CONFIDENTIAL UNTIL PUBLISHED

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

Fulvestrant for untreated hormone-receptor positive locally advanced or metastatic breast cancer

Produced by Southampton Health Technology Assessments Centre

Authors Dr. Keith Cooper, Senior Research Fellow (Health Economics), SHTAC
Ms. Neelam Kalita, Research Fellow (Health Economics), SHTAC
Ms. Petra Harris, Research Fellow, SHTAC
Mr. Olu Onyimadu, Research Fellow (Health Economics), SHTAC
Dr. Wendy Gaisford, Research Fellow, SHTAC
Dr. Joanna Picot, Senior Research Fellow, SHTAC

Correspondence to Dr. J. Picot
Southampton Health Technology Assessments Centre
University of Southampton
First Floor, Epsilon House
Enterprise Road, Southampton Science Park
Southampton SO16 7NS

Date completed 10 July 2017

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 16/54/04.

Declared competing interests of the authors

None from the authors. Professor Robert Coleman has received a lecture honorarium from a national breast cancer meeting (held in 2016) that was indirectly funded by AstraZeneca and Dr Eleni Karapanagiotou has received a travel grant from AstraZeneca.

Acknowledgements

We are very grateful to Professor Robert Coleman, Emeritus Professor of Medical Oncology, University of Sheffield and Dr Eleni Karapanagiotou, Consultant Medical Oncologist, Guy's and St Thomas' NHS Foundation Trust, who offered clinical advice and comments on the draft report. We would also like to thank: Karen Welch, Information Scientist, SHTAC, for appraising the literature search strategies in the company's submission, and for running searches where necessary; and Jonathan Shepherd, Principal Research fellow, for acting as internal editor for the ERG report.

Rider on responsibility for report

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

This report should be referenced as follows:

Cooper, K., Kalita, N., Harris, P., Onyimadu, O., Gaisford, W., Picot, J. Fulvestrant for untreated hormone-receptor positive locally advanced or metastatic breast cancer: A Single Technology Appraisal. Southampton Health Technology Assessments Centre (SHTAC), 2017.

Contributions of authors

Dr. Keith Cooper, critically appraised the health economic systematic review, critically appraised the economic evaluation, and drafted the report; Ms. Neelam Kalita, critically appraised the health economic systematic review, critically appraised the economic evaluation, and drafted the report; Ms. Petra Harris, critically appraised the clinical effectiveness systematic review and drafted the report; Mr. Olu Onyimadu, critically appraised the health economic systematic review, critically appraised the economic evaluation, and drafted the report; Dr. Wendy Gaisford, critically appraised the clinical effectiveness systematic review and drafted the report; Dr. Joanna Picot, critically appraised the clinical effectiveness systematic review, drafted the report, project managed the review and is the project guarantor.

Word count: 43,365

Commercial in confidence (CIC) information in blue
Academic in confidence (AIC) information in yellow.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABC	Advanced breast cancer
AE	Adverse event
AI	Aromatase inhibitor
AIC	Akaike information criteria
BIC	Bayesian information criteria
CBR	Clinical Benefit Rate
CHMP	Committee for Medicinal Products for Human Use
CI	Confidence interval
CR	Complete response
CS	Company submission
CSR	Clinical study report
DoCB	Duration of clinical benefit
DoR	Duration of response
EDoCB	Expected duration of clinical benefit
EMA	European Medicines Agency
EQ-5D	EuroQoL5 Dimensions questionnaire
ER	Oestrogen receptor
ERG	Evidence review group
FACT-B	Functional Assessment of Cancer Therapy – Breast questionnaire
HER2	Human epidermal growth factor receptor 2
HR	Hazard ratio
HR+	Hormone receptor-positive
HRQoL	Health-related quality of life
HRT	Hormone replacement therapy
HTA	Health technology assessment
ICER	Incremental cost-effectiveness ratio
IM	Intramuscular
IPD	Individual patient data
ITT	Intention-to-treat
KM	Kaplan-Meier
MMRM	Repeated measures mixed effects regression models
NHS	National Health Service

NICE	National Institute of Health and Care Excellence
NMA	Network meta-analysis
NR	Not reported
ORR	Objective response rate
OS	Overall survival
PAS	Patient Access Scheme
PD	Progressed disease
PFS	Progression-free survival
PgR	Progesterone Receptor
PR	Partial response
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSS	Personal Social Services
QALY	Quality-adjusted life year
RCT	Randomised controlled trial
RECIST	Response Evaluations Criteria in Solid Tumours
SAE	Serious adverse event
SD	Standard deviation
SE	Standard error
SLR	Systematic literature review
SMC	Scottish Medicines Consortium
SmPC	Summary of Product characteristics
TOI	Trial Outcome Index
TTP	Time to treatment progression

SUMMARY

Scope of the company submission

The company's submission (CS) reflects the scope of the appraisal issued by the National Institute for Health and Care Excellence (NICE). The submission assesses the clinical effectiveness and cost effectiveness of fulvestrant for untreated hormone-receptor positive locally advanced or metastatic breast cancer in comparison to aromatase inhibitors (AIs) (anastrozole and letrozole) or, when these are not tolerated or are contraindicated, tamoxifen.

Summary of submitted clinical effectiveness evidence

The company's systematic review of clinical effectiveness identified two relevant randomised controlled trials (RCTs) of fulvestrant:

- The FIRST trial (phase II, open label, non-inferiority trial) compared fulvestrant (500 mg) versus anastrozole (1 mg) in postmenopausal women with hormone-receptor positive (HR+) advanced breast cancer (ABC) who had either never received endocrine therapy for advanced disease or who had received previous adjuvant endocrine therapy for ABC completed at least 12 months prior to randomisation into the study.
- The FALCON trial (phase III, double blind, superiority trial) compared fulvestrant 500 mg versus anastrozole 1 mg in postmenopausal women with oestrogen receptor positive (ER+) and/or progesterone receptor positive (PgR+) ABC who had not previously been treated with any endocrine therapy.

In these trials anastrozole is considered the standard of care.

There are some differences between the trials:

- in terms of patient inclusion criteria, the chief differences were the requirement in FALCON for all participants to be endocrine therapy naive and also human epidermal growth factor (HER2) negative. In FALCON patients were allowed to have received one line of prior chemotherapy for ABC whereas prior chemotherapy for ABC was not permitted in the FIRST trial.
- in terms of design, the chief differences were that FALCON was a double blind phase III trial, whereas FIRST was open-label phase II trial and the trials had different primary outcomes [clinical benefit rate (CBR) in FIRST and progression-free survival (PFS) in FALCON]

Both trials were judged by the Evidence Review Group (ERG) to be of good methodological quality. The ERG believes that the company has identified all the relevant RCTs of fulvestrant.

There are no head-to-head RCTs of fulvestrant versus tamoxifen or letrozole so the company conducted a Bayesian fixed-effect network meta-analysis (NMA) to perform an indirect treatment comparison. The company's systematic review identified a further four RCTs for inclusion in the NMA initially, of which three compared anastrozole versus tamoxifen (the North American trial; the TARGET trial and a trial by Milla-Santos et al.) and one compared letrozole versus tamoxifen (the PO25 trial). The North American and TARGET studies were prospectively designed to allow for combined data analysis and the combined data are described as NorthAmTarget in the CS. The Milla-Santos trial was subsequently excluded from the NMA as its inclusion led to heterogeneity, used a dose of tamoxifen not recommended by the European Medicines Agency (EMA) and reported the outcomes of interest only for a subset of participants.

The CS summarises the methodological and patient characteristics for all six trials (two for fulvestrant, four for other comparators) that were identified for inclusion in the NMA. Individual patient data (IPD) were available for the two fulvestrant trials and also the combined NorthAmTARGET data set. This enabled the company to select patient data from the FIRST and NorthAmTARGET trials that matched the criteria of the FALCON trial in respect of ER+/PgR+ status and endocrine treatment naive status. Only aggregate data were available for the PO25 study which therefore could not be matched to FALCON. The possible advantages and disadvantages of this matching process were not discussed in the CS. The ERG understands that by matching to FALCON it was possible to exclude participants [redacted] (except for study PO25). Although the ERG has concerns about whether there may be unknown potential disadvantages to this matching approach, the ERG has concluded that these would likely be outweighed by the benefits of reduced heterogeneity in the NMA. The company used appropriate methods to investigate whether there was a constant relative treatment effect over time. The company concluded that methods for NMA that rely on the assumption of proportional hazards were inappropriate and therefore used an alternative method (Ouwens et al.). Fixed-effect NMA results are presented for the outcomes of PFS and overall survival (OS) and these inform the economic model. The company provided reasons for the use of a fixed-effect model and why it was not possible to run the NMA using a random-

effects model. The ERG accepts that with few trials in the network and the features of the modelling methodology the company have used, a random-effects NMA has not been possible but this does leave a concern that uncertainty in the NMA outcomes may be under-represented.

The CS reports the effects of fulvestrant treatment across a range of outcomes relevant to the NICE scope and company decision problem which are summarised below.

PFS (FALCON trial primary outcome) and time to progression (TTP, FIRST secondary outcome) both favoured fulvestrant and in both cases the difference in medians (fulvestrant versus anastrozole) was statistically significant. However, the median PFS in the FALCON trial was only 2.8 months longer in the fulvestrant arm than the anastrozole arm [hazard ratio (HR) = 0.80, 95% confidence interval (CI) 0.64 to 1.0, p = 0.049] which the two clinical experts the ERG consulted did not believe was a clinically significant difference. Median TTP in the FIRST trial was 10.3 months longer in the fulvestrant arm than the anastrozole arm (HR 0.66, 95% CI 0.47 to 0.92, p = 0.01).

Results from the PFS fixed-effect NMA indicated that fulvestrant is statistically significantly better than anastrozole and tamoxifen is statistically significantly worse than anastrozole.

OS is a secondary outcome for the FALCON trial and OS data are immature. A median OS could not be calculated at the time of PFS analysis. There was a slightly lower proportion of deaths in the fulvestrant arm than the anastrozole arm (29% vs 32% respectively) but the difference is not statistically significant. The CS states OS will be re-assessed after a longer follow-up period. In the FIRST trial, OS was added as a secondary outcome after TTP data were analysed. Results of the OS analysis (undertaken after approximately 65% of deaths had occurred) demonstrated a lower proportion of deaths in the fulvestrant arm (61.8% versus 71.8% in the anastrozole arm). The difference in median survival times, 54.1 months with fulvestrant in comparison to 48.4 months with anastrozole, is statistically significant (HR 0.70, 95% CI 0.50 to 0.98, p=0.04) but as the analysis was not originally specified confirmation of the results from the ongoing FALCON trial is needed.

The only statistically significant differences observed in the OS fixed-effect NMA related to the letrozole versus anastrozole comparisons, but the direction of the differences is not consistent.

Subgroup analyses conducted for PFS and OS indicated that across the subgroups tested, the results were consistent with those of the whole study populations of FIRST and FALCON. Consideration of subgroups of people with visceral disease and those with non-visceral disease (if evidence allows) was included in the company's decision problem and the largest numerical difference in the reported hazard ratios for PFS of subgroups was observed for the visceral disease versus no visceral disease at baseline but the company do not discuss this subgroup result in the CS.

CBR was the primary outcome for the FIRST trial and a secondary outcome of the FALCON Trial. CBR was defined in both trials as the proportion of all randomly assigned patients with a best overall response of either complete response (CR), partial response (PR) or stable disease (≥24 weeks). In both trials the CBR numerically favoured fulvestrant, but the difference between fulvestrant and anastrozole was not statistically significant (as the FIRST trial was designed as a non-inferiority trial it was not powered to detect a statistically significant difference in CBR).

Among the other secondary outcomes the results were either similar between treatment arms or favoured the fulvestrant arm but with no statistically significant differences between fulvestrant and anastrozole.

Included among the other secondary outcomes was health-related quality of life (HRQoL) which was only assessed in the FALCON trial, using both the EuroQoL5 Dimensions-3L (EQ-5D-3L) and Functional Assessment of Cancer Therapy – Breast (FACT-B) questionnaires. Data from the EQ-5D-3L was used to inform the economic model. Results from both questionnaires indicated that HRQoL was similar between treatment arms and maintained in both arms during treatment.

Adverse events (AEs) were reported from both the fulvestrant trials. The proportions of AEs and serious adverse events were similar between treatment arms of both trials. The proportion of patients discontinuing study treatment due to an AE was also similar in the fulvestrant and anastrozole treatment groups. Deaths considered to be related to AEs were reported (3% in each arm of the FALCON trial, 3% in the fulvestrant arm and 4.9% in the anastrozole arm of the FIRST trial) but none of these were reported as being causally related to study treatment.

Summary of submitted cost effectiveness evidence

The CS includes:

- A review of submissions made to national reimbursement agencies and technology assessment organisations of treatments for locally advanced or metastatic breast cancer
- An economic evaluation undertaken for the NICE STA process to assess fulvestrant for post-menopausal women with locally advanced or metastatic hormone receptor-positive (HR+) breast cancer who have not received endocrine therapy.

A review was conducted by the company to identify submissions, of locally advanced or metastatic breast cancer, to national reimbursement agencies and technology assessment organisations, including the Canadian Agency for Drugs and Technologies in Health (CADTH), the pan-Canadian Oncology Drug Review, NICE, Pharmaceutical Benefits Advisory Committee (PBAC) and Scottish Medicines Consortium. However, the company only selected submissions published by NICE for review. The company identified 10 technology appraisals relating to advanced or metastatic breast cancer.

Five of these 10 submissions relate to first-line therapy but none relate to the same population as in the current submission. As the company did not search for published cost-effectiveness literature, the ERG completed a search of published cost-effectiveness studies. We identified two studies that compared fulvestrant in patients previously treated with an AI or anti-oestrogen therapy.

The company constructed a cohort partitioned survival model in Microsoft Excel. The model compared first-line treatment with fulvestrant compared to anastrozole, letrozole and tamoxifen for post-menopausal women with HR+ locally advanced or metastatic breast cancer. The model had a lifetime horizon of 30 years, with discounting at 3.5% per annum for costs and benefits, a four-week cycle length and a half-cycle correction. The perspective of the analysis is the National Health Service and Personal Social Services. The model has three health states: PFS, ‘progressed disease’ (PD) and ‘death’.

The model uses clinical effectiveness data from head-to-head trials comparing fulvestrant and anastrozole (FIRST and FALCON). Fulvestrant is compared to letrozole and tamoxifen via an indirect treatment comparison. The model uses parametric survival modelling to fit survival curves using results from the company’s NMA. The company considered that the most

appropriate method for extrapolating best fit to the observed data for PFS was to use a generalised gamma distribution. The Weibull distribution was chosen as the most appropriate method of extrapolating OS. The model derives the proportion of patients in the PD health state as the difference between the PFS and OS curves. Treatment duration was assumed to be until objective disease progression.

