterative Parameter Estimation (IPE) algorithm, were deemed inappropriate since "both of these methods critically rely on a limiting assumption of the "common treatment effect"—that is, the treatment effect received by "switchers" must be the same (relative to the time the treatment is taken for) as the treatment effect received by patients initially randomized to the experimental arm. This assumption is not valid for the JGDG trial since the treatments were different. In the experimental arm, the treatment was OlaDox followed by Ola monotherapy after discontinuation of Dox (without progression) versus Ola monotherapy in the "switchers" (following progression) in the control arm. Therefore, it is unreasonable to assume a "common treatment effect" since the treatment regimens were different. (Eli Lilly submission, Appendix 27, p. 213)

We agree with the decision made by Lilly not to consider the RPSFTM and IPE methods.

The company also investigated the possibility of using Inverse Probability of Censoring Weights (IPCW) and the two-stage methods, both of which do not suffer from this problem.

IPCW method

A IPCW model, examined by Lilly, included 6 baseline covariates: lines of previous treatments (0 vs. 1+), histological subtype (LMS vs. non-LMS), ECOG (0 vs. 1+2), advanced age (< 65 vs. ≥ 65 years), gender, cumulative time from randomization to progression (see Eli Lilly submission, Appendix 27, p. 214 for further details).

Two-stage method

The company used a Weibull parametric model to obtain adjusted survival of those patients from the Dox arm who switched to Ola monotherapy post-progression (Eli Lilly submission, Appendix 27, p. 216). The covariates from the IPCW analysis together with the indicator of

crossover were examined, and Weibull was selected based on model fit, and the lowest AIC and BIC values.

Both methods, the IPCW and the two-stage method, estimated counterfactual survival times (i.e., survival times that would have been observed in the absence of treatment switching).

The resulting adjusted HRs are shown in Table 56 together with the results from the main (ITT) analysis and a naïve method where Dox patients were censored at the time of initiation of Ola monotherapy.

Table 56: Summary of HR from all analyses for comparisons within each row of values

	ITT (Main Result)	Censoring at Time of Initiation of New Treatment	ITT (With Covariates From IPCW)[a]	IPCW	Two Stage
Unstratified	0.517		0.517	0.458	0.495
95% CI	0.341-0.786		0.309-0.865	0.268-0.781	0.305-0.755
Stratified[b]	0.463	0.425	0.560	0.482	0.440
95% CI	0.301-0.710	0.193-0.933	0.332-0.945	0.279-0.832	0.267-0.699

Key: CI, confidence interval; ECOG, Eastern Cooperative Oncology Group; HR, hazard ratio; IPCW, inverse probability of censoring weights; ITT, intention to treat.
Notes: a, Adjusting for covariates age, sex, baseline ECOG, previous lines of treatment, histological tumour type, and time to progression; b, Stratified for previous lines of treatment and histological tumour type.
Source: Eli Lilly submission, Appendix 27, p.218.

The company acknowledged the limitations of IPCW and two-stage methods. IPCW method, for example, is more sensitive than other methods to small sample sizes and extreme switching proportions. In JGDG, there were 67 patients who received Dox treatment, and 45% of those patients switched to Ola monotherapy. The sample size and proportion of switchers in JGDG trial were similar to those in the treatment switching analysis reported by Latimer et al. (2015),[43] where the IPCW method was not considered appropriate. The authors argued: "…including several covariates within statistical models may result in convergence issues when the patient and event numbers are small. In addition, because the weights are applied to control group of non-switchers to adjust for the censoring of switchers, the method can become highly prone to error because the number of control group non-switchers is small."

The company stated that "despite the limitations that could potentially bias the results of the analyses, all four methods estimated similar HRs. Based on these supportive analyses, the conclusion is that control-arm patients who received Ola monotherapy had similar OS as compared to control-arm patients who did not receive Ola monotherapy. Therefore, receiving Ola monotherapy after progression on Dox monotherapy has not altered the hazard of death compared to receiving other treatments and therefore does not alter the conclusions about

the trial comparison of OS between OlaDox and Dox arms." (Eli Lilly submission, Appendix 27, p. 218)

Therefore, the company did not adjust for post-progression Ola monotherapy in Dox patients in the main analysis, and no sensitivity analysis addressing potential uncertainty related to treatment switching was reported in the company's submission.

Table 57: Baseline characteristics for patients in the Dox arm according to whether subsequently received Ola monotherapy (ITT Population)

	Dox arm patients N = 67 (%)	Dox arm patients with subsequent Ola monotherapy N = 30 (%)	Dox arm patients that did not receive subsequent Ola monotherapy N = 37 (%)
Gender (n %)			
Male	33 (49.3)	14 (46.7)	19 (51.4)
Female	34 (50.7)	16 (53.3)	18 (48.6)
Age group (n %)			
< 65 years	43 (64.2)	23 (76.7)	20 (54.1)
≥ 65 years	24 (35.8)	7 (23.3)	17 (45.9)
ECOG PS (n %)			
0	38 (56.7)	19 (63.3)	19 (51.4)
1	26 (38.8)	11 (36.7)	15 (40.5)
2	3 (4.5)	0	3 (8.1)
Histology tumor type (CRF) (n %)			
Leiomyosarcoma	27 (40.3)	11 (36.7)	16 (43.2)
Other	40 (59.7)	19 (63.3)	21 (56.8)
Number of lines of previous treatment (CRF) (n %)			
0	47 (70.1)	24 (80.0)	23 (62.2)
1 or more	20 (29.9)	6 (20.0)	14 (37.8)

Following our request for the data on baseline characteristics of those patients in the Dox arm who did not switch to Ola monotherapy, this information was provided by the company (see Table 57). The patients who switched to Ola monotherapy were younger than those who did not (76.7% of switchers were <65 years old while only 54.1% of non-switchers fall into this age category). A higher proportion of patients who switched had ECOG PS of 0 (63.3% switchers vs. 51.4% in the non-switchers group). Switchers were less exposed to previous treatments than those who did not switch (80% vs. 62%, respectively). Therefore, the company's statement that "the group receiving Ola had similar baseline prognostic characteristics as those on the control arm who have never received Ola" (Eli Lilly submission, Appendix 27, p. 211) is not supported by data.

We believe that the results of the analyses of treatment switching and conclusion drawn by the company should be considered with caution.

In response to our questions, the company acknowledged that "the trial was not designed (sample size, hypothesis testing, etc.) to assess the efficacy of Ola monotherapy vs. Dox monotherapy; therefore no inferences can be drawn from these results."

Even though we are satisfied with Lily's decision not to adjust OS for switching, there then remains the question of whether to cost for the Ola monotherapy in the Dox arm, and this will be discussed in Section 5.3.7.1.3.

At the Decision Problem meeting on 17th October 2016, the ERG requested all the relevant raw data used in the company's analysis of the effect of treatment switching on survival, which would allow the ERG to validate the results. The company, however, did not provide the data requested by the ERG, hence, no validation has been performed.

5.3.5.2 OlaDox vs. IfoDox (indirect comparison)

In the absence of head-to-head data comparing OlaDox vs IfoDox, a network meta-analysis was conducted using fractional polynomials. All trials in the evidence network, with the exception of the JGDG, included 1st-line patients only. Therefore, the fractional polynomial models were fitted to the JGDG ITT dataset without including a covariate for line of therapy (see Section 4.3 for further details on the methodology). Eli Lilly wrote (Eli Lilly submission, p. 169):

> *"The parameters of the survival functions were applied in the economic model to calculate the survivor proportion in each model cycle. The … median estimates are compared with the trial data …Note that the average outcomes for Dox across the trials included in the NMA were applied, rather than the outcomes for the Dox arm of the JGDG study specifically, in order to reflect outcomes for the advanced STS population as a whole as observed across the trials in the NMA. The treatment effects for the other interventions (as estimated by the fractional polynomials NMA) were applied to these average Dox outcomes. The outcomes differ somewhat from the Kaplan-Meier data but are expected to be more generalisable to the advanced STS population as a whole. In addition, the median PFS and OS from the NMA are not expected to closely match those in the original trials because the NMA provides an adjusted indirect comparison based on average outcomes for Dox rather than a naïve comparison of individual trial arms."*

The median PFS and OS estimates from fractional polynomials used in the company's model are compared with the trial data in Table 58.

Table 58: Fractional Polynomial functions estimated by NMA

	Lilly[1]		ERG[2]	
Parameter	OlaDox	IfoDox	OlaDox	IfoDox
PFS (investigator assessed)				
Scale	[redacted]	[redacted]	[redacted]	[redacted]
Shape 1	[redacted]	[redacted]	[redacted]	[redacted]
Shape 2	[redacted]	[redacted]	[redacted]	[redacted]
Median (years)	[redacted]	[redacted]	[redacted]	[redacted]
Median from trial (years)	[redacted]	[redacted]	[redacted]	[redacted]
OS	[redacted]	[redacted]	[redacted]	[redacted]
Scale	[redacted]	[redacted]	[redacted]	[redacted]
Shape 1	[redacted]	[redacted]	[redacted]	[redacted]
Shape 2	[redacted]	[redacted]	[redacted]	[redacted]
Median (years)	[redacted]	[redacted]	[redacted]	[redacted]
Median from trial (years)	[redacted]	[redacted]	[redacted]	[redacted]

Key: IfoDox, ifosfamide + Doxorubicin; OlaDox, olaratumab + Doxorubicin; OS, overall survival; PFS, progression-free survival.
Notes: [1]Estimated from median values of coefficients of fractional polynomials estimated in NMA; [2]Estimated from mean values of coefficients of fractional polynomials estimated in NMA;[3] Based on Tap et al. (2016)[16]; [4] Based on Judson et al. (2014)[21]
Source: Eli Lilly submission, Table 49, p. 154

Importantly, the survival curves in the company's model, presented in Figure 24 and Figure 25 below, are based on median values of coefficients of fractional polynomials estimated in NMA, not mean values, and, when replaced corrected, increase the base-case ICER from [redacted] to [redacted] per QALY. This constitutes item 2 of the PenTAG base case (Table 70, p 157).