Utility estimates were taken from the company's FALCON trial, in which quality of life values from the EQ-5D 3L questionnaire were collected. Fulvestrant is administered intramuscularly into the buttocks in the first month and then monthly thereafter. The recommended dose is two injections (one in each buttock) of 250 mg at a list price of £522.41 for both injections. Comparator treatments consisted of oral treatments taken daily with a cost of between £0.75 and £1.62 per month. The cost of comparator treatments are taken from the pharmaceutical electronic market information tool (eMit) and their doses are as recommended by their Summary of Product Characteristics. Health state costs are based upon those recommended in the NICE Clinical Guidance on ABC. Subsequent therapies were included for second- and third-line treatment.

The results of the economic model are presented as incremental cost effectiveness ratios (ICERs), measured as the incremental cost per quality-adjusted life-year (QALY). For the base case the incremental cost per QALY gained is £34,099 for fulvestrant compared to anastrozole.

Table 1 Base case cost effectiveness results

Treatments	Total costs	Total QALYs	Incremental costs	Incremental QALYs	Incremental ICER (cost per QALY)
Letrozole	£26,221	2.46	--	--	--
Anastrozole	£30,572	2.68	£4,351	0.22	£19,702
Tamoxifen	£32,328	2.47	£1,756	-0.21	Dominated
Fulvestrant	£49,431	3.23	£18,859	0.55	£34,099

This table draws on information presented in Table 30 within the appendix to the company's written response to the clarification questions.

In probabilistic sensitivity analyses, the probability of first-line fulvestrant being cost-effective compared to anastrozole, letrozole and tamoxifen is 1.1% and 26.5% at willingness to pay

thresholds of £20,000 and £30,000 per QALY respectively. At both these willingness to pay thresholds, both letrozole and anastrozole were more cost-effective than fulvestrant.

The company conducted sensitivity analyses and scenario analyses and concluded that the key drivers to the cost-effectiveness results were the OS extrapolation parameters, PFS, utility values and fulvestrant treatment acquisition costs.

Commentary on the robustness of submitted evidence

Strengths

The company's systematic review of clinical effectiveness was of good methodological quality. The ERG does not believe that any key studies of fulvestrant or of potential comparators are missing. Two RCTs provide evidence for the effectiveness of fulvestrant versus anastrozole for people with untreated hormone-receptor positive locally advanced or metastatic breast cancer. Three additional RCTs provide evidence, which is used in an NMA for the outcomes of PFS and OS, for the other comparators of interest, letrozole and tamoxifen.

The model structure is representative of the clinical pathway for patients with advanced or metastatic breast cancer. The company used methods for the economic evaluation that are consistent with NICE methodological guidelines. The company's clinical trial collected EQ-5d 3L HRQoL data.

Weaknesses and Areas of uncertainty

The initial phase II study FIRST demonstrated a clinically significant and statistically significant improvement with fulvestrant in TTP in comparison to anastrozole. The pivotal phase III trial FALCON demonstrated a statistically significant improvement in PFS but the magnitude of the improvement was not as great as that observed in the FIRST study (median TTP 10.3 months longer with fulvestrant versus median PFS 2.8 months longer). Furthermore, clinical advice to the ERG was that a median PFS of 2.8 months longer with fulvestrant would not be considered clinically significant. The median OS in FIRST was almost 6 months longer with fulvestrant versus anastrozole but median OS in the FALCON study has not yet been reached. The ERG is concerned that the OS benefit in FALCON may mirror that of PFS and not be as great as observed in the FIRST study. This has an impact on the cost-effectiveness modelling.

There is no direct evidence comparing fulvestrant to either letrozole or tamoxifen so the company conducted an NMA. For all comparisons in the NMA except fulvestrant versus anastrozole there was only one data set (although in one case this single data set was obtained from two replicate trials). The company were unable to conduct a random-effects meta-analysis so it is possible that uncertainty in the outcomes of the NMA is not adequately represented.

The model results are sensitive to changes in the estimation of overall survival. The OS data from the FALCON trial are immature. There is some uncertainty in what the cost-effectiveness estimates would be if complete FALCON OS were available. The ICERs are likely to be higher when the full results of the FALCON become available.

Summary of additional work undertaken by the ERG

We conducted a number of scenario analyses to examine the robustness of the company's base case economic analyses. These are:

- Scenario 1: Varying the parametric survival distribution for overall survival
- Scenario 2: Varying the treatment effectiveness of fulvestrant by changing the scale parameter of the OS
- Scenario 3: Varying units of resource use and costs associated with progression-free and PD health states
- Scenario 4: Varying the proportion of patients receiving endocrine therapy, chemotherapy and targeted therapy as second-line treatment
- Scenario 5: Exclusion of PO25 trial and Milla-Santos study from the network-meta analysis used to obtain PFS and OS curves, and assuming that anastrozole and letrozole have similar clinical efficacy
- Scenario 6: Change of administration cost for fulvestrant, assuming that all patients receive treatment in an outpatient setting

The ICERs were mostly not particularly sensitive to the selection of the parametric distribution for extrapolating OS curve, with the exception of the Gompertz distribution which produced an ICER of £59,953 per QALY for fulvestrant vs anastrozole. The ERG considered that the Gompertz distribution had a poor fit to the observed data. Changing the treatment effectiveness of fulvestrant by varying the OS scale parameter increased the ICERs of fulvestrant vs comparators considerably. Decreasing the value of incremental scale parameter of the OS curve increased the ICERs. For instance, an incremental scale parameter of [redacted] (near the

upper range of the confidence interval) increased the ICER of fulvestrant vs anastrozole to £208,231 (an increase of £174,132 per QALY from the base case ICER) whereas the ICER increased to £40,761 per QALY (an increase of £6,662 from the base case ICER) when the incremental scale parameter was set at [redacted]. The incremental results obtained from scenario 3 to 6 were comparable to the company's base case results, with ICERs ranging between £32,084 and £35,496 per QALY for fulvestrant vs anastrozole.

For the ERG base case, we combined scenarios 3, 4, 5 and 6. The incremental results obtained are presented in Table 2. The ERG base case ICER for fulvestrant vs anastrozole is £33,455 per QALY which is slightly less than the company's base case ICER of £34,099 per QALY.

Table 2 ERG base case results

Treatments	Total costs	Total QALYs	Incremental costs	Incremental QALYs	Incremental ICER (cost per QALY)
Letrozole	£11,336	2.68	--	--	--
Anastrozole	£11,356	2.68	--	--	--
Tamoxifen	£11,852	2.47	£496	-0.21	Dominated
Fulvestrant	£29,866	3.23	£18,510	0.54	£33,455

1 Introduction to ERG Report

This report is a critique of the company's submission (CS) to the National Institute of Health and Care Excellence (NICE) from AstraZeneca on the clinical effectiveness and cost effectiveness of fulvestrant for untreated hormone-receptor positive locally advanced or metastatic breast cancer. It identifies the strengths and weakness of the CS. Clinical experts were consulted to advise the Evidence Review Group (ERG) and to help inform this review.

Clarification on some aspects of the CS was requested from the company by the ERG via NICE on 01 June 2017. A response from the company via NICE was received by the ERG on 16 June 2017 and this can be seen in the NICE committee papers for this appraisal.

2 BACKGROUND

2.1 Critique of company's description of underlying health problem

The ERG considers that the CS provides a clear and fairly accurate overview of the prevalence, cause and prognosis of breast cancer (CS pp. 26-27), including the impact of the disease on patients, carers and society (CS pp. 31-34). The CS details the different subcategories of breast cancer based on the expression of hormone receptors (which may also be termed endocrine receptors) for oestrogen and progesterone, and the human epidermal growth factor 2 receptor (HER2) (CS p27). A detailed explanation of the role of endocrine receptors is provided, including the receptor for the subgroup of interest [oestrogen receptor positive (ER+)]. The CS highlights that approximately 6% of women at initial presentation have advanced breast cancer (ABC). However, the NICE final scope[1] suggests that around 13% of women with invasive breast cancer have locally advanced or metastatic disease when diagnosed (of which around 5% have Stage IV ABC according to expert opinion). The CS states that of these patients, a panel of UK breast cancer oncologists estimated that 40% have visceral disease and our clinical experts concur with this estimate (the NICE final scope suggests this figure to be around 35%). The CS acknowledges that ER+ HER2 negative ABC is largely incurable; 44% of women die within five years of diagnosis, rising to over 70% in patients with Stage IV disease.

The subgroups of ABC of interest in the NICE final scope[1] are people with visceral and non-visceral disease. The CS acknowledges that visceral metastatic breast cancer (defined as metastasis to internal organs of the body, including liver, lungs or brain) confers a worse

prognosis than bone metastasis alone, which is not normally immediately life-threatening. Around 15% to 30% of women with ABC go on to develop brain metastases, with a median survival time of three to six months from development (CS p. 27).

2.2 Critique of company's overview of current service provision

The CS provides a clear and accurate overview of how ER+ breast cancer is currently managed in patients with ABC. An illustration of the NICE treatment pathway for ABC[2] is provided (CS p. 31) which was designed and adapted to illustrate the suggested positioning of fulvestrant in the pathway. Figure 1 shows this pathway omitting fulvestrant.

The target population of the submission are post-menopausal women with ER+ ABC, without life threatening disease, who receive endocrine therapy [with the aromatase inhibitor (AI) anastrozole or letrozole] in the first instance, or tamoxifen if AI's are not tolerated or are contraindicated under current guidance.[2] The CS states (CS p. 31) that NICE have recently recommended that women suffering recurrence or progression after a first line of AI therapy may be switched to a second line AI such as exemestane (potentially in combination with everolimus).[3] However, this population group are outside of the NICE scope[1] for this submission. For those with life-threatening disease or requiring early symptom relief, NICE Clinical Guideline CG81[2] recommends chemotherapy. The CS states that there is an absence of detailed data on how many lines of different endocrine therapies are typically administered in the UK, which a panel of UK Breast Cancer Oncologists estimates to be around 2.5 lines and our clinical experts concur with this estimate (CS p. 32).

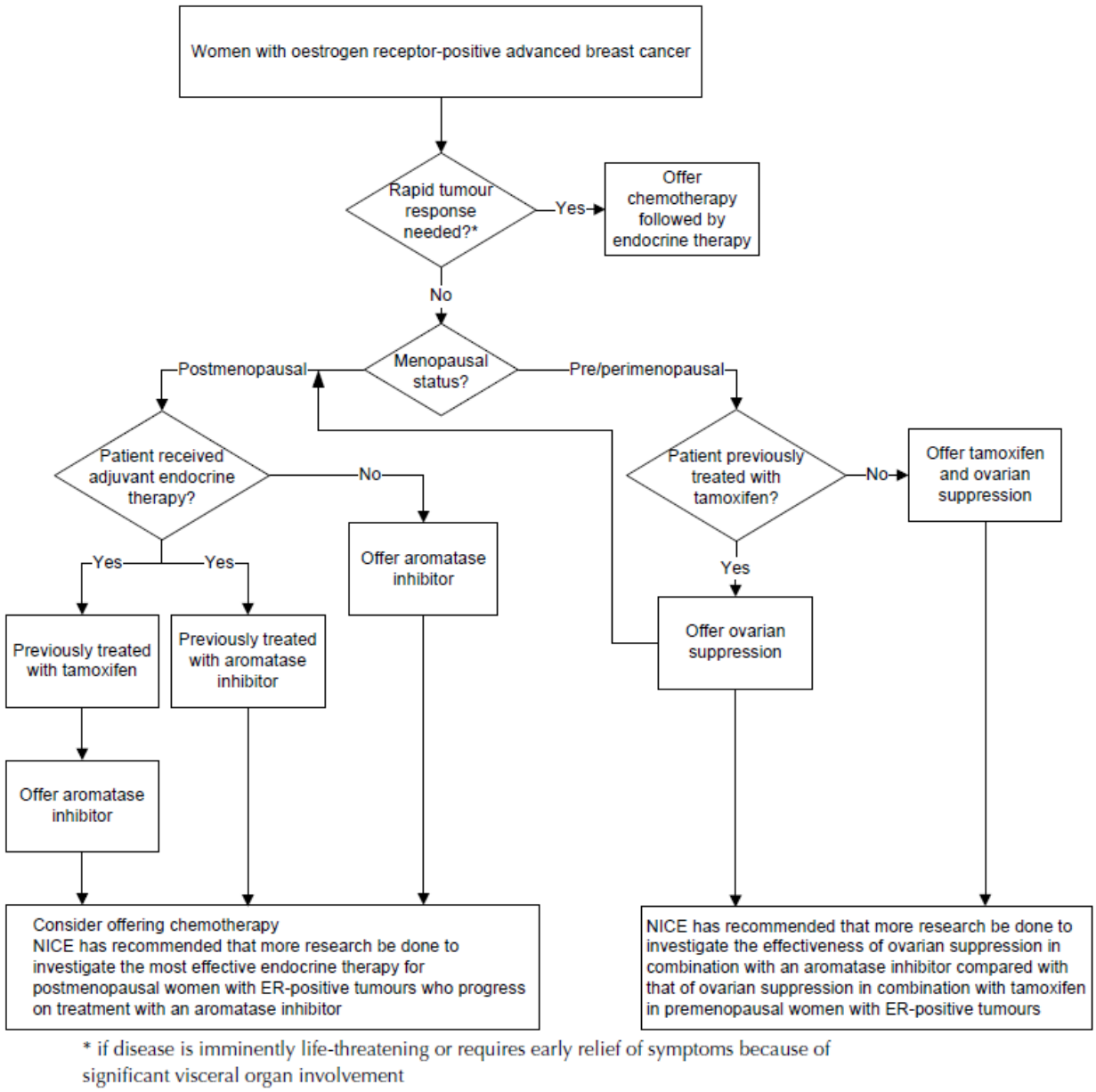


Figure 1 Treatment pathway for women with oestrogen receptor-positive ABC (from NICE CG81)

2.3 Critique of company's definition of decision problem

Population

The population specified in the company's decision problem is post-menopausal people with locally advanced or metastatic hormone receptor-positive (HR+) breast cancer, who have not received endocrine therapy. The patient population matches that specified in the final scope issued by NICE.[1]

Intervention

In accordance with the final scope,[1] the intervention described in the company's decision problem is fulvestrant (brand name: Faslodex®). Fulvestrant is a selective oestrogen receptor degrader and works by binding to endocrine receptors and downregulating oestrogen receptor (ER) protein expression in human breast cancer cells.