Of note, the mean values of the coefficients result in longer PFS in both arms compared to predictions based on medians (Table 58). Overall survival based on the mean coefficients, however, is slightly shorter for IfoDox.

OS after end-of-trial follow-up

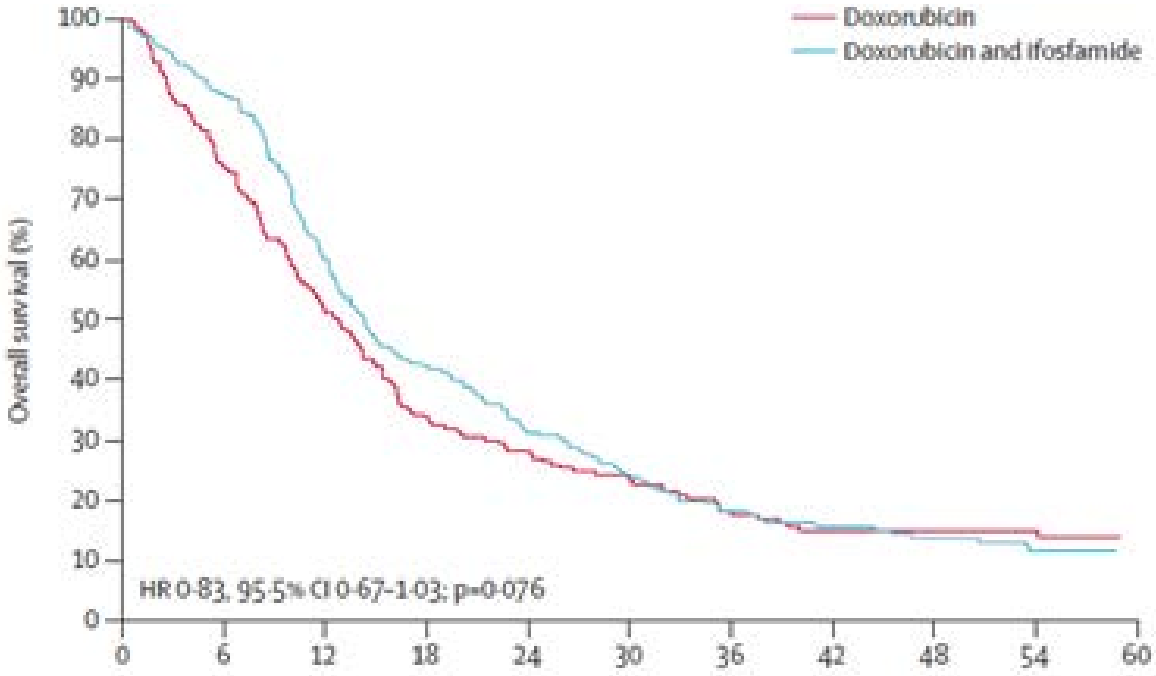
Eli Lilly stated (Eli Lilly submission, p. 159):

> *"For OlaDox, and IfoDox, Kaplan-Meier OS estimates were available for a period of approximately 47 months from JGDG (Figure 12) and 60 months from the Judson 2014 study. For IfoDox, the Kaplan-Meier OS curve converged with the Dox curve at approximately 30 months, and the appearance of the curves suggested that there was no reduction in the risk of death between 30 and 60 months for patients who received IfoDox compared with patients who received Dox. Therefore, for IfoDox, the economic model assumes the same hazards for death as for Dox after the follow-up period of the Judson*

2014 trial; the long-term hazards for Dox were estimated using the fractional polynomial function for Dox and STS-adjusted age/sex-specific mortality data. The same approach is taken for OlaDox (assuming no treatment effect after trial follow-up).

Kaplan-Meier curves of overall survival from Judson et al. (2014)[21] are shown on Figure 23.

Figure 23: Overall survival for IfoDox and Dox (Judson et al., 2014)



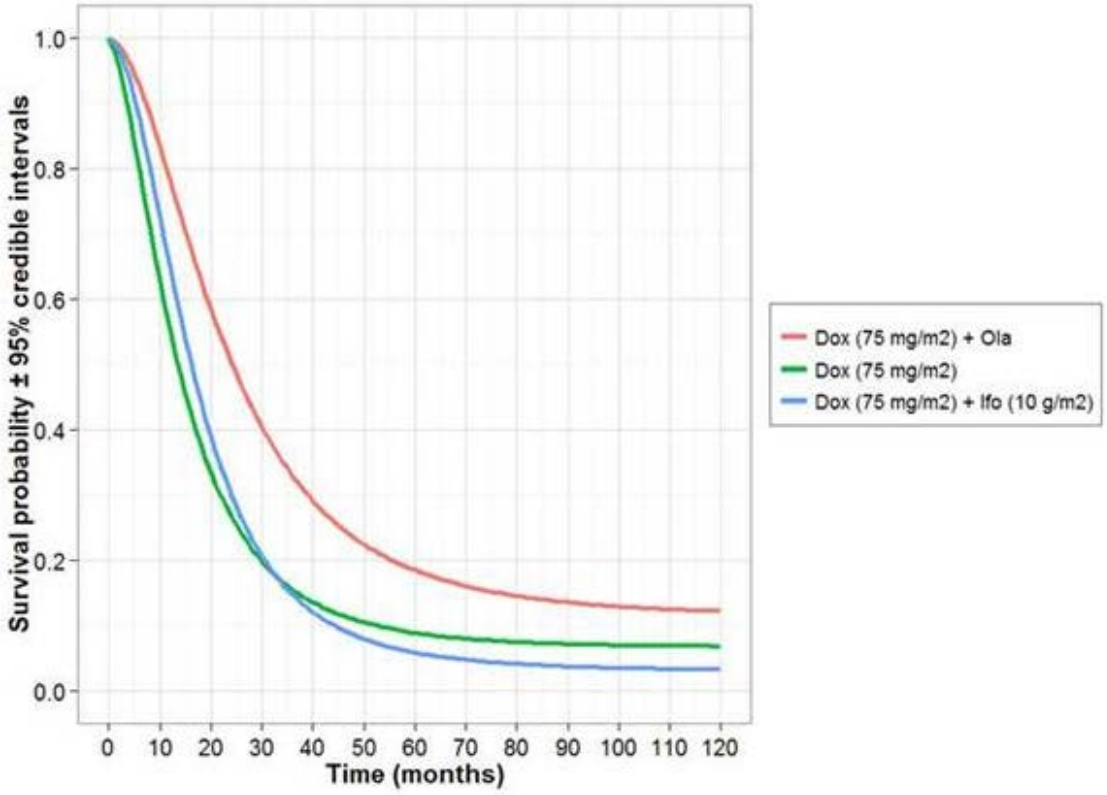
Source: Judson et al. (2014)[21]

Eli Lilly wrote (Eli Lilly submission, p. 159):

The model allows the user to set the trial follow-up period individually for each intervention; the fractional polynomial function for the intervention is applied up to this time; thereafter, the fractional polynomial function for Dox and STS-adjusted age/sex-specific mortality data are applied (i.e. the model assumes no treatment effect for OlaDox or IfoDox versus Dox after the time specified as the end of trial follow-up, see 0). For the base-case analysis, the indirect comparison assumes no treatment effect for OlaDox or IfoDox versus Dox after 32 months. This time-point was selected for consistency with the base-case direct comparison.

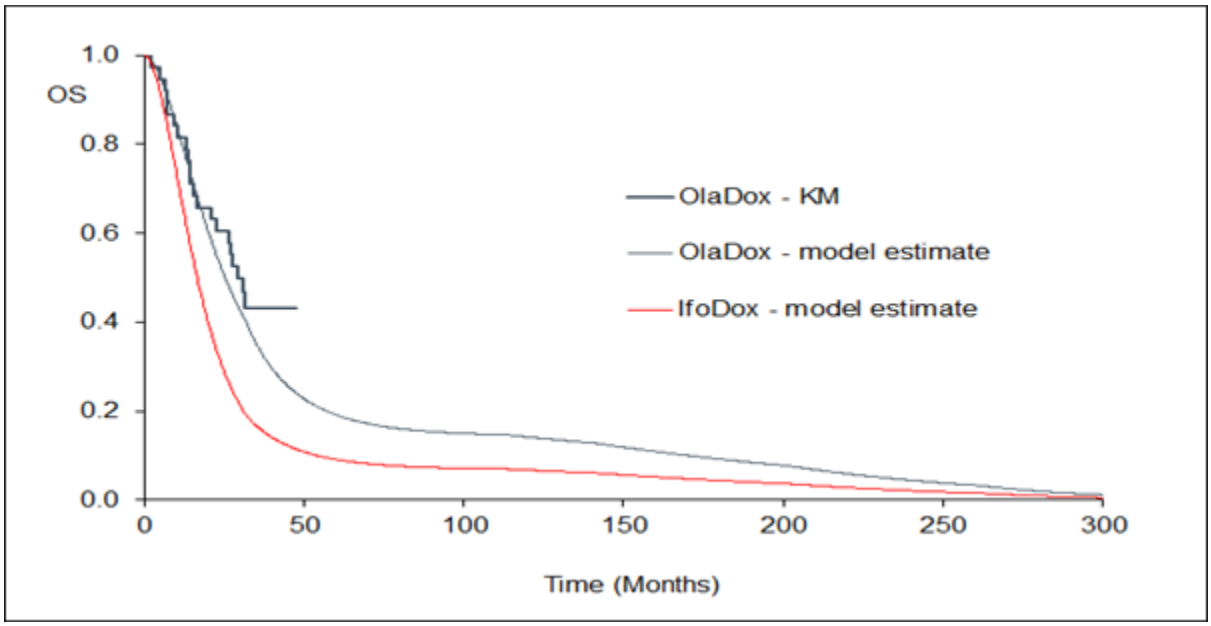
The resulting OS functions for the indirect comparison with IfoDox in 1st-line treatment are presented in Figure 24.

Figure 24: Fractional polynomial survival probability curves for each treatment (NMA report)



Key: HR=1 after 32 months
Source: Lilly's submission, Figure 38, p. 160

Figure 25: Model OS prediction for OlaDox and IfoDox, first-line treatment (indirect comparison using fractional polynomials)



Key: Dox, doxorubicin; IfoDox, ifosfamide + doxorubicin; KM, Kaplan-Meier; OlaDox, olaratumab + doxorubicin; OS,overall survival; STS, soft tissue sarcoma.

Notes: Fractional polynomial functions for OlaDox and IfoDox were used up to 32 months; thereafter, the model assumed no treatment effect (the fractional polynomial for Dox is applied for both treatment arms). STS-adjusted age-specific mortality rates were applied wherever these gave a higher risk of death than the survival functions. KM data shown in the figure are for first-line treatment with OlaDox in the JGDG study.

We are generally satisfied with the NMA for OlaDox and IfoDox, conducted by Lilly (see Section 4.3, p 73 for a detailed critique).

5.3.6 Health related quality of life

Since HRQoL data was not reported in JGDG study, Eli Lilly conducted a systematic literature review to identify published health-state utility estimates (see section 5.1.3.2, p 90 for further details).

Three publications, Reichardt et al. (2012),[2] Amdahl et al. (2014)[3] and Delea et al. (2014),[4] fulfilled the requirements of the NICE reference case. HSUVs from the Reichardt study were deemed most appropriate for inclusion in the model.

5.3.6.1 Health related quality of life used in previous NICE TAs and other sources

In TA185[1] (Yondelis (trabectedin) for the treatment of advanced soft tissue sarcoma), utility estimates from a patient population with advanced nonsmall cell lung cancer, reported by Nafees et al. (2008)[44] were used in the absence of STS specific data (Table 59). According to the company submission for TA185,[1] "Clinical expert input recommended that in the absence of STS specific data, lung cancer could act as an appropriate proxy disease, based on comparable prognosis and disease stage." (p.74)

The mean age of the patient population in this study was 40.51 years (SD=14.91); 38% of patients were females. Utility estimates reported by Nafees et al. (2008)[44] were based on the Standard Gamble technique. The utility decrement for progressive disease was estimated from a mixed model analysis with random effects.

Table 59: Utility values used in TA185

Health Effect	Parameter estimate	Standard error	Source
Stable	0.6532	0.0222	Nafees et al. (2008)
Progressive (decrement)	-0.1798	0.0217	Nafees et al. (2008)

Source: TA185[1]

5.3.6.2 Health related quality of life reported in sources selected during the systematic literature review

Health state utility values identified in systematic literature review are shown in Table 60.

Table 60: Utility values reported in Reichard, Amdahl and Delea

	Source					
	Amdahl et al. (2014) [3]		Delea et al. (2014)[4]		Reichardt et al. (2012)[2]	
Health state	*Parameter estimate (SE)*				*Utility mean (SD)*	
	Pazopanib	Placebo	Pazopanib	Placebo		
Pre-progression	0.674 (0.015)[1]	0.678 (0.024)[1]	0.674 (0.015)	0.678 (0.024)	1st- line treatment	0.72 (0.31)
					2nd- line treatment	0.64 (0.33)
					3rd- line treatment	0.77 (0.14)
					After chemotherapy	0.77 (0.14)
Post-progression	0.568 (0.044) [1]	0.636 (0.040)[1]	0.349[2] (mean value based on the utility decrement of 0.239 and the utility in the terminal phase of the disease reported by Shingler et al. (2013) [29])	0.344[2] (mean value based on the utility decrement of 0.253 and the utility in the terminal phase of the disease reported by Shingler et al. (2013) [29])	Progressive disease (on or off chemotherapy)	0.56 (0.27)

Notes: 1, As stated in the source, Amdahl et al. (2014) [3], the utilities were estimated from a phase III randomized controlled trial, PALETTE, assessing pazopanib vs.placebo for the treatment of patients with advanced/metastatic STS (n= 369) who had received prior treatment with chemotherapy (reported by Graaf et al. (2012)[45]). PD utility was assessed using the utility in pre-progression state and the decrement of 0.239 (both based on PALETTE study), and the utility in the terminal phase of the disease reported by Shingler et al. (2013) [29]). However, the numerical values for PD utilities, reported in the source and shown in the table, could not be replicated based on the description of the strategy in the main text; 2, The numerical values for PD utilities were not provided in the source, and were calculated by the ERG using the strategy described in the source, Delea et al. (2014).[4]

As stated in Amdahl et al. (2014),[3] "In the palette trial, the EuroQol Group's EQ-5D (Rotterdam, Netherlands) was assessed only at baseline and week 4; the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire (qlq-C30) was also assessed at weeks 8 and 12. A mapping algorithm was therefore developed using data from the EQ-5D and qlq-C30 at baseline and week 4 to predict EQ-5D utility values at weeks 8 and 12 from the qlq-C30 at those later assessments. The observed and mapped utility values were then combined to calculate mean utility values for each group for all pre- and post-progression assessments. The mean time from progression to post-progression utility assessment in palette was limited to approximately 1 week in both groups. For that reason, the mean differences in utility for post- compared with pre-progression in palette reflect only declines in utility values immediately after progression; they do not reflect the declines in utility that would be expected over the entire post-progression period. As a consequence, post-progression utility values for pazopanib and placebo were calculated *by combining treatment group–specific estimates of the mean decrement in utility post-progression in palette* (reflecting the period immediately after progression) *with an estimate of utility in the terminal phase of the disease*. The latter estimate was based on the estimated utility value for progressive disease from a vignettes study (mean ± standard error: 0.263 ± 0.0231)".

The company wrote on p. 172 of their submission that "Amdahl and Delea … reported similar pre- and post-progression utility values from patients receiving second-line treatment from the PALETTE study on pazopanib".

We note that the utilities for pre-progression reported in these cost-effectiveness studies were indeed the same, and the same strategy was employed to estimate the utility values for post-progressive disease state. However, PD utilities presented in Table 3, p. 5 in Amdahl et al. (2014)[3] are not consistent with their main text (p. 2), where the strategy for utility estimation was described. Moreover, Delea et al. (2014)[4] did not present numerical values for utilities in PD; they only presented utility decrements of post-progression vs. pre-progression (Table 1, Delea et al. (2014)[4]) along with the description of the same strategy for estimation PD utilities as in the Amdahl study. We have estimated the relevant utilities based on the above strategy (see Table 60).

The median age in PALETTE study was 51.9 and 56.7 in the placebo and treatment arms, respectively; 56% and 60% were females; all patients had WHO performance status of 0-1.

Importantly, in both cost-effectiveness studies, Amdahl et al. (2014)[3] and Delea et al. (2014)[4], utilities were among the most influential model parameters identified in deterministic sensitivity analyses.

5.3.6.3 Health related quality of life in the company's model

The HSUVs of 0.72 and 0.56 were assumed in the company's base case for pre- and post-progression states, respectively. They were adopted from the study by Reichardt et al. (2012),[33] which reported health state utility values for patients with mSTS and metastatic bone sarcoma with *favourable response* (complete response, partial response or stable disease) according to the WHO or RECIST 1.0 criteria *after any line of chemotherapy*. The mean age at metastatic disease diagnosis was 49.5 (SD = 17.1), and the majority of patients (99 out of 120) had mSTS.

The company acknowledged that the study selection criterion (that patients have a *favourable response*) may have resulted in higher utility values than would be expected for all patients.

Notably, the patients in the Reichardt study were substantially younger than in the JGDG trial (49.5 years (SD = 17.1) vs. 56.8 years (SD=12.5) and 58.3 years (SD=12.5) in the OlaDox and Dox arms, respectively), which might lead us to question their relevance.

Along with the utilities for the 1st-line pre-progressive disease and progressive disease, which were used in Lilly's model, Reichardt et al. (2012)[33] reported a utility value for 2nd -line pre-progression (0.64). The company, however, did not use this estimate in patients who discontinued the study treatment but have not yet progressed.

The utility values in Eli Lilly's main analysis were higher than the estimates used in TA185:[1] PFS and PD utilities in the HTA were 0.65 and 0.47, respectively, while in the company's main analysis the utilities were 0.72 and 0.56. Importantly, the estimates in TA185[1] were obtained from a substantially younger patient population with the mean age of 40.51 years (see section 5.3.6.1**,** p 143.).

The company argued (p. 173) that "the study by Reichardt [33] reports health state utility values by line of chemotherapy and health state (progression-free and progressed). Health State Utility Values (HSUVs) from this study were therefore considered most appropriate for inclusion in the model, although the study had some limitations."

We believe that the eligibility criterion (of having a *favourable response* to *any line of treatment*) and substantially younger age of the patient population in the Reichardt study compared to the population in JGDG may significantly overestimate the health utilities and thus bias cost-effectiveness. In terms of age similarity of respective patient populations, the utility estimates from Amdahl et al. (2014)[3] and Delea et al. (2014)[4] are more relevant to the decision problem under consideration. These values, however, were based on patients in

the second-line STS setting, and were estimated from vignettes, which are considered very low grade evidence.