The Summary of Product Characteristics (SmPC) (2009),[4] indicates that fulvestrant (500 mg) is administered by two pre-filled syringes each containing 250 mg fulvestrant in 5ml solution. It is administered by slow intramuscular (IM) injection (1-2 minutes/injection), one in each buttock (gluteal area). The recommended dose is 500 mg at intervals of one month, with an additional 500 mg dose given two weeks after the initial dose. It is currently indicated for the treatment of postmenopausal women with ER+, locally advanced or metastatic breast cancer for disease relapse on or after adjuvant anti-oestrogen therapy, or disease progression on therapy with an anti-oestrogen.

The current marketing authorisation for the 500 mg dose was received from the European Medicines Agency (EMA) on 16th March 2010 and was launched in the UK on 3rd June 2010. Fulvestrant is currently being considered for a change in the marketing authorisation to enable its use for the treatment of ER+, locally advanced or metastatic breast cancer in postmenopausal women;

[redacted]

- *with disease relapse on or after adjuvant anti-oestrogen therapy, or disease progression on anti-oestrogen therapy*

Committee for Medicinal Products for Human Use (CHMP) opinion is expected in [redacted]

[redacted] with full marketing authorisation anticipated in [redacted]

The recommended dose regimen for the proposed new indication is: 500 mg to be slowly delivered IM as two 5 ml injections, one in each buttock, on days 1, 15, 29 and once monthly thereafter. The intervention described in the decision problem is appropriate for the National Health Service (NHS) and reflects its draft licence indication.

Comparators

The comparators described in the company's decision problem are AIs (such as anastrozole and letrozole) and tamoxifen (if AIs are not tolerated or are contraindicated). The comparators match those in the NICE scope[1] and are appropriate for the NHS.

Outcomes

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

- overall survival (OS)
- progression-free survival (PFS)
- response rate
- adverse effects of treatment
- health-related quality of life (HRQoL).

Clinical expert opinion agreed that outcomes are appropriate and clinically meaningful.

Economic analysis

The CS states that the economic analysis specified in the decision problem is the same as the final scope issued by NICE[1] (CS p. 11 Table 1) and the ERG agrees. The company have conducted a cost-utility analysis with a time horizon of 30 years (CS p. 142). Given the starting age of the modelled cohort is 63.5 years, this time horizon is appropriate for considering differences in costs and outcomes between treatments for patients with untreated hormone-receptor positive locally advanced or metastatic breast cancer. Costs are considered from the NHS and Personal Social Services (PSS) perspective.

No Patient Access Scheme (PAS) discount has been proposed and none of the comparators are subject to a PAS.

Other relevant factors

The NICE scope[1] states that, if the evidence allows, subgroups of people with visceral disease and people with non-visceral disease should be considered. The CS presents subgroup analyses for the visceral disease and non-visceral disease subgroups in section 4.8 (CS p. 77) alongside other subgroup analyses. In the FIRST trial post-hoc subgroup analyses were

conducted for time to progression (TTP) and OS, whereas in the FALCON trial subgroup analyses were prespecified and are presented for PFS and OS.

No equity or equality issues were specified in the final scope[1] or identified by the company. The ERG is not aware of any issues related to equity or equality in the use of fulvestrant in patients with untreated hormone-receptor positive locally advanced or metastatic breast cancer.

3 CLINICAL EFFECTIVENESS

3.1 Critique of company's approach to systematic review

3.1.1 Description of company's search strategy

The company's submission (CS) reports the following 3 literature searches:

- Clinical effectiveness: CS Appendix A (database inception - January 2017)
- Cost effectiveness: CS Section 5.1 (May 2016)
- HRQoL: CS Section 5.4.3 (October 2013 to June 2016)

The clinical effectiveness searches represent a one-step approach designed to identify trials relating to fulvestrant and comparator drugs, negating the need for additional indirect comparison searches. Simultaneous searches of Embase and Medline were undertaken on the host Embase.com. Separate searches were conducted in The Cochrane Library and on Pubmed to identify in-process records. The list of drug terms in the search strategy is comprehensive, using a mix of free text and descriptor terms, including an appropriate clinical trials filter. However there is an error in the Embase.com tabulation, where sets are incorrectly linked (line 37: where sets 1 and 10 and 36 should have been combined rather than 1 and 10 and 37), and also in the Cochrane search (line 37: where lines 4 and 37 should have been combined rather than lines 7 and 36), although these could be typographical errors in the CS. The Pubmed search is linked correctly. The ERG ran a search for fulvestrant trials from database inception on Embase and Medline and found no additional relevant trials. Pertinent conference proceedings were searched by the company which the ERG deemed sufficient. It is stated in section 4.14 that there are no other ongoing studies to provide further relevant evidence due to be completed within the next year. The ERG concurs, after conducting an ongoing trials search on the UK Clinical Trial Gateway (UKCTG), clinicaltrials.gov, the World Health Organisation International Clinical Trials Registry Platform (WHO ICTRP), Prospero and Astra Zeneca's website.

The cost effectiveness searches took an unusual approach as the company elected to search only for health technology assessments (HTA) of therapies for locally advanced or metastatic breast cancer. The review included a search of national reimbursement and technology assessment organisations [Canadian Agency for Drugs and Technologies in Health (CADTH, Canada); the Canadian Oncology Drug Review; National Institute for Health and Clinical Excellence (NICE, England and Wales); Pharmaceutical Benefits Advisory Committee (PBAC,

Australia); Scottish Medicine Consortium (SMC, Scotland)]. The websites were searched in May 2016 for any HTA in breast cancer and those related to advanced or metastatic breast cancer were included. The company did not include any additional inclusion or exclusion criteria. The ERG undertook checks on the CEA Registry and also on HTA and NHS Economic Evaluation Database (NHSEED) within the Cochrane Library. Medline and Embase were additionally searched over the last 10 years applying a standard cost filter to fulvestrant and to three chosen key comparator drugs: anastrozole, letrozole and tamoxifen. Two extra records were identified from the search results by a senior health economist and checked by a second health economist (see section 4.2).

The HRQoL searches were run to identify studies published between October 2013 and June 2016 on Embase. The ERG ran update searches (2016-2017) on Medline and Embase. ScHARR's Health utility Database (ScHARRHUD) was also searched and four additional records were identified from search results by a senior health economist (further detail is provided in Section 4.3.6.)

In summary the searches are of a reasonable quality, transparent and the CS contains Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow charts of results. Despite documentation errors within the search strategies, further searching by the ERG did not produce additional relevant clinical effectiveness evidence. Although the ERG identified additional cost and quality of life studies these were either not relevant or not suitable for use in the appraisal.

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

The company provides an overview of the inclusion/exclusion criteria for the systematic literature review (SLR), evaluating the clinical effectiveness of fulvestrant (CS Table 8, p. 36). The criteria were designed to capture not only studies of fulvestrant, but also any studies that might be pertinent to the wider evidence network and relevant if an NMA had to be undertaken. The company's inclusion/exclusion criteria limited study design to randomised controlled trials (RCTs) (irrespective of blinding status) and only English language publications were eligible.

Population

The population of the SLR was restricted to female, post-menopausal patients (≥18 years of age) with HR+, HER2 negative, locally advanced or metastatic breast cancer. The CS states

that HER2 receptor testing was not usually carried out in regular clinical practice until the mid-2000s and so therefore eligibility was not restricted to a HER2 negative population in order not to exclude some important comparators (CS p. 38). Additional inclusion/exclusion criteria were applied to full-text copies of studies, to exclude pre-menopausal females (studies including both pre- and post-menopausal females could be included if sub-group data for the post-menopausal population was reported, CS p. 38). This ensured that the included population was in line with that stated in the final NICE scope.[1] The final inclusion criteria for treatment were aligned with the FALCON trial, including patients with either:

- locally advanced disease not amenable to surgery or radiotherapy of curative intent (patients may have had one line of cytotoxic chemotherapy, following which they must remain unsuitable for therapy of curative intent)

or

- metastatic disease (patients may have had one line of cytotoxic chemotherapy as previous treatment of breast cancer but must show progressive disease prior to enrolment.

Although the NICE scope[1] focuses on an endocrine treatment-naive population, the criteria of SLR allowed for the inclusion of studies in which >70% of patients met that criterion (CS p. 39).

Intervention

The inclusion criteria listed hormonal and chemo- or biologic therapies, including fulvestrant. The CS SLR therefore includes a wider variety of interventions than indicated in the NICE scope.[1] The CS states that both licensed and investigational pharmacological treatments for HR+ [expressing the ER and/or the progesterone receptor (PgR)] locally advanced or metastatic breast cancer were included in the SLR (based on recommendations of clinical guidelines, searching of clinicaltrials.gov and by expert input). As the inclusion criteria of the SLR were designed to also identify studies for the NMA, this may explain the variety of drugs included in the criteria.

Comparators

The inclusion criteria for the comparators in the submission were:

- any included intervention
- any other pharmacological intervention and
- placebo/best supportive care.

This is much broader than the criteria specified by the NICE final scope[1] for comparators, which is limited to AIs (such as anastrozole and letrozole) and tamoxifen (if AIs are not tolerated or are contraindicated). Again the broad scope was designed to also identify studies for the NMA.

Outcomes

The company has listed all the outcomes specified in the final scope in their decision problem. These outcomes are appropriate and clinically meaningful to patients, and the ERG considers that the company has included all important outcomes in the decision problem.

The CS provides a PRISMA diagram illustrating the number of records identified and included/excluded records at each stage of the SLR screening processes (CS Fig. 5, p. 42). Out of 12,498 records found by database searching, a final number of 44 studies (based on 91 publications) were data extracted. However, a further 38 studies were excluded from the wider evidence network (exclusion reasons were documented for most of these studies in CS Appendix B, missing exclusion reasons were supplied in the company's response to clarification question A1). This left six studies that were included in the main evidence review. Two of these studies provided heat to head comparisons between fulvestrant and anastrozole (Section 3.1.3). The other four studies included the other comparator drugs (letrozole and tamoxifen) and these contributed data for the NMA (Section 3.1.7).

The company did not specify treatment setting as an inclusion criterion nor place any limits on inclusion relating to the quality of the RCTs, which is appropriate. Overall, the ERG considers the inclusion criteria reasonable.

The CS does not highlight any potential bias in their selection of studies. Overall, the ERG considers that the eligibility criteria used in the SLR review were appropriate and matched the decision problem (notwithstanding the need for additional criteria to conduct an NMA).

3.1.3 Identified studies

The company's SLR included two RCTs, FIRST[5-7] and FALCON,[8] comparing fulvestrant and the AI anastrazole for the treatment of ABC. For the NMA a further four RCTs were identified, of which three compared anastrazole versus tamoxifen; the North American trial;[9] the TARGET trial[10] and a trial by Milla-Santos et al.[11] The fourth study was the PO25 trial[12,13] which

compared letrozole versus tamoxifen. RCT reports were provided electronically on 19th May as part of the submission reference pack.

Summary details of the two fulvestrant RCTs, FIRST[5-7] and FALCON[8] are presented in the CS (Sections 4.3 to 4.5, pp. 48-66). These studies were sponsored by the company.

- The differences between the two RCTs in terms of trial designs,-population, eligibility criteria, setting, intervention, outcomes and pre-planned subgroups are summarised in the CS (Table 14, p. 55).
- A diagram covering trial design, intervention description and initial patient numbers was provided for both the FIRST (CS Figure 7, p. 49) and the FALCON study (CS Figure 8, p. 51).
- QUORUM flow diagrams describing numbers randomised and drop out data are provided (FIRST RCT, CS Figure 9, p. 61, FALCON RCT, CS Figure 10, p. 64).
- For both studies, power/sample size calculations were provided (CS pp. 57-58).
- Analysis sets for the FIRST RCT are described on CS pp. 62-63 and summarised in CS Table 16, p63. For the FALCON trial they are described on CS pp. 65-66 and summarised in CS Figure 11, p66. There was an intention-to-treat (ITT) analysis for key outcomes in both trials.

The ERG has summarised some of the key trial characteristics in Table 3 and key patient characteristics in Table 4.

The fulvestrant trials were both multi-centre RCTs that were conducted across a variety of countries (Table 3). The FIRST trial was a phase II open-label non-inferiority trial with CBR as the primary outcome whereas FALCON was a phase III double-blind superiority trial with PFS as the primary outcome. The United Kingdom (UK) is listed one of the countries from which patients were recruited for the FIRST trial and the ERG believes that patients were also recruited from the UK for the FALCON trial because, although this is not explicitly stated, two of the listed study investigators who participated in the study have a UK location. The number of patients who were included from the UK in each trial is not stated.

Table 3 Summary of key design characteristics of the fulvestrant trials

Trial	FIRST[5-7]	FALCON[8]
Trial design	PHASE II open-label multicentre RCT.	PHASE III double blind multicentre RCT
Enrolled population	Postmenopausal women with ER + &/or PgR+ ABC who had either, never received endocrine therapy for advanced disease or had received previous endocrine therapy for early disease completed ≥12 months prior to randomisation.	Postmenopausal women presenting with ER +ve &/or PgR+, HER2-ve ABC who had never received endocrine therapy for breast cancer.
Number of centres (location)	62 (Brazil, Bulgaria, Czech Republic, France, Italy, Poland, Spain, United Kingdom, & United States).	113 (20 countries in Asia, Europe, North America, South America & South Africa).
Intervention (n in arm)	Fulvestrant (n=102)	Fulvestrant (n=230)
Comparator (n in arm)	Anastrazole (n=103)	Anastrazole (n=232)
Data analysis points	Primary analysis of CBR: 6 months after last patient randomly assigned[6] Data cut-off 10th January 2008	Primary analysis of PFS: planned for when 306 progression events had occurred (68% of planned sample size)[8] Data cut off 11th April 2016
	Follow-up analysis for TTP: planned for when 75% of patients had discontinued treatment.[5] Data cut-off 26th March 2010	Interim OS analysis: conducted at the same time as PFS analysis above[8]
	OS analysis: Protocol amendment to assess OS after approximately 65% of patients had died.[7] Data cut-off 15th July 2014	OS analysis: planned for when approximately 50% of patients have died (not yet reported). [redacted]
Primary outcome	CBR (non-inferiority)	PFS (superiority)

ABC, advanced breast cancer; CBR, clinical benefit rate; ER+, oestrogen receptor positive; HER2-ve, human epidermal growth factor 2 receptor negative; PFS, progression-free survival; PgR+, progesterone receptor positive; RCT, randomised controlled trial.

The CS states that baseline characteristics of enrolled patients were well balanced between treatment groups (CS Table 12, p. 44). Whilst this generally appears to be the case for FALCON (CS Table 17, p. 65), in FIRST (CS Table 15, p. 62), the fulvestrant treatment arm includes a

higher percentage of women with prior endocrine therapy completed >12 months prior to randomisation than the comparator arm (28% vs 22% anastrozole). The ERG believes difference is unlikely to have caused an imbalance in outcomes between the trial arms.