Utility values in TA185[1] were not directly related to soft tissue sarcoma.

Therefore in our base case, we adopt the estimates from Reichardt et al. (2012)[33] as in Lilly's base case. We conduct a sensitivity analysis for the utility values from Amdahl et al. (2014)[3] and Delea et al. (2014)[4] (Section 5.4.2).

5.3.6.4 Disutilities due to adverse events

Eli Lilly assumed zero disutility for hepatic toxicity based on Nafees et al. (2008).[44] In the source, however, this was an arbitrary assumption not supported by any data. We were advised by our clinical expert that about 50% of such events require hospitalisation. Therefore, we think that the assumption of no disutility of the hepatic toxicity does not reflect the quality of life of mSTS patients experiencing such an event.

Assuming disutility for hepatic toxicity to be the same as for fatigue increases the ICERs only slightly. Therefore in our base case, we adopt the same disutility for this AE as in the submission.

5.3.6.5 Age-related disutilities

We note that Lilly did not account for age-associated disutilities in their analysis.

5.3.7 Resources and costs

5.3.7.1 Drug acquisition

Ola and G-CSF (filgrastim) are dosed based on patient's weight. The doses of other drugs, considered in this appraisal, are given proportional to body surface area (BSA).

Eli Lilly assumed a mean weight of 77.3 kg referencing GeDDiS trial which was conducted mainly in the UK. A mean BSA of 1.91m^2 was taken from Health Survey of England (2013).

We could not verify the mean weight in the reference provided by the company since it is available only as an abstract. From personal communication with the lead author, [redacted] by the lead author.

The mean weights of patients in OlaDox and Dox arms from JGDG study were 85.8 kg (SD=23.00) and 82.5 kg (SD=23.40), respectively (CSR, Table JGDG.14.6, p. 230).

Assuming the mean weight of 82.5 kg (as in Dox arm) results in the ICERs of [redacted] and [redacted] per QALY for the comparisons with Dox and IfoDox, respectively; the mean

weight of 85.8 kg (observed in OlaDox patients) results in the ICERs of [redacted] and [redacted] per QALY.

We believe that for consistency between costs and outcomes, costing of drug acquisition should be based on the characteristics of patients from the trials used in the analysis.

Adopting the conservative estimate of the mean weight of 82.5 kg from JGDG constitutes item 4 of the PenTAG base case (Table 70).

Eli Lilly assumed a mean BSA of 1.91 m^2 based on Health Survey of England, (2013).

The mean BSA in OlaDox and Dox arms in JGDG study were 2.0 (SD=0.30) and 1.9 (SD=0.26), respectively (CSR, Table JGDG.14.6, p. 230). The BSA of 2.0 results in the ICERs of [redacted] and [redacted] per QALY for the comparisons with Dox and IfoDox, respectively.

An overall and gender specific mean BSA values for patients receiving chemotherapy for various cancers in the UK, was reported by Sacco et al. (2010).[46] The reported mean BSA (1.79 m^2, 95% CI 1.78–1.80), was lower than the estimate used by the company.

Assuming the mean BSA of 1.8, based on Sacco et al. (2010),[46] and the gender-specific incidence of STS reported by Cancer Research UK [8] increases the ICER for OlaDox vs. IfoDox from [redacted] to [redacted] per QALY, and has virtually no effect on the base-case ICER for OlaDox vs. Dox, changing it from [redacted] to [redacted] per QALY.

We have found only slight discrepancies in Dox prices from Table 27, p 107, with those reported in eMIT, which have a negligible effect on the ICERs.

5.3.7.1.1 OlaDox vs. Dox

We note that in the 1st-line base-case analysis, Lilly modelled the mean number of administrations of Ola in the OlaDox arm (including Ola monotherapy after discontinuation of Dox treatment) of [redacted], which was estimated from patients in JGDG trial with no prior treatment (Table 29, p 108). Notably, the mean number of administrations for *all* patients receiving Ola in OlaDox arm was [redacted] (CSR, page 167, Table JGDG12.12).

The ERG is satisfied with the assumption on the mean number of Ola administrations in the base-case analysis.

5.3.7.1.2 OlaDox vs. IfoDox

We note that in the OlaDox vs. IfoDox analysis, Lilly assumed that the regimes, Dox 60mg/m^2 + Ifo 9g/m^2 and Dox 75mg/m^2 + Ifo 10g/m^2, have similar efficacy. Our clinical expert advised us that in terms of efficacy, these regimens would not be significantly different.

As stated in the company's submission (Eli Lilly submission, p. 179):

> *the treatment costs for the IfoDox regimen were calculated based on the dose commonly used in UK practice, 9mg/m^2 of Ifo over 3 days and 60mg/m^2 of Dox, as recommended by advisors at the UK STS advisory board. Data on mean number of IfoDox cycles was sourced from the publication by Judson et al 2014 [26] (4.419 cycles of IfoDox and 4.096 cycles of Dox). The treatment costs for IfoDox were calculated by multiplying the planned dose by number of administrations per cycle (three for ifosfamide since it is administered over 3 days (including 1 Dox administration on day 1) and the mean number of cycles from Judson et al 2012 [26]. There was no data available in the literature to estimate the extent of dose reduction on IfoDox. The default dose data for IfoDox in the model therefore only reflects the planned dose and may overestimate the drug cost. A sensitivity analysis was performed to explore the impact of dose reduction for IfoDox: 20% decrease in IfoDox dose increased the ICER from [redacted] to [redacted] per QALY.*

The ERG believes that for consistency, the same assumption on treatment dose (the assumption of dose reduction) should have been used for both treatments in the base-case analysis.

Eli Lilly assumed the price of £91.32 for 1000mg of Ifo and £179.88 for 2000mg of Ifo referring to BNF 70 (Sept 2015 - March 2016) (see Table 27, p107, and Table 69 of Eli Lilly's submission, p. 192). However, the most recent prices based on BNF (February 2017)[47] are £66.08 and £130.04 for 1g and 2g vials, respectively. The assumption of the most recent prices for Ifo increases the ICER from [redacted] to [redacted] per QALY, which constitutes item 5 of the PenTAG base case (Table 70, p 157).

In the submission, the price for 1000mg vial of Mesna was £29.41, and £13.41 for 400mg vial. However, the most recent prices for Mesna (according to BNF, February 2017[47]) are £9.77 and £3.95 for 1000mg and 400mg vial, respectively. Assuming the most recent prices of Mesna increases the base-case ICER from [redacted] to [redacted] per QALY. This is item 6 of the PenTAG base case (Table 70, p 157).

There is a discrepancy in the prices for 0.3mg and 0.48mg vials of G-CSF reported in the submission and BNF,[47] which affects the ICER only slightly changing it from [redacted] to [redacted] per QALY. Therefore for our base case, we adopt the prices for G-CSF used by Lilly.

5.3.7.1.3 Ola monotherapy in Dox patients

We note that in the base-case ITT analysis, the Ola monotherapy was not costed in Dox patients who switched to Ola.

NICE TSD DSU 16[42] states that "the ITT approach represents an accurate economic evaluation of the RCT because it models exactly what happened in the trial; survival estimates are not adjusted for treatment switching, but the cost of the treatments switched to are included in the analysis. This might be described as a "full ITT" analysis. The usefulness of the technique for use in HTA is uncertain, given the economic evaluation decision problem…To address the economic evaluation decision problem, it would be preferable to accurately adjust survival estimates for switching and to exclude the costs of switching treatments." (p. 15)

Eli Lilly argued: "The cost of Ola monotherapy for patients randomised to Dox who received post-progression Ola in JGDG have not been included in the economic model as use of Ola monotherapy will not occur in routine practice."

The ERG is satisfied with the approach taken by Lilly.

5.3.7.1.4 Vial sharing

In both comparisons, OlaDox vs. Dox and OlaDox vs. IfoDox, no vial sharing was assumed in the base case. We think that this assumption is reasonable since, with a rare cancer, vial sharing is less likely. This was confirmed by our clinical expert.

A sensitivity analysis making the assumption of no wastage has been performed, and the ICERs were only slightly lower than those in the base case.

5.3.7.1.5 Availability of vial sizes

In the base-case analysis, availability of both vial sizes of Ola, 500mg and 190mg, was assumed. As it was explained in the company's submission, "conditional marketing authorisation has been granted for the 500mg vial of olaratumab. A regulatory submission for the 190mg vial is planned for [redacted], with anticipated marketing authorisation in [redacted] and UK product availability in [redacted]." (Eli Lilly submission, p. 22)

In response to our question on the availability of different vial sizes, the company wrote:

[redacted]

The ERG was advised by NICE to model what would be available in the NHS at the time of the recommendation. Therefore, in our base-case, we assume that only 500 mg vial of olaratumab is available. This assumption increases the base-case ICER for OlaDox vs. Dox

from [redacted] to [redacted] per QALY; and from [redacted] to [redacted] per QALY for OlaDox vs. IfoDox comparison. This constitutes item 7 of the PenTAG base case (Table 70).

We also conducted a scenario analysis, assuming the availability of both vial sizes (see Section 5.4.2).

5.3.7.1.6 Pharmacy drug preparation costs

We note that the company did not include any costs related to drug preparation by a hospital pharmacist.