The CS contains a summary of the baseline characteristics of participants with fuller information reported in the published papers for the two trials. The ERG added some information from these publications to the table of baseline characteristics that are relevant to the assessment [e.g. receptor status (including HER2 status) and numbers of participants with visceral disease] (Table 4). In the FIRST trial a greater proportion of participants in the anastrozole group had any visceral disease (56% vs 47% in the fulvestrant group) and a greater proportion had lung metastases (41% vs 29% in the fulvestrant group). In the FALCON trial a smaller proportion of women in the anastrozole arm were aged 65 years or more (39% vs 47% in the fulvestrant group) and a greater proportion had 'Other non-visceral' as a disease site (35% vs 26% in the fulvestrant arm).

There were some differences between these two trials in baseline characteristics (see Table 4) The mean age of participants in the FALCON trial was slightly lower (63 years vs 67.1 years FIRST) and FALCON included more women categorised as 'Asian' or 'Black or Other' (24% vs 3% FIRST) than the FIRST trial. A slightly lower percentage of women in the FALCON trial had locally advanced disease (13% vs 18% FIRST), and therefore correspondingly slightly more had metastatic disease compared to women in the FIRST trial (87% vs 82% FIRST). The majority of women in both trials had ER+, PgR+ breast cancer (FIRST 76% vs Falcon 77%), and about 19% of women in the FIRST trial had a positive HER2 status (HER2 positive breast cancer patients were excluded from the FALCON trial). Nearly 10% more participants in FIRST trial received adjuvant chemotherapy compared to FALCON (26% vs 13.5% FALCON) and 23% of the FALCON population received radiotherapy (a characteristic not reported by FIRST). Clinical advice to the ERG indicated that in UK practice it would be unusual to give chemotherapy before endocrine treatment.

In summary, there are two important differences between the FIRST and FALCON trial populations, which are both a consequence of their different inclusion and exclusion criteria. As noted previously, a quarter of all participants in the FIRST trial (n=52) had received prior endocrine therapy (in all but one case this occurred >12 months prior to randomisation). In FALCON, previous hormonal treatment was an exclusion criterion so 99.4% of the population

were endocrine therapy naïve (protocol errors meant that three women who had previously received endocrine therapy were included). As noted in the published paper for the FALCON study,[8] the aim of limiting the included population to women who were prior endocrine therapy naive was to avoid reducing the efficacy of anastrozole in the control group through exposure to prior adjuvant endocrine therapy. The second important difference, also noted above, is that approximately 19% of women in the FIRST trial had a positive HER2 status whereas these patients were excluded from the FALCON trial. HER2 positive breast cancers are typically more aggressive and spread more quickly than HER2 negative breast cancers. The HER2 positive breast cancer patients in FIRST might have been expected to have less favourable outcomes than the HER2 negative patients but this is not commented on in the CS or the FIRST trial publications.[5-7] The other differences between the trials in patient baseline characteristics and treatment experience appear to be minor.

Table 4 Baseline characteristics of patients in the included fulvestrant RCTs

	FIRST		**FALCON**	
Patient demographic and disease characteristics	**Fulvestrant 500 mg (n=102)**	**Anastrozole 1 mg (n=103)**	**Fulvestrant 500 mg (n=230)**	**Anastrozole 1 mg (n=232)**
Gender, % Female	100	100	100	100
Age (years), Mean (SD)	67 (9)	68 (9)	63.8 (9.86)	63.3 (10.38)
Median	66	68	64.0	62.0
Range	40–89	48–87	38-87	36-90
Race, n (%)	**Caucasian**		**White**	
	97 (95.1)	102 (99)	175[a] (76)	174[a] (75)
	Black		**Black or other**	
	3 (2.9)	0	19[a] (8)	24[a] (10)
	Other		**Asian**	
	2 (2.0)	1 (1)	36[a] (16)	34[a] (15)

Patient demographic and disease characteristics	FIRST		FALCON	
	Fulvestrant 500 mg (n=102)	**Anastrozole 1 mg (n=103)**	**Fulvestrant 500 mg (n=230)**	**Anastrozole 1 mg (n=232)**
Receptor status,[b] n (%)				
HR+	102 (100.0)	103 (100.0)	230 (100.0)	232 (100.0)
ER+, PgR+	78 (76.5)	78 (75.7)	175 (76%)	179 (77%)
ER+, PgR–	19 (18.6)	19 (18.4)	44 (19%)	43 (19%)
ER+, PgR unknown	1 (1.0)	3 (2.9)	10 (4%)	7 (3%)
ER–, PgR+	3 (2.9)	3 (2.9)	1 (<1%)	3 (1%)
ER unknown, PgR+	1 (1.0)	0	0	0
HER2 status[b] n (%)	**2+/3+**		**Positive**	
	19 (18.6)	19 (18.4)	0	1 (<1%)
Negative	48 (47.1)	49 (47.6)	230 (100%)	231 (100%)
Unknown	35 (34.3)	35 (34.0)	0	0
Disease stage, n (%)				
Locally advanced only	19 (18.6)	18 (17.5)	28[a] (12)	32[a] (14)
Metastatic	83 (81.4)	85 (82.5)	202[a] (88)	200[a] (86)
Measurable disease	89 (87.3)[b]	93 (90.3)[b]	193[a] (84)	196[a] (84)
Site of disease[b] n (%)				
Any visceral disease	48 (47.1)	58 (56.3)	135 (59%)[c]	119 (51%)[c]
Previous treatment modalities[d,e]				
Prior endocrine therapy n (%)				
None	73 (71.6)	80 (77.7)	228 (99.1)[f]	231 (99.6)[f]
Completed ≤12 months prior to randomisation	1 (1.0)	0	2[a] (1)	1[a] (<1)
Completed >12 months prior to randomisation	28 (27.5)	23 (22.3)		

Patient demographic and disease characteristics	FIRST		FALCON	
	Fulvestrant 500 mg (n=102)	Anastrozole 1 mg (n=103)	Fulvestrant 500 mg (n=230)	Anastrozole 1 mg (n=232)
Prior chemotherapy, n (%)				
None	73 (71.6)	78 (75.7)	152 (66)[f]	151 (65)[f]
Any chemotherapy	NR	NR	78[f] (34)	81[f] (35)
Advanced disease[g]	NR	NR	36[a] (16)	43[a] (19)
Adjuvant chemotherapy	29 (28.4)	25 (24.3)	35[a] (15)	27[a] (12)
Neo-adjuvant	NR	NR	11[a] (5)	16[a] (7)
Any radiotherapy, n (%)	NR	NR	53[a] (23)	50[a] (22)

Table based on CS Table 15 (p. 62) and Table 17 (p. 65).

ER, Oestrogen receptor; HER2, Human epidermal growth factor 2; HR+, Hormone receptor postive; NR, Not reported; PgR, Progesterone; SD, Standard deviation.

[a] number of participants obtained from trial publication[8].

[b] obtained from trial publications (FIRST,[6] FALCON[8]).

[c] Includes patients with site of baseline disease as any of the following: adrenal, bladder, CNS, oesophagus, liver, lung, peritoneum, pleura, renal, small bowel, stomach, pancreas, thyroid, colon, rectal, ovary, biliary tract, ascites, pericardial effusion, spleen, or pleural effusion.

[d] Previous study treatment, as deemed by the sponsor to be relevant to the interpretation of the results.

[e] Therapies prior to enrolment.

[f] Calculated by ERG.

[g] Includes 1L, 2L, 3L, metastatic and palliative chemotherapies.

Both FALCON and FIRST met the inclusion criteria of the submission and all relevant RCTs appear to have been identified in the CS.

The CS did not list any additional ongoing trials which included fulvestrant as a treatment that was not used in combination with another drug and this appears to be correct.

Information is given on the additional four RCTs in the NMA, though in less detail than the two fulvestrant RCTs.

- A summary of methodological characteristics is provided in CS Table 25 (p. 85)
- A summary of outcomes reported by the studies is provided in CS Table 23 (p. 83) and the definitions of PFS and TTP are compared in CS Table 24 (p. 84).

- Clinical characteristics of the participants in the trials are summarised in CS Tables 26 to 28 (pp. 87-89).

The CS did not include any non-randomised or non-controlled studies relevant to the decision problem.

3.1.4 Description and critique of the approach to validity assessment

The CS provided a quality assessment of all the included RCTs (CS Table 12, p. 44 to 45). While an assessment was provided (described as the NICE critical appraisal checklist), the table includes a Jadad score (known to be 1 to 5) and an Allocation Concealment Grade (presumed to be either A or B) for the RCTs. The use of quality assessment scores is discouraged by The Cochrane Collaboration.[14] The CS does not provide any information about the grading scales employed for the Allocation Concealment Grade or indeed interpretation of the final grading score supplied for each of the RCTs (FIRST: 2B and FALCON: 5A).

The ERG has used the NICE criteria for assessing the risk of bias in RCTs based on criteria from the Centre for Reviews and Dissemination for systematic reviews,[15] hence omitting the Jadad score and Allocation Concealment Grade.

The ERG's quality assessment mostly agrees with that of the company (see Table 5). However, contrary to the company, the ERG thinks that it is unclear if there was any bias in relation to allocation concealment in the FIRST trial, mainly due to insufficient details being reported in the CS and the publications. The trial used randomisation cards, but it is unclear if these were in sealed opaque envelopes. The use of unsealed cards or insufficiently opaque envelopes could leave the trial at a potential high risk of bias due to inadequate allocation concealment. The ERG also differs to the company in the assessment of baseline differences for the FIRST and FALCON trials. In both trials there were some differences in the treatment groups at the outset of the trials (as detailed in Table 5). The ERG sought the opinion of the clinical experts for their view on whether these differences could have had an impact on the reported outcomes. Their view was that in the FIRST trial the higher proportion of those with 'any visceral disease' in the anastrozole arm would be associated with worse prognosis. In the FALCON trial the difference in age was unlikely to have an impact and the anastrozole arm might have had a slightly better prognosis as a greater proportion had 'other non-visceral' disease this arm. The ERG agrees with the company's assessment of there being a potential high risk of bias in the FIRST trial in

relation to blinding, especially with regard to outcome assessors, as this was an open label trial. While a blinded independent review was carried out for the primary endpoint, it is unclear if this was applied to other endpoints. Both studies used an ITT analysis and although neither reported how missing data was dealt with in the analyses, missing data appeared to be balanced between trial arms. Overall, both studies appear to have been well conducted.

Table 5 Company and ERG assessment of trial quality

NICE Quality Assurance Criteria for RCT[15]	**Judgements* FIRST**		**Judgements* FALCON**	
1. Was the method used to generate random allocations adequate?	CS:	Low risk	CS:	Low risk
	ERG:	Yes, low risk	ERG:	Yes, low risk
ERG comment: FIRST – CS states central randomisation (CS p. 44) but this information is not present in any of the published papers. [redacted] [16] FALCON - computer generated randomisation scheme.[8]				
2. Was the allocation adequately concealed?	CS:	Low risk	CS:	Low risk
	ERG:	Unclear, uncertain risk	ERG:	Yes, low risk
ERG comment: FIRST - One of the publications states that patients were randomised sequentially using randomisation cards, with the clinical study team unaware of the randomisation scheme.[5] It is unclear if the cards were sealed in opaque envelopes, as unsealed randomisation cards or insufficiently opaque envelopes would leave the study at a potentially high risk of allocation bias. FALCON – integrated voice or web response system.[8]				
3. Were the groups similar at the outset of the study in terms of prognostic factors, e.g. severity of disease?	CS:	Low risk	CS:	Low risk
	ERG:	Unclear, uncertain risk	ERG:	Unclear, uncertain risk
ERG comment: FIRST – around 6% more women in the fulvestrant arm had previous endocrine therapy for early disease compared to the comparator arm (28% vs 22% anastrozole), but fewer had any visceral disease (47% vs 56% anastrozole) and lung metastases (29% vs 41% anastrozole). FALCON - a smaller proportion in the anastrozole				

NICE Quality Assurance Criteria for RCT[15]	Judgements* FIRST		Judgements* FALCON	
arm were aged 65 years or more (39% vs 47% fulvestrant) and a greater proportion had 'Other non-visceral' as a disease site (35% vs 26% fulvestrant).				
4. Were the care providers, participants and outcome assessors blind to treatment allocation? If any of these people were not blinded, what might be the likely impact on the risk of bias (for each outcome)?	CS:	High risk	CS:	Low risk
	ERG:	No, high risk	ERG:	Yes, low risk
ERG comment: FIRST - was an open label study, although a blinded independent review was carried out by a radiologist at BioImaging Technologies to provide assurances that the open-label design did not bias the results of the tumour assessments in this study. The independent reviewers' evaluation was used to corroborate the local investigator read analysis of the primary endpoint (CBR) and concordance was high, but it is unclear if this applied to other outcomes. FALCON – was a double-blind, double-dummy trial and publication states that blinding included those assessing outcomes.[8]				
5. Were there any unexpected imbalances in drop-outs between groups? If so, were they explained or adjusted for?	CS:	Low risk	CS:	Low risk
	ERG:	No, low risk	ERG:	No, low risk
ERG comment: Both trials had higher discontinuation in the comparator arm largely due to disease progression/worsening condition. Participant numbers and reasons for discontinuations were detailed in both studies.				
6. Is there any evidence to suggest that the authors measured more outcomes than they reported?	CS:	Low risk	CS:	Low risk
	ERG:	No, low risk	ERG:	No, low risk
ERG comment: Outcomes listed in the methods sections of the published papers match those presented in the results. The ERG did not check trial protocols.				
7. Did the analysis include an ITT analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	CS:	Low risk	CS:	Low risk
	ERG:	Yes & Unclear (low risk)	ERG:	Yes & Unclear (low risk)
ERG comment: Although there were insufficient details of how missing data was dealt with for both trials, missing data appears balanced between groups in each trial hence risk of bias is likely to be low.				

CS judgements taken from CS Table 12, p. 44.

3.1.5 Description and critique of company's outcome selection

As stated earlier the outcomes in the CS match those listed in the NICE scope[1] (CS p. 11), those being OS, PFS, objective response rate (ORR), AEs of treatment and HRQoL. For the FALCON study, data are reported from one analysis conducted following the 11th April 2016 data cut off (the point of PFS analysis). For the FIRST trial, data are available from three time points:

- the first data cut-off (10th January 2008)[6]
- the first follow-up (26th March 2010)[5]
- the final assessment of OS (15th July 2014)[7]

The ERG found that it was not always clear in the CS which of these three different points for FIRST was being used for reporting different outcomes.