5.3.7.2 Drug administration

"Olaratumab is administered as an IV infusion over 60 minutes at an infusion rate not exceeding 25mg/minute on days 1 and 8 of a 21 day cycle".(Source: Eli Lilly submission, p. 24). According to the Summary of Product Characteristics, premedication with an H1 antagonist (e.g., diphenhydramine) and dexamethasone (or equivalent medicinal products) should be given, intravenously, 30–60 minutes prior to the olaratumab doses on days 1 and 8 of cycle 1 in all patients. For subsequent cycles, premedication with an H1 antagonist (e.g., diphenhydramine) should be given intravenously 30–60 minutes prior to each dose of olaratumab.

The company states that "Doxorubicin is administered as an IV infusion one hour after olaratumab on day 1 of each 21 day cycle…. if premedication was required prior to the first doxorubicin infusion, this was to be given after the completion of olaratumab infusion, not before the olaratumab infusion. This premedication may have been administered within the hour that followed completion of the olaratumab infusion)."

"All doxorubicin doses administered (…for both arms of the Phase 2 portion) to patients in Cycle 5 through Cycle 8 were preceded by administration of dexrazoxane….Dexrazoxane treatment … may have been administered within the hour that followed completion of the olaratumab infusion" (CSR)

In Table 61, administration costs for OlaDox and comparators used in the company's and our base-case analyses are presented. .

Table 61: Administration costs for OlaDox and comparators

	Lilly		ERG	
Regimen	**Cost per administration**	**Delivery code**	**Cost per administration**	**Delivery code**
Ola + Dox, day 1	£329.32	SB13Z (Daycase); NHSRC [35])	£407[1]	SB14Z (daycase)
Ola + Dox + Dex, day 1	£329.32	SB13Z (Daycase); NHSRC [35])	£407[1]	SB14Z (daycase)
Ola, day 1	£185.53	SB12Z (outpatient); NHSRC[35])	£337[1]	SB13Z (daycase)
Ola, day 8	£204.47	SB15Z (outpatient); NHSRC[35])	£361[1]	SB15Z (daycase)
Dox, day 1	£185.53	SB12Z (outpatient); NHSRC[35])	£253[1]	SB12Z (daycase)
Dox + Dex, Day 1	£185.53	SB12Z (outpatient); NHSRC[35])	£253[1]	SB12Z (daycase)
IfoDox, per cycle	£1,781.86*	DZ17V (EI) (Respiratory Neoplasms without Interventions, with CC Score 0-3)	£792	Assumed £264/day for 3 days; based on Guest et al. (2013);[28] converted to GBP and inflated to the cost year 2016 using HCHS indices.

Notes: *Calculated as National average Unit cost of £1518.3 divided by average length of stay 2.56 days = £594 /day x 3 days; [1]https://www.gov.uk/government/publications/nhs-reference-costs-2014-to-2015

Source: Eli Lilly submission, Table 62, p.183, CS

The company assumed that the cost of premedications for Ola and Dox would be included in the delivery HRG. Therefore, a separate cost is not included in the model. The ERG believes that this assumption reflects UK practice.

With regard to the duration of drug administration, the company did not justify the choice of HRG codes for delivery of different regimes (Table 62). We were advised by our clinical expert that OlaDox administration (including premedication for Ola and Dox) may take 2.5-3 hours, and administration of Ola monotherapy (with premedication for Ola) may take up to 90 minutes. Therefore, using SB14Z delivery code for costing OlaDox administration, and SB13Z for administration of Ola would be more appropriate (see Table 62for currency description and explanation).

Table 62: Relevant HRG codes

Currency	Currency Description	Explanation
SB12Z	Deliver Simple Parenteral Chemotherapy at First Attendance	Overall time of 30 minutes nurse time and 30 to 60 minutes[1]
SB13Z	Deliver more Complex Parenteral Chemotherapy at First Attendance	Overall time of 60 minutes nurse time and up to 120 minutes[1]
SB14Z	Deliver Complex Chemotherapy, including Prolonged Infusional Treatment, at First Attendance	Overall time of 60 minutes nurse time and over two hours[1]
SB15Z	Deliver Subsequent Elements of a Chemotherapy Cycle	Delivery of any pattern of outpatient chemotherapy regimen, other than the first attendance, i.e. day 8 of a day 1 and 8 regimen or days 8 and 15 of a day 1, 8 and 15 regimen.[1]

Key: 1, Guidance for the collection of the reference costs (item 161, Table 9, pp. 37-38): https://www.gov.uk/government/publications/nhs-reference-costs-collection-guidance-for-2014-to-2015

Source: https://www.gov.uk/government/publications/nhs-reference-costs-2014-to-2015

We were also advised by our clinical expert that all therapies considered in the submission, except IfoDox, are administered in daycase settings. Therefore, we estimated the cost of drug administration based on this advice.

Lilly costed administration of IfoDox using the code, DZ17V (Respiratory Neoplasms without Interventions, with CC Score 0-3), which is not directly relevant to the disease in question.

Hospitalisation cost for chemotherapy infusion (per day) used in the analysis of Dox/Ifo compared to Trabectedin by Guest et al.(2013)[28] was €213 in Spain, €238 in Italy and €288 in Sweden in 2010/2011 prices. In our main analysis, we assumed hospitalisation costs for IfoDox infusion based on this study.

Using the corrected HRG codes and unit costs increases the base-case ICER for OlaDox comparison against Dox from [redacted] to [redacted] per QALY, and for the comparison against IfoDox from [redacted] to [redacted] per QALY. This constitutes item 8 of the PenTAG base case (Table 70).

5.3.7.3 Cost of post-progression treatment

In the JGDG study, patients with advanced STS received up to 4 lines of systemic anticancer therapy after the study treatments detailed in Table 63 along with the estimates from our clinical expert.

Table 63: The number of subsequent lines of therapy in OlaDox and Dox arms

Number of subsequent lines	OlaDox Arm (N = 66)	Dox Arm (N = 67)	ERG
Any treatment	44 (66.7)	46 (68.7)	50%
1 regimen	18 (27.3)	20 (29.9)	30%
2 regimens	12 (18.2)	14 (20.9)	15%
3 regimens	9 (13.6)	6 (9.0)	<5%
4 regimens	1 (1.5)	2 (3.0)	<5%
> 4 regimens	4 (6.1)	4 (6.0)	<5%

Subsequent therapies used in JGDG most frequently are listed in Table 64.

Table 64: Systemic anticancer therapy received after study treatment in study JGDG

Therapy	OlaDox Arm (N = 66)	Dox Arm (N = 67)	ERG
Ola	0	30 (44.8)	
Gem	15 (22.7)	11 (16.4)	10% in combination with Docetaxel
Pazopanib	15 (22.7)	10 (14.9)	<5%
Docetaxel	14 (21.2)	8 (11.9)	10% in combination with Gem
Dacarbazine	12 (18.2)	8 (11.9)	<2%
Trabectedin	11 (16.7)	3 (4.5)	25%
Investigational drug	8 (12.1)	2 (3.0)	5%
Ifo	8 (12.1)	8 (11.9)	10%

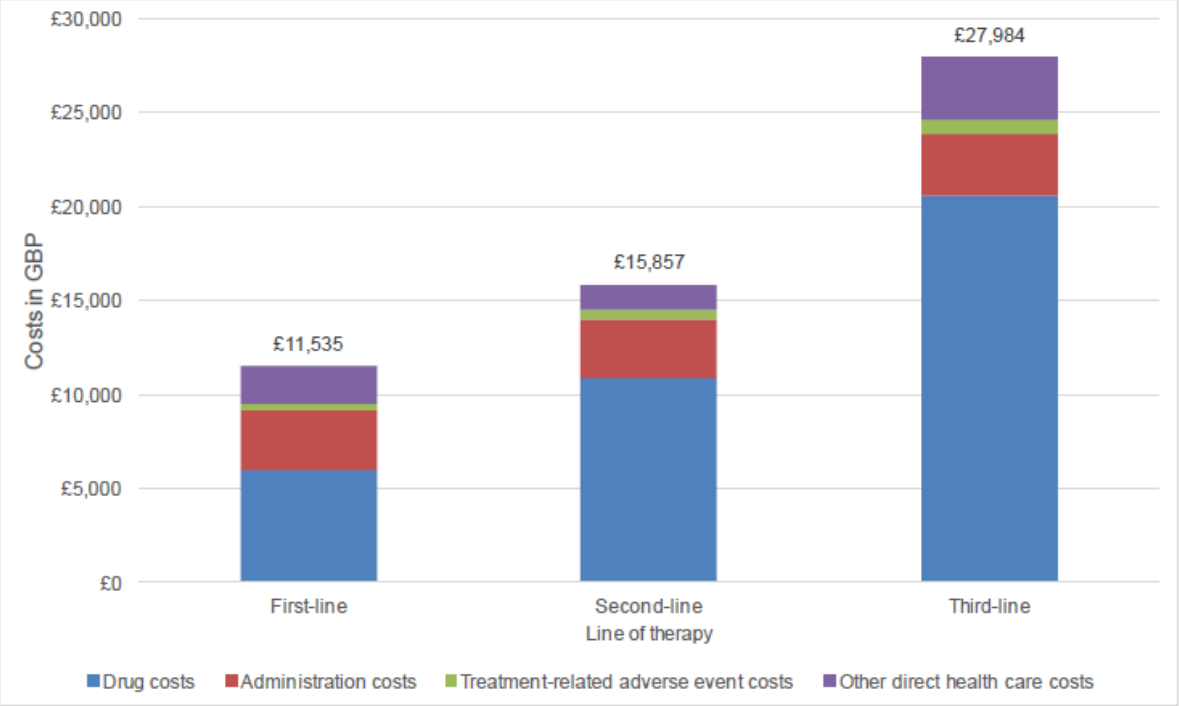
Note: based on Table 51, p. 156, Eli Lilly's submission and our clinical expert's opinion

For simplicity, in the base case, the model assumes that total post-progression cost is independent of survival post-progression, i.e., "cost is identical in both arms." (Eli Lilly submission, p. 190)

Table 34 presents a summary of the total costs of active therapy after first progression. Our clinical experts advised us that the cost of subsequent anticancer treatment is likely to be higher in patients who survive longer; and that the therapy is not given to patients with progressive disease continuously, stopping if prognosis is less than 3 months.