The primary outcome in the FALCON trial was PFS, defined as time from randomisation until objective disease progression as defined by RECIST 1.1 (Response Evaluation Criteria in Solid Tumours), surgery or radiotherapy to manage worsening of disease or death by any cause (in the absence of progression) (CS Table 24 p. 84).[8]). PFS was not reported by the FIRST trial, instead the trial reported TTP (a secondary outcome), defined as time from randomisation to the time of the earliest evidence of objective disease progression or death from any cause prior to documented progression (CS Table 24 p. 84).

The primary outcome in the FIRST trial was clinical benefit rate (CBR), a secondary outcome in the FALCON trial. This was defined in both RCTs as the proportion of all randomly assigned patients who had best overall response of complete response (CR), partial response (PR) or stable disease (≥24 weeks), CS, p. 50.[6,8]

OS was defined as time from randomisation until death by any cause in both trials.[7,8] However, this was not a planned outcome in the original protocol of the FIRST trial, but added as an addendum after TTP results were analysed following ongoing monitoring and at which time approximately 65% of deaths had occurred (CS p 69).[7]

Key secondary outcomes were CR, PR, safety, HRQoL (FALCON only), ORR and duration of response (DoR; defined as time from response through to progression in FIRST and expected

DoR (EDoR) for the FALCON study) (CS p. 48 and 50). ORR was defined as the proportion of patients with a best overall response of CR or PR assessed only in patients with measurable disease at baseline in both trials.[6,8]

Other outcomes presented in the CS (not contained in the NICE final scope[1]) reported by both trials were median duration of clinical benefit (DoCB) and progressive disease. In addition, the CS presented outcomes such as expected duration of clinical benefit (EDoCB) and median time to onset of response for the FALCON trial.

Ten pre-specified adverse events (AEs) are presented for the FIRST trial, while AEs for FALCON were graded according to Common Terminology Criteria for Adverse Events [CTCAE], version 4·0 (no reference provided in either the CS or the published paper[8], but these are published by the U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute[17]). Incidence of serious adverse events (SAEs) and deaths are presented for FIRST at final data cut-off (65% OS), and for FALCON limited to frequency of ≥5% in any treatment group. Discontinuations due to AEs are presentenced in the CS by organ class for both trials.

For the NMA, outcomes were limited to PFS and OS.

All outcomes covered in the trials were reported in the CS.

Only the FALCON trial measured HRQoL, using the EuroQoL5 Dimensions Questionnaire-3L (EQ-5D-3L, visual analogue scale and health index score) (CS, p 72) reported at baseline (week 0) and the end of the trial (week 156). The ERG notes that the EQ-5D is a validated, generic measure of HRQoL and is NICE's favoured HRQoL measure.[18] The index was calculated using the utility value set for the UK (CS p. 148). Utilities for both time spent in PFS and progressed disease (PD) health states were calculated from the EQ-5D and used in the cost-effectiveness analysis (see Section 4.3.6 for more information). HRQoL was also assessed by the Functional Assessment of Cancer Therapy – Breast (FACT-B) questionnaire which is a self-reported questionnaire, comprising four general FACT-B subscales (physical well-being, functional well-being, social well-being and emotional well-being) along with the Breast Cancer-Specific subscale that assesses symptoms/concerns of particular relevance to breast cancer such as body image, arm swelling and tenderness,[19] assessed on a 5-point Likert scale (0 = Not at all to

4 = Very much). The main outcome measure from the FACT-B questionnaire was the trial outcome index (TOI), summarising the physical well-being, functional well-being and breast cancer subscales (CS p 72). The FACT-B is also a validated measure of HRQoL.

Overall, we consider that the CS appropriately addresses the outcomes listed in the NICE scope.[1]

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

The CS reports the results for all the relevant primary and secondary outcomes listed in CS Table 14 (CS p. 55) for the FIRST and FALCON trials. An interim analysis of OS is presented for the FALCON trial which was done at the time of the analysis of the primary outcome, PFS (CS p. 58).

The CS reports the statistical methods used to analyse data and details about power calculations (CS section 4.4 p. 57 to 60). Both trials were adequately powered.

Efficacy results are presented in the CS in terms of percentages, odds ratios or hazard ratios (HRs) with 95% confidence interval (CIs) and p-values. The number of participants included in the analyses is clearly identified and where only a percentage value is reported, the number of participants with an event can usually be calculated. HRQoL data are presented in CS Figure 16 and CS Figure 17 (CS p. 73 to 74), however these data are not presented in tabular form (values need to be read from the figures). The company confirmed in response to the ERG and NICE clarification question A14, that neither the FIRST nor the FALCON trial were designed to formally allow cross-over (treatment switch) between trial arms.

Analysis sets

The CS describes four different analysis sets for the FIRST trial which are summarised in CS Table 16 (CS p. 63), and three analysis sets for the FALCON trial which are summarised in CS Figure 11 (CS p. 66). The ITT analysis set (described as the full analysis set for the FIRST trial) includes all randomised patients analysed in the group to which they were assigned, regardless of actual treatment received.

In most cases the number of participants contributing data to the analyses matches one of the analysis sets described. The two exceptions are in CS Table 19 (CS p. 70, 'Summary of

additional secondary outcomes in FIRST') where the number of patients for the FALCON trial was given in error (confirmed by the company in response to the ERG and NICE clarification question A2) and in the reporting of FALCON HRQoL data, where number of patients at baseline for mean TOI does not match any of the FALCON analysis sets (CS Figure 16, p. 73) and numbers of patients for the EQ-5D-3L analysis are not provided (CS Figure 17, p. 74). In response to the ERG and NICE clarification question A3, the company indicated that although analyses of the TOI were performed on the ITT analysis set, however only [redacted] participants provided evaluable forms at baseline. The company also supplied a reference (which had been omitted in error from the original reference pack) that provided details of the number of patients contributing data to the EQ-5D-3L analysis.

Safety outcomes were analysed using the safety analysis set in both trials. This was defined as all randomised participants who received at least one dose of the trial drug (including placebo fulvestrant or placebo anastrozole). The proportion of participants excluded from the safety outcomes analyses was very low [FIRST trial 1/205 (0.49%); FALCON trial 2/462 (0.43%)].

Subgroups

The CS includes results from subgroup analyses. For the FALCON study analysis of subgroups was performed as defined by covariates and the same approach appears to have been taken in the FIRST study. For the FIRST trial, these were not pre-planned and not part of the initial analysis as published in 2009 by Robertson et al.,[6] however five covariates were pre-defined for the next analysis of TPP published in 2012[5] [age, receptor status, visceral involvement, prior chemotherapy, and measurable disease (CS p. 77)]. In addition six patient exploratory subgroups were prespecified for the post-hoc OS analyses[7] [age, receptor status, visceral involvement, prior chemotherapy, measurable disease, prior endocrine therapy (CS p. 78)].

For the FALCON trial, PFS analyses were pre-specified for six subgroups in the study protocol (if numbers permitted) (ER+ and PgR+ at baseline, metastatic disease at baseline, use of bisphosphonates, measurable disease, prior chemotherapy for locally advanced or metastatic breast cancer, geographic region). Some amendments were then made to the planned analyses before unblinding, which included adding two further subgroups [prior oestrogen containing hormone replacement therapy (HRT) and visceral disease]. The amendments at this stage also included adding subgroup analysis for OS.

The clinical experts the ERG consulted thought the choice of subgroups was on the whole appropriate. One clinical expert thought that the use of bisphosphonates and measurable disease were unlikely to be important prognostic factors whereas other important prognostic factors such as performance status, lactate dehydrogenase (LDH) and number of sites of disease had not been considered.

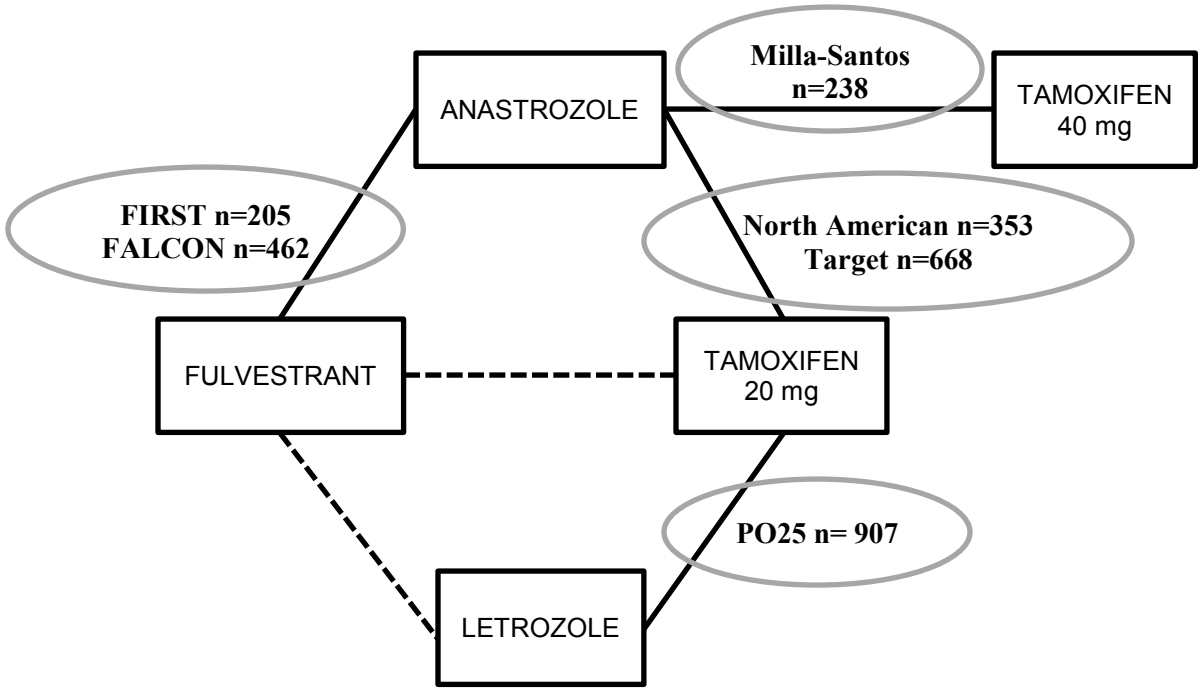
3.1.7 Description and critique of the company's approach to the evidence synthesis

A narrative review of the evidence from the key fulvestrant studies FIRST[5-7] and FALCON[8] is presented in the CS Section 4 (pp. 36 to 81 and 119 to 123). Where possible the ERG has checked key data presented in the CS against those in the publications and identified a few minor discrepancies, most of which appear to be typographical errors.

No pair-wise meta-analyses for the outcomes of interest are presented in the CS for FIRST and FALCON, only network meta-analysis (NMA) was conducted. The ERG and NICE asked the company to provide results for pairwise comparisons of interventions in the NMA (clarification question A10) and these were provided.

Although there were two RCTs that compared fulvestrant with anastrozole, there were no RCTs that compared fulvestrant with the other possible comparators, letrozole and tamoxifen. The company therefore conducted an NMA incorporating trials of anastrozole versus tamoxifen and one trial of letrozole versus tamoxifen to enable an indirect comparison of fulvestrant with letrozole and tamoxifen.

A diagram showing the core network of relevant evidence is presented in the CS (Figure 22, p. 82) however this does not show the indirect comparisons being made. The direct comparisons, evidence for these, and the indirect comparisons being made are shown in Figure 2.



Interventions are shown in rectangular boxes, available trial evidence is shown in oval shapes. Solid lines indicate direct comparisons, dashed lines indicate indirect comparisons.

Figure 2 Network of studies included in the NMA

Of the six trials contributing data to the network, patient level data were available for the two trials conducted by the company (FIRST and FALCON) and also for a combined data set for the two trials that compared anastrozole and tamoxifen 20 mg, TARGET and North American (which were both supported by grants from AstraZeneca and were prospectively designed to allow for combined data analysis). The combined data set is referred to as NorthAmTarget in the CS. For the single trials available for the anastrozole versus tamoxifen 40 mg comparison (Milla-Santos et al.[11]) and the letrozole versus tamoxifen comparison (PO25 trial[12]) the original patient level data were not available.

3.1.7.1 Outcome measures used in the NMA

The outcome measures reported by the six trials contributing to the network of evidence are summarised in a CS table which has been reproduced below (Table 6). All six trials report PFS (or TTP), and five trials report OS (the FALCON trial is not shown as reporting OS, presumably because those data are not yet mature), ORR, CBR, OS rate, and PFS rate. DoR and TTF were reported by two trials and safety outcomes by four trials.

Table 6 Outcomes reported across the studies to be included in the NMA (CS Table 23, p. 83)

Study Name	PFS	OS	ORR	CBR	OS rate	PFS rate	DoR	TTF	Safety
FALCON trial[8]	✓	-	✓	✓	✓	✓	✓	-	✓
FIRST study[6]	✓	✓	✓	✓	✓	✓	-	✓	✓
Milla-Santos 2003[11]	✓	✓	✓	✓	✓#	✓#	-	-	✓
North American trial[9]	✓	✓**	✓	✓	✓**#	✓#	✓	✓	✓
TARGET trial*[10]	✓	✓**	-	-	✓**#	✓	-	-	-
PO25 trial*[12]	✓	✓	✓	✓	-	-	-	-	-

OS: Overall survival; PFS: Progression-free survival; ORR: Overall response rate; CBR: Clinical benefit rate; DoR: Duration of response; TTF: Time to treatment failure
*Studies reporting subgroup data of interest
**OS data reported from combined analysis of North American trial and TARGET trial and not for individual studies
#Data reported graphically and were captured using Engauge software

Only PFS and OS were selected for the NMA. The CS does not provide a rationale for selecting just these two outcomes, but the ERG presumes this was because only these outcomes contribute to the economic model inputs. This is therefore considered reasonable.

The CS states that the heterogeneity of the six studies forming the evidence network was assessed for the PFS and OS outcomes and a discussion of the results for heterogeneity assessment for each comparison per outcome is provided (CS p. 83).

For the NMA of PFS, only one of the six studies (FALCON) actually reported PFS, the remainder reported TTP. The definitions of PFS and TTP are presented in a table (CS Table 24 p. 84). Theoretically TTP differs from PFS in that it should be defined as the time from randomisation until objective tumour progression i.e. it does not include deaths. However, in practice, death (either from breast cancer or from any cause) is often also counted as an event.[20] All the definitions of TTP for the five trials reporting this outcome are provided in the CS and include deaths. Death was specified as being due to any cause in the FALCON, FIRST and PO25 trials, but the types of death contributing to TTP in the Milla-Santos, North American and TARGET trials were not described. The Milla-Santos trial did not provide the starting point for the assessment of TTP, whereas in the other five trials the interval was calculated from the time of randomisation. The ERG agrees that the definitions of PFS and TTP for the six studies

are similar, but notes that Milla-Santos only calculated TTP for patients with a clinical benefit (CR or PR or stable disease ≥ 24 weeks).

The CS does not report the definitions for OS from the six studies. The ERG has checked these and for five of the studies these are the same (from random assignment to death from any cause) whereas the Milla-Santos trial measured OS from the beginning of treatment. The time difference between randomisation and the beginning of treatment in the Milla-Santos trial is not known, so there is the potential for OS to differ slightly from the other studies for this reason.

3.1.7.2 Methodological quality of NMA trials

The CS summarises the methodological and clinical characteristics of the participants in the six studies available for inclusion in the NMA in CS Table 25 and Tables 26 to 28 respectively (p. 85 and pp. 87 to 89).

In addition to the methodological summary presented in CS Table 25 (p. 85) a critical appraisal of the trials is presented in a table earlier in the CS (CS Table 12, p. 44-45). The ERG have done its own critical appraisal of the trials, based on NICE criteria, and this is presented alongside that of the company in Table 7. No judgements of a high risk of bias were made however for many items the judgement was 'unclear' because the details necessary to determine the risk of bias were not reported in the published papers.

Table 7 Quality assessment of the included NMA studies

(see section 3.1.4 Table 5 for the QA of the FIRST and FALCON trials)

QA Criteria for RCT	Milla-Santos[11]		North American[9]		TARGET[10]		PO25 trial[12]	
1. Was the method used to generate random allocations adequate?	CS:	Not clear	CS:	Adequate[a], low risk of bias	CS:	Adequate[a], low risk of bias	CS:	Adequate[b], low risk of bias
	ERG:	Unclear	ERG:	Unclear	ERG:	Unclear	ERG:	Yes, low risk of bias
ERG comments: The North American and TARGET papers[9,10] do not state the method used to generate random allocations. [a] Central randomisation; [b] Computer generated randomisation								
2. Was the allocation adequately concealed?	CS:	Grade B	CS:	Grade A	CS:	Grade A	CS:	Grade A
	ERG:	Unclear	ERG:	Unclear	ERG:	Unclear	ERG:	Unclear
ERG comments: None of the studies provide the information needed to determine whether allocation was adequately concealed.								
3. Were the groups similar at the outset of the trial in terms of prognostic factors, e.g. severity of disease?	CS:	Low risk	CS:	Low risk	CS:	Low risk	CS:	Low risk
	ERG:	Yes, low bias risk	ERG:	Unclear	ERG:	Unclear	ERG:	Yes, low risk of bias
ERG comments: The North American trial and Target trial publications state~~s~~ that the groups were well balanced however some differences are apparent in baseline characteristics, but these are not discussed in the papers (e.g. in both trials differences in proportions with metastatic disease at different sites such as the liver, and in the North American trial an approximate 9% difference in the proportion with measurable disease).								
4. Were the care providers, participants and outcome assessors blind to treatment allocation? If any of these people were not blinded, what might be the likely impact on the risk of bias (for each outcome)?	CS:	Not clear	CS:	Double blind, low risk	CS:	Double blind, low risk	CS:	Double blind, low risk
	ERG:	Unclear	ERG:	Yes, low risk of bias.	ERG:	Yes, low risk of bias.	ERG:	Yes, low risk of bias.

ERG comments: The North American, TARGET and PO25 studies had appropriate placebo tablets in each arm (i.e. intervention A + placebo B; placebo A + intervention B). In these three trials it was unclear if outcome assessors were blinded. PO25 trial – states that internal Novartis data evaluation committee reviewed in a blinded fashion all tumour assessment and overall response data.								
5. Were there any unexpected imbalances in drop-outs between groups? If so, were they explained or adjusted for?	CS:	Low risk	CS:	Not clear	CS:	Not clear	CS:	Not clear
	ERG:	No, low bias risk	ERG:	Unclear	ERG:	Unclear	ERG:	No, low risk of bias
ERG comments: Milla-Santos flow diagram indicates all patients completed the trial. North American & Target trials list the reasons why participants might be withdrawn but it is unclear if all withdrawals are described in the papers and there are no flow diagrams. PO25 does not present a flow diagram, however all participants appear to be accounted for with those excluded from the ITT analysis appearing balanced between groups.								
6. Is there any evidence to suggest that the authors measured more outcomes than they reported?	CS:	Not clear	CS:	Not clear	CS:	Not clear	CS:	Not clear
	ERG:	Unclear	ERG:	Unclear	ERG:	Unclear	ERG:	Unclear
ERG comments: The protocols for these studies were not referenced and so it has not been possible to ascertain whether there were any measured outcomes that were not reported in the published papers or indeed more outcomes than stated in the protocol.								
7. Did the analysis include an ITT analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	CS:	Low risk	CS:	Low risk	CS:	Low risk	CS:	Low risk
	ERG:	Unclear; Unclear	ERG:	Yes, low risk of bias; Unclear	ERG:	Yes, low risk of bias; Unclear	ERG:	Yes, low risk of bias; Not applicable
ERG comments: Milla-Santos - CS states that the safety and efficacy analysis was done using ITT population, but this is not reported in the publication. In Trial PO25 all participants appear to be accounted for i.e. no missing data.								

CS judgements taken from CS Table 12, p. 44 and Table 25, p. 85

3.1.7.3 Assessment of NMA heterogeneity

The ERG has summarised key trial characteristics in Table 8.

Table 8 Summary characteristics of trials included in the NMA

Trial	FIRST[5-7]	FALCON[8]	Milla-Santos[11]	North American[9]	TARGET[10]	PO25[12,13]
Trial design	PHASE II open-label multicentre RCT	PHASE III double blind multicentre RCT	Phase III single centre RCT (blinding unclear)	PHASE III double blind multicentre RCT	PHASE II double blind multicentre RCT	PHASE III double blind multicentre RCT
Enrolled population	Postmenopausal women with ER+ &/or PgR+ ABC[a]	Postmenopausal women with ER+ &/or PgR+, HER-ve ABC[b]	Postmenopausal women with ER+ ABC[c]	Postmenopausal women with ER+ &/or PgR+ or of unknown receptor status ABC[d]	Postmenopausal women with ER+ &/or PgR+ or of unknown receptor status ABC[d]	Postmenopausal women with ER+ &/or PgR+ or of unknown receptor status ABC[e]
Number of centres (location)	62 (Brazil, Bulgaria, Czech Republic, France, Italy, Poland, Spain, United Kingdom, & United States).	113 (20 countries in Asia, Europe, North America, South America & South Africa).	1 (NR, but presumed to be Spain)	97 (USA & Canada)	83 (Europe, Australia, New Zealand, South America, & South Africa)	201 (29 countries. Countries NR)

Trial	FIRST[5-7]	FALCON[8]	Milla-Santos[11]	North American[9]	TARGET[10]	PO25[12,13]
Intervention (n in arm)	Fulvestrant (n=102)	Fulvestrant (n=230)	Anastrazole (n=121)	Anastrazole (n=171)	Anastrazole (n=340)	Letrazole (n=453)
Comparator (n in arm)	Anastrazole (n=103)	Anastrazole (n=232)	Tamoxifen, 40 mg (n=117)	Tamoxifen, 20 mg (n=182)	Tamoxifen, 20 mg (n=328)	Tamoxifen, 20 mg (n=454)
Primary outcome	CBR	PFS	NR[f]	TTP & ORR	TTP & ORR	TTP

ABC, advanced breast cancer; CBR, clinical benefit rate; ER+, oestrogen receptor positive; n, number; NR, not reported; ORR, objective response rate; PFS, progression-free survival; PgR+, progesterone receptor postive; RCT, Randomised controlled trial; TTP, time to progression.

[a] who had either, never received endocrine therapy for advanced disease or had received previous endocrine therapy for early disease completed ≥12 months prior to randomisation.

[b] who had never received endocrine therapy for breast cancer.

[c] who had not had previous therapy for advanced disease and no previous hormonal adjuvant therapy.

[d] Prior adjuvant chemotherapy or hormonal therapy for early BC was permitted, provided tamoxifen had not been received within 12 months prior to randomisation.

[e] No prior endocrine therapy for advanced breast cancer was permitted. Patients with disease relapse or recurrence during adjuvant anti-oestrogen therapy or who were within 12 months of completing such therapy were excluded.

[f] Not reported but main endpoints stated as overall response and clinical benefit.

The clinical characteristics summarised in CS tables 26 to 28 highlight some differences between the trial populations which are summarised in the CS on p. 86. The ERG presents selected characteristics in Table 9 and the ERG's view is summarised below:

- Measurable disease at baseline - in the North American trial the proportion of participants with measurable disease at baseline was lower than in most other studies (less than 80% in both arms), whereas it was approximately 85% or above in four of the other studies (CS Table 26 states not reported for PO25, but the ERG believes it can be deduced from details provided in the paper[12] that approximately 85% or more had measurable disease).
- Metastatic disease at baseline - a lower proportion of participants had metastatic disease at baseline in the FIRST and FALCON trials than in the other studies.
- Metastatic sites - participants in the FALCON and FIRST trials were less likely to have bone metastases than participants in the other trials, whereas those in the TARGET trial were less likely to have visceral metastases than the other trials.
- Endocrine therapy naïve - just fewer than 75% of the FIRST trial participants were endocrine therapy naïve. In the other studies 80% or more of the participants were endocrine therapy naïve and in two, FALCON and Milla-Santos, 99% or more were endocrine therapy naïve.
- Receptor status - the proportion of participants with breast cancer known to be HR+ where reported was lower in the North American, PO25 and TARGET trials than in the FALCON, FIRST and Milla-Santos trials where 100% of participants were known to have HR+ breast cancer.

Whilst some of the characteristics (e.g. being endocrine naive, lacking visceral metastases, absence of metastatic disease) are associated with better outcomes in individual patients, it is difficult to make cross-trial comparisons.

Table 9 Selected characteristics of patients included in the NMA trials

Trial	**FIRST[5-7]**	**FALCON[8]**	**Milla-Santos[11]**	**North American[9]**	**TARGET[10]**	**PO25[12,13]**
Age, years (median unless stated otherwise)	Ful = 66	Ful = 64	Ana = 60.2 Mean	Ana = 68	Ana = 67 Mean	Let = 65
	Ana = 68	Ana = 62	Tam = 60.6 Mean	Tam = 67	Tam = 66 Mean	Tam = 64
	Ful = 84	Ful = 87	Ana = NR[a]	Ana = 68	Ana = 89	Let = NR

Trial		FIRST[5-7]	FALCON[8]	Milla-Santos[11]	North American[9]	TARGET[10]	PO25[12,13]
Measurable disease at baseline, %		Ana = 85	Ana = 90	Tam = NR[a]	Tam = 77	Tam = 87	Tam = NR
Metastatic disease at baseline		Ful = 88	Ful = 81	Ana = 100	Ana = 99	Ana = 99	Let = 93
		Ana = 86	Ana = 83	Tam = 100	Tam = 99	Tam = 100	Tam = 92
Endocrine naive participants, %		74.6	99.4	100	80	89	82
HR status, %	**HR+**	100	100	100	89	45	66
	ER+	97	99	100	85	43	NR
	PgR+	80	77	NR	69	26	NR
	HER-ve	47	99.8	NR	NR	NR	NR

This table draws on information presented in CS Tables 26, 27 and 28 pp. 87-89
[a] The Milla-Santos publication does indicate that patients were required to have bi-dimensional measurable lesions or evaluable lytic bone metastases.
Ana, anastrozole; ER+, oestrogen receptor positive; Ful, fulvestrant; HER-ve, human epidermal growth factor receptor negative; HR, hormone receptor; Let, letrozole n, number; NR, not reported; PgR+, progesterone receptor positive; Tam, tamoxifen.

Some aspects of the clinical characteristics appear broadly similar between the trial populations (e.g. age), whereas others were not sufficiently reported across the studies to make a judgement about how similar the populations were (performance status, HER2 status, Prior surgery &/or radiotherapy, Race and disease-free interval). As previously noted, the CS states that eligibility was not restricted to populations known to be HER2 negative because HER2 receptor testing was not routinely conducted until the mid-2000s. This means that HER2-positive patients may have been included in the comparator studies and if HER2 positive patients were included and they responded less well to treatment this would disadvantage the comparator studies.

The CS does not present a 'bottom-line statement' following the assessment of heterogeneity, however as an NMA was conducted, the ERG presumes that the company judged the studies were sufficiently similar to combine. The CS does state that studies were similar with respect to age of participants, patient performance status, largely similar in race, comparable in terms of metastatic sites of the disease and the majority of patients (>70%) were endocrine treatment naïve. However, the population in PO25 may not have been as similar to the other studies (because a third of patients were not HR+).

CS p. 90 presents further considerations for NMA and at this point explains that the Milla-Santos trial was excluded from the final network (because it was the only trial to use the higher 40 mg dose of tamoxifen, which is not approved by the EMA and because its inclusion led to heterogeneity in the NMA of both the OS and PFS outcomes, which may have been because these were calculated only for patients with clinical benefit and not for all patients). Furthermore, for the OS outcome only, OS results for the North American and TARGET trials combined were available, and hence combined OS results from these trials were considered in the OS analysis. From the description provided in the CS, it also seems that the North American and TARGET trial data were combined for the PFS analysis, although this is not explicitly stated.

ERG summary on the heterogeneity of trials:

- There are some differences between the trials, both in terms of methodology and participants.
- The exclusion of the Milla-Santos trial from the network seems appropriate due to the use of the higher dose of tamoxifen in this trial and because OS and PFS were not calculated for the full trial population in this trial.
- The inclusion of FIRST, FALCON, North American, TARGET and PO25 trials appears appropriate despite some evidence of heterogeneity in trial participants between the trials, but this could potentially be accounted for by using a random effects model.

3.1.7.4 NMA statistical methods

The methods of the NMA are presented in CS section 4.10.1 (pp. 90 - 99).