According to the results of Lilly observational study reported by Mytelka et al. (2016)[36], the drug costs increase substantially with the line of treatment (see Figure 26). Therefore, the company's assumption made in the base-case (that the cost of subsequent treatment does not depend on survival (see Table 34, p 111), are not supported by the results of the Eli Lilly observational study.

Figure 26: Average Per-Patient Health Care Costs Related to Advanced STS by Line of Therapy, Overall and by Component



Key: GBP, Great Britain Pounds; STS, soft tissue sarcoma.
Source: Mytelka et al. (poster) reporting the results of Lilly observational study

The results of Lilly's observational study and the opinion of our clinical experts suggest that the assumption of the cost of subsequent treatment increasing with the mean time from progression to death is reasonable. Assuming that if the mean time from progression to death is doubled, the treatment cost is multiplied by 1.5 (which, we think, is a conservative assumption) changes the ICER for the main comparison from [redacted] to [redacted] per QALY and from [redacted] to [redacted] per QALY for the comparison with IfoDox; assuming that if the mean time from progression to death is doubled, the total cost is also doubled, increases the ICER further to [redacted] and to [redacted] per QALY, respectively.

Since there is a high uncertainty associated with the treatment costs post-progression, we explore the effect of these assumptions in sensitivity analyses.

5.3.7.4 Disease monitoring

5.3.7.4.1 Results from Eli Lilly observational study

Health care resource use and costs from Lilly study were reported by Mytelka et al. (2016)[36] (see Table 65 and Figure 26).

Table 65: Results from Lilly observational study

Resource	% patients	Mean (SD)
Outpatient visits	Nearly all	26.4 (48.4)
Outpatient palliative care visits	24%	5.2 (6.0)
Outpatient nurse visits	49%	5.7[3.6]
ED visits	18%	3.2 [2.0]
Inpatient stays	17%	4.1 (4.9)

5.3.7.4.2 The frequency of cardiac monitoring

In the model, it is assumed that all patients have ECHO and MUGA scans on the schedule shown in Table 31. We were advised by our clinical expert that in UK practice these tests are performed at baseline only unless the patient is symptomatic or has risk factors. Correction of this assumption has a negligible effect on the base-case ICER.

5.3.7.4.3 Follow-up visits and imaging costs

The ERG is satisfied with the estimates of the frequencies of follow-ups and tests used in the company's analysis (Table 66).

Table 66: Frequency and resource use for regular follow-up visits and imaging

Frequency of follow-up visits (number of months between each visit)	***Value***
0-5 years	3
5-7 years	6
After 7 years	12
Resource use for each regular follow-up visit	***% patients***
Outpatient visit and physical examination	100%
Computerised tomography scan	92%
Positron emission tomography	9%
Magnetic resonance imaging	14%

5.3.7.4.4 Adverse reaction unit costs and resource use

It was observed that patients treated with OlaDox experienced an increase in the frequency of side effects, and it was reflected in the economic model. We are generally satisfied with the approach taken by Lilly to costing of treatment associated with AEs. However, we identified some inconsistencies in unit costs reported in the submission and in a referenced

source. Nonetheless, we pursue this matter no further, because we find that ICERs change only incrementally when we use the updated unit costs.

5.3.7.5 Health-state costs and resource use

Costs that are associated with managing patients with STS and that are not specifically related to systemic therapy are listed in Table 67. Since JGDG was primarily a US study and resource use would not be generalisable to the UK, the model uses data from the Eli Lilly observational study in the base-case analysis. Resource use data collected included outpatient visits, accident and emergency visits, inpatient hospitalisations, radiotherapy, inpatient long-term care facility, hospice, pain control, supportive drugs, blood transfusions, oxygen, nutritional support and diagnostics. Imaging data were collected but were not included in the health state costs as the cost of imaging is modelled separately in the routine monitoring costs. Total costs after discontinuation of 1st-line therapy include other direct health care costs from end of 1st-line treatment to death/end of follow-up for patients who received no further active systemic treatment after 1st-line therapy, as well as other direct health care costs from start of second-line treatment to death/end of follow-up for those who did receive further lines of treatment.

The same average weekly health state costs are applied to both treatment arms in the model for the progression-free (£131, Lilly Obs study 2016, Lilly data on file 10, 2016) and progressed health states (£35, Lilly Obs study 2016, Lilly data on file 10, 2016[34]), i.e., it is assumed that there are no differences in resource-use not specifically related to systemic therapy.

Table 67: Health state costs in the economic model

Health state costs	Lilly observational study , Lilly data on file 10, first-line analysis	Amdahl et al. (2014)[3]	TA185	Delea et al. (2014) [4]
Progression-free	£131/week	£92/month	£85.96/month	CA$ 213/month
Progressed (excluding radiotherapy/surgery costs)	£35/week	£185/month	£171.91/month	CA$ 426/month

According to model predictions, undiscounted disease management costs per patient on Dox in pre- and post-progression were £4,899 and £4,930, respectively (discounted costs from the model are shown in Table 68). In the Eli Lilly study, however, health-care costs (Figure 26) in 3rd line patients were substantially higher than in 1st line patients.

Table 68: Mean total expected lifetime costs, per patient (discounted)

Line of treatment/Resource	OlaDox, £	Dox, £
1st-line		
Cardiac monitoring during and after treatment	[redacted]	[redacted]
Regular follow-up visits and imaging	[redacted]	[redacted]
Other direct costs	[redacted]	[redacted]
Total:	[redacted]	[redacted]
Subsequent lines of active systemic treatment		
Direct healthcare costs	[redacted]	[redacted]
Regular follow-up visits and imaging	[redacted]	[redacted]
Other direct costs	[redacted]	[redacted]
Total:	[redacted]	[redacted]

Source: Company's model

In TA185,[1] medical management costs from Judson et al.[26] were used assuming £85.96 and £171.91 per month in pre-and post-progression states, respectively. Amdahl et al. (2014) [3] used similar costs, with higher costs incurred in post-progression state (Table 67). A Canadian study by Delea et al. (2014)[4] also reports higher health-state costs in post-progression. Using the estimates from Amdahl et al. or TA185,[1] however, has a negligible effect on the base-case ICER. Therefore, we are not pursuing this further.

5.3.7.6 Post-study treatment radiotherapy and surgery in JGDG trial

Eli Lilly reports that 18.2% and 7.5% of patients in OlaDox and Dox arms, respectively, received post-study radiotherapy treatment; 7.5% of patients in OlaDox arm underwent surgery, while only 1.5% of patients in Dox arm had a resection. The company argues that "the modest imbalances between the arms are unlikely to significantly influence the difference in OS observed in the study" (Eli Lilly submission, Appendix 28, p.220).

Radiotherapy and surgery resource use was based on the Lilly's observational study.

Since surgery with palliative intent may bring some survival benefit to patients and improve their quality of life, the ERG believes that, for internal consistency, the costs of these procedures should be included.

Inclusion of radiotherapy and surgery costs increases the base-case ICER for OlaDox vs. Dox from [redacted] to [redacted] per QALY, and from [redacted] to [redacted] per QALY for the comparison against IfoDox. As the patient population in JGDG study was relatively small, there might be an issue of generalisability of the observed differences in the frequencies of radiotherapy and surgery in patients from different study arm post-progression. Therefore, we do not include these costs in our base case.

5.3.8 Checking wiring of the company's model

The wiring of the model was checked in the following way:

- We checked the key formulae in the model.
- We checked that the model outputs were correct when input parameters were set to extreme values.

We did notice some minor inconsistencies between the model and the company's written submission pertaining to the Bayesian fractional polynomial coefficients. The OS NMA estimated parameters differed slightly. The parameters in brackets in Table 69 are the values in the Excel model (3 significant figures), the others are the numbers in the report.

Table 69: Inconsistencies in fractional polynomial coefficients

Parameter	OlaDox	IfoDox
Scale	[redacted]	[redacted]
Shape 1	[redacted]	[redacted]

5.4 Exploratory and sensitivity analyses undertaken by the ERG

5.4.1 Derivation of PenTAG base case

In this section we derive the ERG's base case (Table 70). The impacts of the individual components of our base case on cost-effectiveness are shown, as well as selected combinations of components and finally the base case, which is composed of all components.

Table 70: Derivation of PenTAG base case ICERs (£ per QALY)

					OlaDox vs. Dox	IfoDox
		PenTAG's assumption in the base case	**Lilly's base case**	**Reference**	[redacted]	[redacted]
1	Parametric survival function for OS	Log-normal	Gamma	Section 5.3.5.1.2, p128.	[redacted]	NA as the model uses a fractional polynomial function for the indirect comparison
2	Coefficients of fractional polynomials estimated in NMA	Mean values	Median values	Section 5.3.5.2 p137.	NA	[redacted]
3	Patients' mean weight	82.5 kg	77.3 kg	Section 5.3.7.1, p145	[redacted]	[redacted]
4	Ifo prices	£66.08 and £130.04 for 1g and	£91.32 and £179.88 for 1g and 2g	Section 5.3.7.1.2. p146	[redacted]	[redacted]

		2g vials, respectively	vials, respectively			
5	Mesna prices	£9.77 and £3.95 for 1000mg and 400mg vial, respectively	£29.41, and £13.41 for 1000mg vial and 400mg vial, respectively	Section 5.3.7.1.2. p148	[redacted]	[redacted]
6	Availability of vial sizes for Ola	Only 500 mg vial available	Both vial sizes, 190 and 500 mg, are available	Section 5.3.7.1.5, p148.	[redacted]	[redacted]
7	HRG codes and unit costs	Corrected		Section 5.3.7.2, p149.	[redacted]	[redacted]
Overall: 1+2+3+4+5+6+7	PenTAG base case				[redacted]	[redacted]

5.4.2 Sensitivity analyses

Results of sensitivity analyses conducted by the ERG summarised in Table 71.