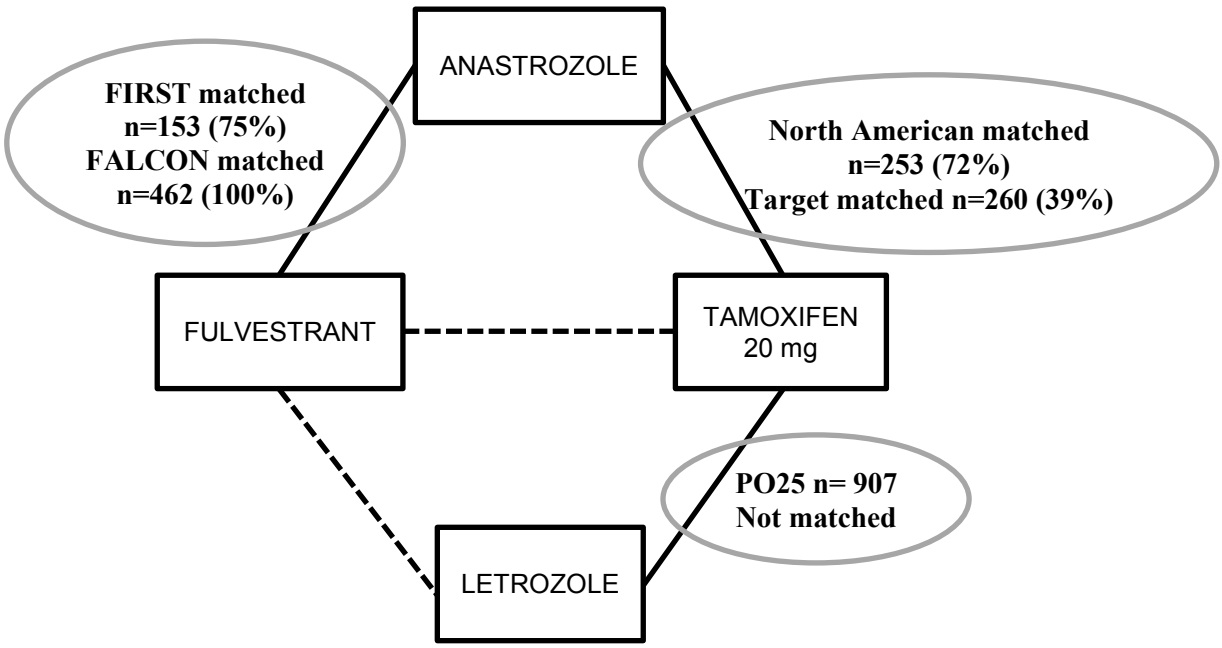
The CS states that the inclusion and exclusion criteria from the FALCON trial were applied to each treatment arm of the FIRST and NorthAmTarget trials for which patient level data were available. The reason the CS gives for this approach was so that the FIRST and NorthAmTarget trials would better match the FALCON trial population (CS pp. 90 - 91). Although not explicitly stated the ERG presumes that the reason studies were matched to FALCON (instead of another study) was because this is the pivotal phase III trial for the proposed new indication for fulvestrant. No details of the matching methodology used were presented in the CS, so the ERG and NICE sought clarification from the company. In response to clarification question A8, the company explained that criteria were applied so that only data for ER+/PgR+ patients plus endocrine treatment naive patients would be

included. The ERG assumed that the matching process would decrease the sample size of the FIRST and NorthAmTarget trials, but this was not commented on in the CS and patient demographics or clinical characteristics for these trials following the matching process were not presented. The ERG and NICE therefore asked the company to provide this information in clarification question A9. The company were able to supply information regarding the number of study participants retained after matching and this is summarised below (Table 10 and Figure 3). As can be seen the effect of matching in decreasing the numbers of participants included in the NMA was most pronounced for the TARGET trial where only 39% of participants met the matching criteria. The company do not comment on this but the ERG believes that this is because for 55% of participants in TARGET had unknown ER and unknown PgR status.[10]

Table 10 Patient numbers in the NMA studies before and after matching to the FALCON trial.

	Trial									
	FALCON		FIRST		North American		TARGET		NorthAm-Target	
Treatment arm	Ful	Ana	Ful	Ana	Ana	Tam	Ana	Tam	Ana	Tam
ITT population, n	230	232	102	103	171	182	340	328	511	510
Matched to FALCON, n (%)	230 (100)	232 (100)	73 (72)	80 (78)	119 (70)	134 (74)	132 (39)	128 (39)	251 (49)	262 (51)

This table draws on information presented in the company's written response to clarification question A9, Table 3. Note trial PO25 is not included because only aggregate data were available.
Ana, anastrozole; Ful, fulvestrant; ITT, intention-to-treat; Let, letrozole; n, number; Tam, tamoxifen



Interventions are shown in rectangular boxes, available trial evidence is shown in oval shapes. Solid lines indicate direct comparisons, dashed lines indicate indirect comparisons. Numbers and proportion of participants remaining after matching are provided for the FIRST, North American and Target studies.

Figure 3 Network of final set of studies and patient numbers after matching

As noted above, although a matching process with the FALCON trial was undertaken, the CS does not report the demographic and clinical characteristics of the matched populations. The ERG and NICE therefore sought clarification from the company (clarification question A9) and in response, baseline characteristics for the matched trial populations were provided. Where possible the ERG has compared the baseline characteristics for the matched trial populations and those for the whole trial populations reported in CS Tables 26-28. Where the ERG is confident that the definition of characteristics correspond (e.g. age, visceral disease, measurable disease) the baseline characteristics of the matched and whole trial population data are very similar. As expected, the matching process increased homogeneity between the FALCON and the FIRST and NorthAmTarget trials but heterogeneity remained with the PO25 trial for which only aggregate level data were available (Table 11).

The ERG is not aware of any published methodological guidance that addresses the issue of matching individual patient data (IPD) from one trial (in this case FALCON) to IPD from other studies (in this case FIRST and NorthAmTarget). Whilst it is clear that the matching process allows for the exclusion of participants [redacted]

[redacted] to create a more homogeneous population in the NMA the ERG is concerned about potential disadvantages, for example if matching creates scope for bias as randomisation has been broken. For this reason the ERG and NICE asked the company in clarification question A12 to provide results using all study data from FIRST and NorthAmTarget (i.e. to conduct the analysis without undertaking the matching process). The company declined to do this because approximately a third of the patients in the resulting network (560/1688) would be outside the scope for the appraisal. The ERG has concluded that, the known advantages of matching in [redacted] reducing heterogeneity in the NMA (at least for the trials that could be matched) are likely to outweigh potential disadvantages. A similar conclusion was reached in a previous STA for fulvestrant (TA239[21]) in which only a subgroup of one trial met the decision problem. The ERG for that STA believed that the advantages of decreased heterogeneity outweighed the disadvantages of reduced power.

Table 11 Baseline characteristics of participants in the FIRST and NorthAmTarget studies after matching in comparison to the unmatched PO25 study.

Characteristic n (%) unless stated otherwise	**FALCON** baseline		**FIRST** (matched to FALCON)		**NorthAmTarget** (matched to FALCON)		**PO25** baseline (no IPD, unmatched)	
	Ful n=230	Ana n=232	Ful match n=73	Ana match n=80	Ana match n=251	Tam match n=262	Let n=453	Tam n=454
Median age, years	64	62	67	69	67	66	65	64
ER+ and/or PgR+	220 (96) 10 un[a]	225 (97) 7 un[a]	73 (100)	80 (100)	251 (100)	262 (100)	294 (65) 156 un	305 (67) 149 un
Visceral disease	135 (59)	119 (51)	33 (45)	43 (54)	103 (41)	132 (50)	194 (43)	208 (46)
Bone only disease	24 (10)	24 (10)	2 (3)	2 (3)	53 (21)	50 (19)	69 (15)	72 (16)
Soft tissue only disease	8 (4)	6 (3)	0	0	53 (21)	45 (17)	113 (25)	116 (25)
No prior chemo	151 (66)	151 (65)	63 (86)	68 (85)	191 (76)	198 (76)	320 (71)	301 (66)
Prior chemo for ABC	36 (16)	43 (19)	0	0	0	0	40 (9)	48 (11)
Prior adjuvant chemo	43 (19)	40 (17)	10 (14)	12 (15)	60 (24)	65 (25)	93 (21)	105 (23)
Prior endocrine therapy[b]	2 (1)	1 (0.4)	0	0	0	0	84 (19)[c]	83 (18)[c]
Measurable disease	193 (84)	196 (84)	69 (95)	78 (98)	195 (78)	208 (79)	-	-
Locally advanced	28 (12)	32 (14)	19 (26)	18 (23)	-	-	145 (32)	146 (32)

This table draws on information presented in the company's written response to clarification question A9, Tables 4,5,6 and 9.

Grey shading indicates the characteristics that were matched. +ve, positive; ABC, advanced breast cancer; Ana, anastrozole; ER, oestrogen receptor; Ful, fulvestrant; IPD, individual patient data, Let, letrozole; n, number; PgR, progesterone receptor; Tam, tamoxifen; un, unknown

[a] The patients noted as being unknown are, according to the published paper ER+ and PgR unknown. Therefore the ERG believes that all FALCON participants are HR+; [b] Adjuvant endocrine therapy for early disease; [c] labelled as prior adjuvant anti-oestrogen therapy in the PO25 trial.

After the inclusion and exclusion criteria of FALCON had been applied to FIRST and NorthAmTarget, Kaplan-Meier (KM) plots of PFS and OS were produced for the matched subgroups of participants. For PO25 the published KM plots for the whole study population were digitised and then patient-level data were reconstructed using a published algorithm.[22]

The OS and PFS data were examined to determine whether there was a constant relative treatment effect over time by visual inspection of the KM plots for PFS (CS Figure 23, p. 92) and OS (CS Figure 24, p. 93), and visual inspection of log cumulative hazard plots for PFS (CS Figure 25, p. 98) and OS (CS Figure 26, p. 99) for each trial. The OS KM plots for the arms of the PO25 trial and the NorthAmTarget trial cross, suggesting that a constant relative treatment effect is unlikely in these studies. In the log cumulative hazard plots, a constant relative treatment effect (i.e. proportional hazards) could be assumed if the two lines for each trial run parallel to each other, but this is not the case for all studies.

The company therefore concluded that methods for NMA that rely on the assumption of proportional hazards being true would be inappropriate. The method used for NMA is one developed by Ouwens et al.[23] In this method the differences in the shape and scale parameters of the parametric survival function used to model PFS or OS between the intervention and each comparator over time are synthesised, and used both for the indirect comparison and to extrapolate the PFS and OS curves beyond the end of trial follow-up (see Section 4.3.5). The parametric distributions used to model the KM data were the Weibull, Gompertz, log-logistic, lognormal and generalised gamma. Although not explicitly stated in CS Section 4.10.1 (pp. 90 - 99), the ERG assumes that the analysis took a Bayesian approach using a Markov Chain Monte Carlo method implemented using the WinBUGs software package (as described by Ouwens et al.[23]).

The shape and scale parameters were calculated for the baseline (reference), which was anastrozole. These baseline parameters were then used as the anchor to obtain the estimates for the shape and scale of the other interventions in the network (i.e. fulvestrant, letrozole and tamoxifen).

If the shape parameter is regarded as fixed between treatment arms, this effectively assumes a proportional treatment effect. This ‘no shape arm’ model was tested in sensitivity analysis for all but the generalised gamma model (which, as a three parameter model, was more complex and therefore not included).

A fixed-effect meta-analysis was undertaken. The rationale for not including a random effects model was the limited number of trials in each network. Whilst the ERG agrees that the number of trials is limited, as noted earlier there is some evidence of heterogeneity in trial participants between the trials, which the ERG thought could potentially be accounted for by using a random effects model. The ERG and NICE therefore asked the company to provide results from a random effects NMA (clarification question A10). In response to clarification question A10, the company provided a more detailed explanation of the reasons why a random effects NMA could not be undertaken. Due to the presence of the pooled NorthAmTarget dataset, the only connection in the network where there are two or more trials is the fulvestrant-anastrozole comparison informed by the FALCON and FIRST trials (as shown in Figure 2). The company cites a recent (2016) paper by Friede et al.[24] which states that, in the Bayesian framework, if the number of studies is small then the choice of prior for the between-trial standard deviation is critical. The company goes on to state that an attempt was made to identify an informative prior (as detailed in the response to clarification question A10) but this proved a "*difficult question to answer*" and therefore they concluded, as before, that the more robust approach was to use a fixed effect meta-analysis.

The ERG accepts that there are few trials in the network and that, with the methodology the company have used for the NMA, a random-effects NMA is not possible, the ERG nevertheless is concerned that the potential uncertainty around the effect estimates may not be adequately represented.

The final consideration regarding the NMA is that for the PO25 trial IPD were not available and thus this trial population could not be matched to the FALCON inclusion and exclusion criteria. Furthermore, crossovers between treatments occurred in this trial which may have confounded the survival analysis and there is now a general agreement that the efficacy and safety of anastrozole and letrozole are equivalent [e.g. NICE CG81[2] states "All aromatase inhibitors appear to be equally effective in terms of primary outcome (overall survival)"]. For these reasons a scenario analysis in the economic model assumes the efficacy of letrozole is equivalent to anastrozole by using the anastrozole curves for letrozole (i.e. efficacy data from the PO25 trial is not used). The ERG had an additional concern regarding whether the data for TTP and OS came from the whole PO25 population or the HR+ subgroup (66%) and clarification was requested on this by the ERG and NICE. The company confirmed in their response to clarification question A6 that data from the full study population were used. Because of the differences between the PO25 study and the others in the network, and the general agreement that the efficacy of anastrozole and letrozole are equivalent, the ERG and NICE requested in clarification question A13 that an analysis omitting study PO25 from

the network be conducted and the company complied with this request (results are discussed in section 3.3).

3.1.7.5 ERG summary of the company's approach to evidence synthesis

The ERG agrees that, in the absence of RCTs comparing fulvestrant with the other comparators of interest, letrozole and tamoxifen, conducting an NMA to enable indirect comparisons between fulvestrant versus letrozole and fulvestrant versus tamoxifen is appropriate.

The ERG believes that the company has identified appropriate published data sources for the NMA. Six trials are included: two RCTs, FIRST and FALCON, of fulvestrant versus anastrozole; three RCTs, of anastrozole versus tamoxifen; and one RCT of letrozole versus tamoxifen. The selection of outcomes for analysis (PFS and OS) is reasonable.

It was difficult to determine the potential for bias in the four trials providing evidence anastrozole versus tamoxifen and letrozole versus tamoxifen because many of the details necessary to determine risk of bias were not reported in the published papers.

There was evidence of differences between the trials both in terms of methodology and participants. The ERG agrees that the exclusion of the Milla-Santos trial from the network is appropriate.

The availability of IPD enabled data from the FIRST and NorthAmTarget trials to be matched to the pivotal FALCON trial population. This enabled participants from the FIRST and NorthAmTarget trials [redacted] to be excluded from the network creating a more homogeneous population in the NMA. Matching was not possible for the PO25 study because only aggregate data were available.

The company used appropriate methods to determine whether the assumption of proportional hazards was true. A constant relative treatment effect was unlikely for some of the studies and therefore a method was used that did not rely on the assumption of proportional hazards.

A fixed-effect meta-analysis was undertaken. The ERG accepts the company explanation that a random-effects NMA was not possible (predominantly due to the low number of trials).

However, this may mean that the potential uncertainty around effect estimates is not adequately represented.