Table 71: Sensitivity analyses

	OlaDox vs. Dox	OlaDox vs. IfoDox	
Utility values from the Amdahl and Delea studies	[redacted]		[redacted]
Weibull OS	[redacted]	NA	
Gompertz OS	[redacted]	NA	
Treatment costs post-progression (2)	[redacted]		[redacted]
Treatment costs post-progression (1.5)	[redacted]		[redacted]
PenTAG base case	[redacted]		[redacted]
Patients mean weight	[redacted]		[redacted]
Availability of both vial sizes	[redacted]		[redacted]
Gamma OS	[redacted]	NA	
Cost of the Ola monotherapy in the Dox arm	[redacted]	NA	

The sensitivity analyses are applied individually to the ERG's base case. They are as follows:

- The utility values taken from the Amdahl and Delea studies are 0.674 and 0.349 for PFS and PPS respectively from Table 60 (p 142)
- Weibull, Gompertz, and gamma OS curves change the ERG's base case from the log-normal OS curve to the respective choice.
- Two scenario analyses assuming that treatment costs in PD depend on survival, and:

    - if the mean time from progression to death is doubled, the treatment cost is multiplied by 1.5;
    - if the mean time from progression to death is doubled, the total cost is also doubled.
- Patients' mean weight adjusts the mean weight of patients back to the Lilly base case of 77.3kgs.
- Vial size sensitivity analysis incorporates the availability of both vial sizes (190mg, 500mg), as per the Lilly's base case.
- Costing the Ola monotherapy in the Dox arm. This is done by multiplying the fraction of Dox patients in the JGDG study who switched to Ola monotherapy by their mean number of infusions (10.6) and costing the infusions and administrations as usual. Administrations are assumed to be Ola monotherapy infusions for costing purposes.

5.4.3 Uncertainty analysis

We conducted probabilistic sensitivity analyses for the PenTAG base case. The results for both comparisons, mean probabilistic ICERs and probabilities of cost-effectiveness of OlaDox at different thresholds are presented below.

5.4.3.1 OlaDox vs. Dox

The mean probabilistic ICER for OlaDox and Dox comparison is [redacted] per QALY gained. The probability of OlaDox being cost-effective at £20,000 and £30,000 per QALY gained is [redacted] At the threshold of £50,000 per QALY, OlaDox is cost-effective with the probability of [redacted]

5.4.3.2 OlaDox vs. IfoDox

The mean probabilistic ICER for OlaDox vs. IfoDox is [redacted] per QALY gained. The probabilities of OlaDox being cost-effective at £20,000 and £30,000 per QALY gained are [redacted] and [redacted] respectively.

5.5 Conclusions of the cost-effectiveness section

Based on the evidence submitted by Ely Lilly and the ERG's review, OlaDox is not cost-effective at the NICE threshold of £30,000 per QALY.

6. End of life

The company argues that their presented evidence supports inclusion into NICE's end of life category; that the life expectancy for the patient population is under 24 months with the comparator(s) and that there is sufficient evidence that the intervention adds at least 3 months additional survival. The company cites the Tap paper (which further references two studies)[48, 49] for these criteria; median OS for the Dox population was 12-16 months. The JGDG study reports a median OS of 14.7 months in the Dox arm and a median OS improvement under OlaDox versus Dox of 11.8 months.

The ERG however notes that using the median is an incorrect interpretation of the NICE criteria and it is inconsistent with their modelling approach. The NICE criterion above indicates that life expectancy, i.e. *mean* not median survival, should be under 24 months. In their base-case their comparators Dox/IfoDox have an (undiscounted) mean life expectancy of 2.32/2.67 years respectively (calculated from Eli Lilly model). As such, their base case would not qualify for the EoL category and the standard £20,000-£30,000 per QALY ICER threshold would be applicable.

The ERG base case, however, produces significantly lower survival outcomes for OlaDox vs. Dox arm; the mean undiscounted life expectancy for Dox falls to 1.83 years (IfoDox unchanged). Based on the above criteria, the OlaDox vs. Dox arm would qualify for end of life under the ERG's base case.

7. Overall conclusions

The company submission identified a single RCT trial (JGDG) which was a phase 2, open-label trial with a small population. However, it was well conducted.

The primary efficacy endpoint for the RCT was an ITT comparison of progression free survival for patients randomised to OlaDox versus patients randomised to Dox. An improvement was demonstrated which reached the planned statistical significance level of 0.1999. Statistical significance was also reached for overall survival, indicating OlaDox to be superior to Dox monotherapy.

With regard to the network meta-analysis, the main analyses included data from patients who received OlaDox as a first line treatment only, and produced an HR for overall survival that was significantly lower for OlaDox compared with Dox monotherapy and for one of the four IfoDox treatments. For two of the IfoDox treatments, there was a borderline significant reduction in the HR for OS and for the fourth IfoDox treatment there was no significant difference in HR.

Estimates of cost-effectiveness of OlaDox therapy for mSTS patients currently suggest poor value for money at a willingness-to-pay threshold of £30,000. Our results indicate that the cost of drug acquisition drives this poor value for money.

The ERG considers that all changes to Lilly's base case made by the ERG cannot be regarded as matters of opinion. They relate, among other things, to overall survival of mSTS patients and costing of drug acquisition and administration.

In summary, there is a potential for clinical benefit from olaratumab + doxorubicin therapy for patients with advanced soft tissue sarcoma but cost of administering this treatment is substantial.

7.1 Implications for research

Further research is needed to remove the potential bias associated with open label studies and for treatment dose to be in keeping with UK clinical practice. Therefore, ideally double-blind studies suitably powered to detect clinically meaningful improvements in survival between OlaDox and Dox or IfoDox would be useful. The statistical analyses planned for the research should account for the likelihood of treatment switching, with adequate data collection to support multiple plausible statistical models. Longer-term follow-up of overall survival in patients on OlaDox would be helpful to inform the cost-effectiveness model and accurate estimates of the utilities of patients in pre- and post-progression would allow less biased estimates of cost-effectiveness.

.

7.1.1 Health related quality of life

This research should also collect HRQoL data measured using a generic (as opposed to condition-specific) and validated instrument, which allow outcomes to be valued using preferences from the general public (preferably EQ-5D) and is preferred for economic analyses. Significant efforts should be made to collect HRQoL data across all patients and across all time points to reflect the full range of quality of life experienced by patients.

References

1. Excellence NIoHaC. Trabectedin for the treatment of advanced soft tissue sarcoma NICE website: NICE; 2010 [cited 2017 6 March]. Technology appraisal guidance [TA185]]. Available from: https://www.nice.org.uk/Guidance/ta185.

2. Reichardt P, Leahy M, Garcia Del Muro X, Ferrari S, Martin J, Gelderblon H. Quality of Life and Utility in Patients with Metastatic Soft Tissue and Bone Sarcoma: The Sarcoma Treatment and Burden of Illness in North America and Europe (SABINE) Study. Sarcoma. 2012;2012:740279.

3. Amdahl J, Manson SC, Isbell R, Chit A, Diaz J, Lewis L, et al. Cost-effectiveness of pazopanib in advanced soft tissue sarcoma in the United Kingdom. Sarcoma. 2014;2014.

4. Delea T, Amdahl J, Nakhaipour H, Manson S, Wang A, Fedor N, et al. Cost-effectiveness of pazopanib in advanced soft-tissue sarcoma in Canada. Current Oncology. 2014;21(6):e748.

5. Van Glabbeke M, Van Oosterom A, Oosterhuis J, Mouridsen H, Crowther D, Somers R, et al. Prognostic factors for the outcome of chemotherapy in advanced soft tissue sarcoma: an analysis of 2,185 patients treated with anthracycline-containing first-line regimens—a European Organization for Research and Treatment of Cancer Soft Tissue and Bone Sarcoma Group Study. Journal of Clinical Oncology. 1999;17(1):150-.

6. Pike J, Clarkson PW, Masri BA. Soft tissue sarcomas of the extremities: How to stay out of trouble. BCMJ. 2008;50(6):310-8.

7. Fletcher CD, Unni KK, Mertens F. Pathology and genetics of tumours of soft tissue and bone: Iarc; 2002.

8. Soft tissue sarcoma incidence statistics [Internet]. 2010 [cited 09/02/2017]. Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/soft-tissue-sarcoma/incidence.

9. Sarcoma, Soft Tissue [Internet]. 2017 [cited 28022017]. Available from: http://www.cancer.net/cancer-types/sarcoma-soft-tissue/introduction.

10. Tap WD, Jones RL, Van Tine BA, Chmielowski B, Elias AD, Adkins D, et al. Olaratumab and doxorubicin versus doxorubicin alone for treatment of soft-tissue sarcoma: an open-label phase 1b and randomised phase 2 trial. The Lancet. 2016 10/06/2016. Epub 10/06/2016.

11. Dennis NF, M. Lawrence, Gill. NCIN Soft Tissue Sarcoma Incidence and Survival Tumours Diagnosed in England Between 1985 and 2009. The university of Birmingham, 2012 2012. Report No.

12. Vraa S, Keller J, Nielsen O, Sneppen O, Jurik A, Jensen O. Prognostic factors in soft tissue sarcomas: the Aarhus experience. European Journal of Cancer. 1998;34(12):1876-82.

13. Trovik LH, Ovrebo K, Almquist M, Haugland HK, Rissler P, Eide J, et al. Adjuvant radiotherapy in retroperitoneal sarcomas. A Scandinavian Sarcoma Group study of 97 patients. Acta Oncologica. 2014;53(9):1165-72.

14. Gough NJ, Smith C, Ross JR, Riley J, Judson I. Symptom burden, survival and palliative care in advanced soft tissue sarcoma. Sarcoma. 2011;2011.

15. National Institute for Health and Care Excellence. Olaratumab in combination with doxorubicin for treating advanced soft tissue sarcoma. 2016.

16. Tap WD, Jones RL, Van Tine BA, Chmielowski B, Elias AD, Adkins D, et al. Olaratumab and doxorubicin versus doxorubicin alone for treatment of soft-tissue sarcoma: an open-label phase 1b and randomised phase 2 trial. The Lancet. 2016.

17. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. 2013.

18. Centre for Reviews and Dissemination. Systematic reviews. University of York: 2009.

19. Santoro A, Tursz T, Mouridsen H, Verweij J, Steward W, Somers R, et al. Doxorubicin versus CYVADIC versus doxorubicin plus ifosfamide in first-line treatment of advanced soft tissue sarcomas: a randomized study of the European Organization for Research and Treatment of Cancer Soft Tissue and Bone Sarcoma Group. Journal of Clinical Oncology. 1995;13(7):1537-45.

20. Le Cesne A, Judson I, Crowther D, Rodenhuis S, Keizer H, Van Hoesel Q, et al. Randomized phase III study comparing conventional-dose doxorubicin plus ifosfamide versus high-dose doxorubicin plus ifosfamide plus recombinant human granulocyte-macrophage colony-stimulating factor in advanced soft tissue sarcomas: a trial of the European Organization for Research and Treatment of Cancer/Soft Tissue and Bone Sarcoma Group. Journal of Clinical Oncology. 2000;18(14):2676-84.

21. Judson I, Verweij J, Gelderblom H, Hartmann JT, Schöffski P, Blay J-Y, et al. Doxorubicin alone versus intensified doxorubicin plus ifosfamide for first-line treatment of advanced or metastatic soft-tissue sarcoma: a randomised controlled phase 3 trial. The lancet oncology. 2014;15(4):415-23.

22. Seddon BM, Whelan J, Strauss SJ, Leahy MG, Woll PJ, Cowie F, et al. GeDDiS: a prospective randomised controlled phase III trial of gemcitabine and docetaxel compared with doxorubicin as first-line treatment in previously untreated advanced unresectable or metastatic soft tissue sarcomas (EudraCT 2009-014907-29). American Society of Clinical Oncology; 2015.

23. Maurel J, López-Pousa A, de las Peñas R, Fra J, Martín J, Cruz J, et al. Efficacy of sequential high-dose doxorubicin and ifosfamide compared with standard-dose doxorubicin in patients with advanced soft tissue sarcoma: an open-label randomized phase II study of the Spanish group for research on sarcomas. Journal of Clinical Oncology. 2009;27(11):1893-8.

24. Guyot P, Ades A, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC medical research methodology. 2012;12(1):9.

25. Therneau TM, Grambsch PM. Modeling survival data: extending the Cox model: Springer Science & Business Media; 2013.

26. Judson I, Verweij J, Gelderblom H, Hartmann JT, Schoffski P, Blay JY, et al. Doxorubicin alone versus intensified doxorubicin plus ifosfamide for first-line treatment of advanced or metastatic soft-tissue sarcoma: a randomised controlled phase 3 trial. Lancet Oncol. 2014 Apr;15(4):415-23. PubMed PMID: 24618336. Epub 2014/03/13. eng.

27. Jansen JP. Network meta-analysis of survival data with fractional polynomials. BMC medical research methodology. 2011;11(1):61.

28. Guest JF, Panca M, Sladkevicius E, Gough N, Linch M. Cost effectiveness of first-line treatment with doxorubicin/ifosfamide compared to trabectedin monotherapy in the management of advanced soft tissue sarcoma in Italy, Spain, and Sweden. Sarcoma. 2013;2013.

29. Shingler SL, Swinburn P, Lloyd A, Diaz J, Isbell R, Manson S, et al. Elicitation of health state utilities in soft tissue sarcoma. Quality of Life Research. 2013;22(7):1697-706.

30. Nemeth B, Vincziczki Á. The Role of Half-Cycle Correction in the Models Used for Health Technology Assessment. Value in Health. 2013;16(7):A592-A3.

31. Centre HaSCI. Health Survey for England 2013-2014. [2015]. Available from: http://www.medicines.org.uk/emc/medicine/20555.

32. Erkisi M, Erkurt E, Ozbarlas S, Burgut R, Doran F, Seyrek E. The use of recombinant human granulocyte colony-stimulating factor in combination with single or fractionated doses of ifosfamide and doxorubicin in patients with advanced soft tissue sarcoma. Journal of chemotherapy. 1996;8(3):224-8.

33. Reichardt P, Leahy M, Garcia Del Muro X, Ferrari S, Martin J, Gelderblom H, et al. Quality of Life and Utility in Patients with Metastatic Soft Tissue and Bone Sarcoma: The Sarcoma Treatment and Burden of Illness in North America and Europe (SABINE) Study. Sarcoma. 2012;2012:740279. PubMed PMID: 22550425. Pubmed Central PMCID: 3329673.

34. Eli Lilly and Co Pvt Ltd. Lilly data on file 10: Lilly Observational study - Health State Costs. 2016.

35. Department of Health (DoH). NHS Reference Costs 2014-2015 2015 [updated 27th November 2015; cited 2016 January 2016]. Available from: https://www.gov.uk/government/publications/nhs-reference-costs-2014-to-2015.

36. Mytelka D, Lorenzo M, Stafkey-Mailey D, D'yachkova Y, Nagar S, Candrilli S, et al. REAL-WORLD TREATMENT PATTERNS AND OUTCOMES FOR PATIENTS WITH ADVANCED SOFT TISSUE SARCOMA RECEIVING SYSTEMIC THERAPY IN THE UNITED KINGDOM. The Ispor Scientific Presentations Database: 2016.

37. Dangoor A, Seddon B, Gerrand C, Grimer R, Whelan J, Judson I. UK guidelines for the management of soft tissue sarcomas. Clinical Sarcoma Research. 2016;6(1):20.

38. Eli Lilly and Co Pvt Ltd. Lilly Data on file 3 : JGDG- Clinical Study Report. A Phase 1b/2 Randomised Phase 2 Study Evaluating the Efficacy of Doxorubicin With or Without a Human Anti-PDGFRα Monoclonal Antibody (IMC-3G3) in the Treatment of Advanced Soft Tissue Sarcoma.

39. Karavasilis V, Seddon BM, Ashley S, Al‐Muderis O, Fisher C, Judson I. Significant clinical benefit of first‐line palliative chemotherapy in advanced soft‐tissue sarcoma. Cancer. 2008;112(7):1585-91.

40. Gutierrez JC, Perez EA, Franceschi D, Moffat FL, Livingstone AS, Koniaris LG. Outcomes for soft-tissue sarcoma in 8249 cases from a large state cancer registry. Journal of surgical research. 2007;141(1):105-14.

41. Maretty-Nielsen K, Aggerholm-Pedersen N, Keller J, Safwat A, Baerentzen S, AB P. Relative mortality in soft tissue sarcoma patients: a Danish population-based cohort study. BMC Cancer. 2014;14(682).

42. Latimer NR, KR A. NICE DSU Technical Support Document 16: Adjusting survival time estimates in the presence of treatment switching. In: Excellence NIfHaC, editor. London2014.

43. Latimer Nicholas R, Abrams Keith R, Amonkar Mayur M, Stapelkamp Ceilidh, Suzanne SR. Adjusting for the confounding effects of treatment switching—the BREAK-3 trial: dabrafenib versus dacarbazine. Oncologist. 2015;20(7):798–805.

44. Nafees B, Stafford M, Gavriel S, Bhalla S, J W. Health state utilities for non small cell lung cancer. Health Qual Life Outcomes. 2008;21(84).

45. van der Graaf WT, Blay JY, Chawla SP, Kim DW, Bui-Nguyen B, Casali PG, et al. Pazopanib for metastatic soft-tissue sarcoma (PALETTE): a randomised, double-blind, placebo-controlled phase 3 trial. Lancet. 2012;19(379(9829)):1879-86.

46. Sacco Joseph J, Botten Joanne, Macbeth Fergus, Bagust Adrian, Peter C. The average body surface area of adult cancer patients in the UK: a multicentre retrospective study. PLoS One. 2010;28(5(1)).

47. Excellence NIfHaC. British National Formulary 2017 [cited 2017 6 March]. British National Formulary database]. Available from: https://www.evidence.nhs.uk/formulary/bnf/current.

48. Judson I, Verweij J, Gelderblom H, et al. Doxorubicin alone versus intensified doxorubicin plus ifosfamide for first-line treatment of advanced or metastatic soft tissue sarcoma: a randomised controlled phase 3 trial. Lancet Oncol. 2014; 15: 415–23.

49. Ryan CW, Schoff ski P, Merimsky O, et al. PICASSO 3: a phase 3 international randomized double blind placebo-controlled study of doxorubicin plus palifosfamide vs. doxorubicin plus placebo for patients in first-line for metastatic soft tissue sarcoma. Eur J Cancer 2013; 49 (suppl 2): S876 (abstr 3802).

List of appendices