3.2 Summary statement of company's approach

The ERG's quality assessment of the company's SLR in the CS is summarised in Table 14. Processes for inclusion or exclusion of studies were conducted by two independent reviewers (CS p. 37), while extraction of included studies was carried out in parallel by two independent reviewers and any discrepancies between them were reconciled by a third independent reviewer (CS p. 41). Included studies were subject to critical appraisal. Overall, the ERG considers the study selection, data extraction and critical appraisal processes to have been adequate, following standard accepted review methodology.

The ERG concludes that the submitted evidence reflects the decision problem defined in the CS and that the overall risk of systematic error in the systematic review appears to be low.

Table 12 Quality assessment (CRD criteria) of CS review

CRD Quality Item: score Yes/ No/ Uncertain with comments	
1. Are any inclusion/exclusion criteria reported relating to the primary studies which address the review question?	1. Yes. Inclusion and exclusion criteria are clearly stated.
2. Is there evidence of a substantial effort to search for all relevant research? i.e. all studies identified	2. Yes. There was a substantial effort to search for all relevant studies and the restriction of the evidence to English Language only is unlikely to have resulted in any missed studies.
3. Is the validity of included studies adequately assessed?	3. Yes. Quality assessment of all of the included trials (including the four comparator trials from the NMA), was assessed using the NICE criteria. Generally, the ERG assessment agreed with the company assessment of the fulvestrant trials, with differences mainly due to insufficient reporting of details, preventing a judgement to be made. This was also the case for the quality assessment for the comparator trials included in the NMA.

4. Is sufficient detail of the individual studies presented?	4. Yes. Methodology, patient characteristics and outcomes of the included studies are generally presented in sufficient detail, although the ERG extracted some additional information from the trial publications. Most of the information for the fulvestrant trials was presented in a separate format, making an overview difficult. The ERG presents combined tables to aid with the interpretation of the two trials and their results.
5. Are the primary studies summarised appropriately?	5. Yes. The primary studies are summarised appropriately for both the fulvestrant trials and the NMA trials, with the majority of details provided in tables and figures in the main body of the CS.

3.3 Summary of submitted evidence

In this section the ERG focuses on the main outcomes of the included phase II FIRST trial and the phase III FALCON trial presented in the CS, supplemented with data from other sources (e.g. trial publications and clinical study reports) if necessary. The outcomes from the NMA are also included in this section.

Where evidence feeds into the economic model this is indicated and cross-references are provided to the economic section of the ERG report.

3.3.1 Summary of PFS (FALCON, primary outcome) and TTP (FIRST, secondary outcome)

The CS presents the PFS results for the FALCON trial (which was the primary outcome for this trial, CS p. 70) and TTP results for the FIRST trial (where TTP was a secondary outcome, CS p. 68). These analyses were undertaken when approximately 306 progression events had occurred in FALCON, and when 75% of patients had discontinued (failed) study treatment in FIRST. In both trials the proportion of events occurring in the fulvestrant arm was lower than that in the anastrozole arm (Table 13) over the study period (approximately 36 months since randomisation in the FALCON trial and approximately 42 months since randomisation in the FIRST trial).

Median PFS in the FALCON trial was 2.8 months longer in the fulvestrant arm than the anastrozole arm. Neither of the clinical experts the ERG consulted felt that this was a clinically significant increase and indicated that much larger increases could be obtained from other new drugs. The improvement in PFS with fulvestrant was statistically significant [HR = 0.797, 95% CI 0.637 to 0.999, p = 0.0486].

Median TTP in the FIRST trial was 10.3 months longer in the fulvestrant arm than the anastrozole arm. The improvement in TTP with fulvestrant would be considered clinically significant and was also statistically significant (HR 0.66, 95% CI 0.47 to 0.92, p = 0.01) (Figure 5).

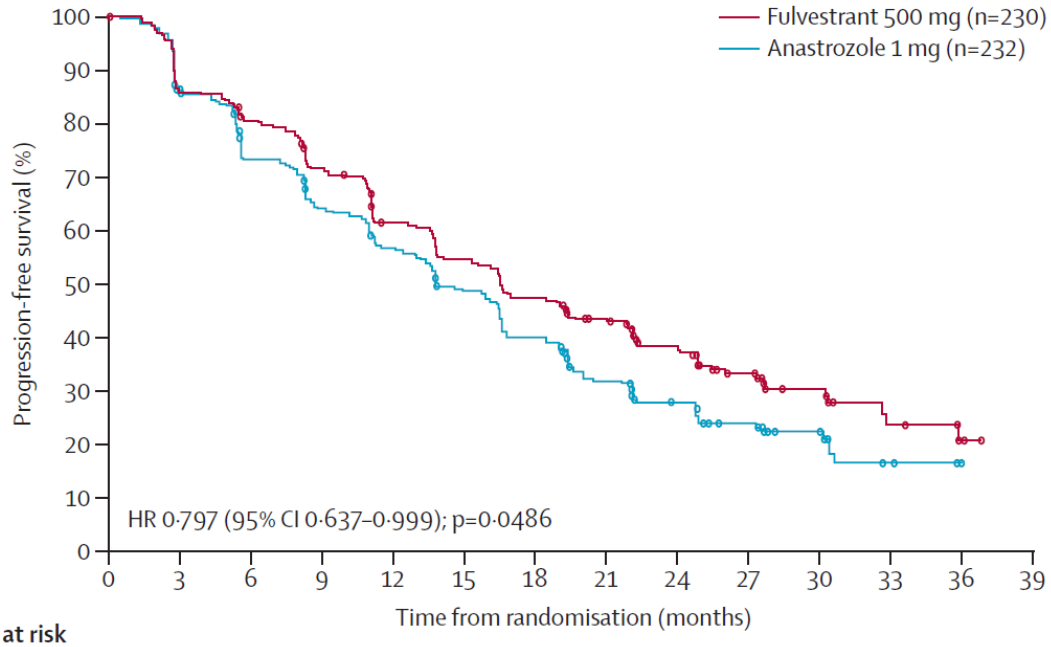
Table 13 PFS results for the FALCON trial and time to progression results for the FIRST trial

	FALCON (PFS)[a]		FIRST (TTP)[a]	
	Fulvestrant (n=230)	**Anastrozole (n=232)**	**Fulvestrant (n=102)**	**Anastrozole (n=103)**
Events, n (%)	143[b] (62)[b]	166[b] (72)[b]	63[b] (61.8)	79[b] (76.7)
HR, (95% CI) p-value	0.797, (0.637 to 0.999) 0.0486		0.66, (0.47 to 0.92) 0.01	
Median PFS (FALCON) or median TTP (FIRST) (95% CI), months	16.6 (13.83 to 20.99)	13.8 (11.99 to 16.59)	23.4	13.1

CI, confidence interval; HR, hazard ratio; n, number; PFS, Progression-free survival; TTP, Time to progression

[a] Median duration of follow-up for PFS in the FALCON study was not reported. Median follow-up for TTP in the FIRST study was 18.8 months in the fulvestrant group and 12.9 months in the anastrozole group.

[b] These data were not reported in the CS but were obtained from the published papers for FALCON[8] and FIRST.[5]

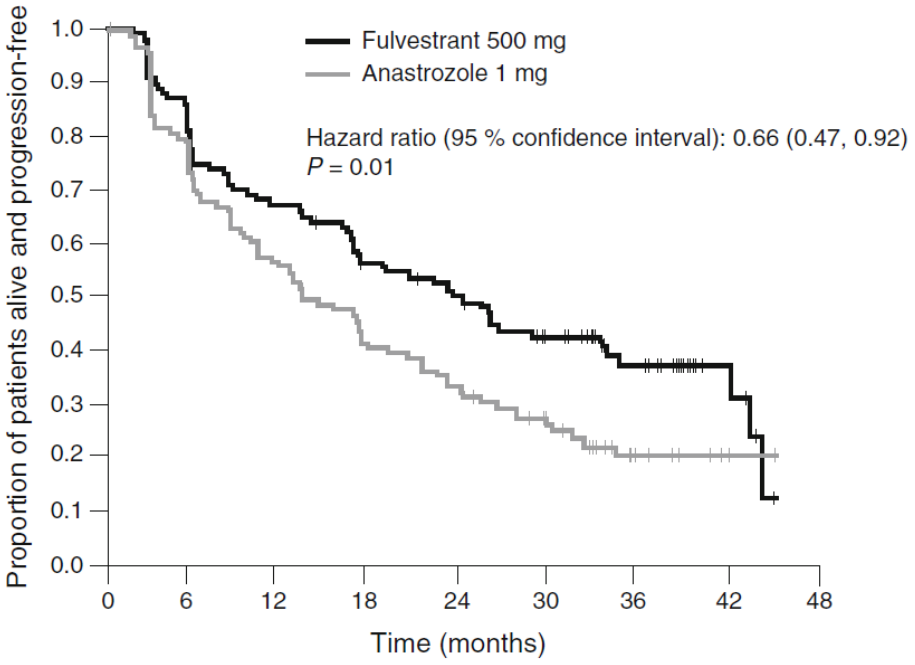


Number at risk

Fulvestrant 500 mg	230	187	171	150	124	110	96	81	63	44	24	11	2	0
Anastrozole 1 mg	232	194	162	139	120	102	84	60	45	31	22	10	0	0

A circle represents a censored observation

Figure 4 Kaplan-Meier plot of PFS in the FALCON trial (CS Figure 14, p. 71)



Patients at risk

Fulvestrant 500 mg	102	74	65	52	45	34	20	6	0
Anastrozole 1 mg	103	69	55	39	30	21	8	2	0

After the primary data cut-off, progression was determined by investigator opinion

Figure 5 Kaplan-Meier plot of TTP in the overall FIRST population (CS Figure 12, p. 68)

KM plots of PFS for FALCON, the matched populations for FIRST and NorthAmTarget, and the full population for PO25 (for which only aggregate data were available) are presented in CS Figure 23. The ERG notes that HRs for the matched PFS data were not presented. PFS is included in the economic model (ERG report section 4.3.5.1).

The ERG notes that the degree of benefit seen with fulvestrant in the FIRST trial (median TTP 10.3 months longer in the fulvestrant arm than in the anastrozole arm) was greater than that observed in the FALCON trial (median PFS 2.8 months longer in the fulvestrant arm than in the anastrozole arm). The CS does not comment on this difference or provide any reasons for it. The ERG has checked to ensure that it is not due to differences in exposure to prior endocrine therapy (the same pattern is observed after matching the FIRST trial data to the key FALCON trial inclusion criteria prior to use in the NMA), nor due to differences in TTP/PFS outcomes in the anastrozole arms of the trials which appear to be broadly similar, nor due a difference in the proportion of events at the time of data analysis between the trials as this is also broadly similar. There is a key difference in the methodology of the two trials that may have had an impact on outcomes, which is that the FIRST trial was not a blinded study, whereas the FALCON trial was conducted with double-blinding. Finally, in the FALCON study publication[8] (but not in the CS) a suggestion is made, based on findings from subgroup analyses, that an enhanced treatment effect with fulvestrant might be seen in patients with non-visceral disease compared to those with visceral disease, but the authors of the paper caution that this observation requires further study. The company's decision problem, in line with the NICE scope[1] for this appraisal, indicates that if the evidence allows, subgroups of people with visceral disease and people with non-visceral disease will be considered. Analyses of subgroups are presented in section 3.3.6 but the CS does not discuss the outcome of these in any detail.

As described earlier (ERG report section 3.1.7), the method used for the NMA synthesises the differences in the shape and scale parameters of the parametric survival function used to model PFS between the intervention and each comparator over time. The baseline (reference) treatment is anastrozole and this is used as the anchor from which the estimates of the shape and scale for the other interventions are then obtained. As shown in Table 14, there were statistically significant differences in the scale parameter for fulvestrant and tamoxifen when compared against anastrozole for four of the five parametric distributions (Weilbull, log-logistic, lognormal and generalised gamma). Statistically significant differences were also seen in the shape parameter for fulvestrant and tamoxifen for the lognormal distribution when compared against anastrozole. Note that a positive estimate of

a difference in log scale indicates that the treatment is better than the reference and conversely, a negative estimate of a difference in log scale indicates that the treatment is worse than the reference. If both the limits of the 95% CI have the same sign this indicates that the difference between the treatment and reference is statistically significant. The ERG has added grey shading in Table 14 to indicate where the statistically significant differences are. Grey shading indicates fulvestrant is statistically significantly better than anastrozole and tamoxifen is statistically significantly worse than anastrozole.

Table 14 Fixed effect NMA PFS results: baseline parametric distribution parameters and difference from baseline for treatment alternatives versus (FALCON) anastrozole (CS Table 29 p. 101)

Weibull	Scale			Shape		
	Estimate	L95%	U95%	Estimate	L95%	U95%
Anastrozole (reference)						
	Estimate	L95%	U95%	Estimate	L95%	U95%
Fulvestrant						
Letrozole						
Tamoxifen						
Gompertz	Scale			Shape		
	Estimate	L95%	U95%	Estimate	L95%	U95%
Anastrozole (reference)						
	Estimate	L95%	U95%	Estimate	L95%	U95%
Fulvestrant						
Letrozole						
Tamoxifen						
Log-logistic	Scale			Shape		
	Estimate	L95%	U95%	Estimate	L95%	U95%
Anastrozole (reference)						
	Estimate	L95%	U95%	Estimate	L95%	U95%
Fulvestrant						
Letrozole						
Tamoxifen						
Lognormal	Scale			Shape		
	Estimate	L95%	U95%	Estimate	L95%	U95%
Anastrozole (reference)						
	Estimate	L95%	U95%	Estimate	L95%	U95%
Fulvestrant						
Letrozole						
Tamoxifen						

Generalised gamma	Scale			Shape		
	Estimate	L95%	U95%	Estimate	L95%	U95%
Anastrozole (reference)	[redacted]					
	Estimate	L95%	U95%	Estimate	L95%	U95%
Fulvestrant	[redacted]					
Letrozole	[redacted]					
Tamoxifen	[redacted]					
Common parameter	Estimate	L95%	U95%	-	-	-
Q	[redacted]					

Abbreviations: L95%, lower limit of the 95% confidence interval; PFS, progression-free survival; U95%, upper limit of the 95% confidence interval

The ERG and NICE asked the company to provide the results of pairwise comparisons (clarification question A10). The company provided these results and the comparison between the fixed-effect NMA and fixed-effect pairwise comparisons (generalised gamma model) indicates the results are almost identical (Table 15).

Table 15 Comparison of PFS results obtained from NMA and pairwise meta-analysis

Intervention	**Ref**		**Scale**	**L95%**	**U95%**	**Shape**	**L95%**	**U95%**
Fulvestrant	Ana	NMA	[redacted]					
		Pairwise	[redacted]					
Letrozole	Ana	NMA	[redacted]					
		Pairwise	[redacted]					
Tamoxifen	Ana	NMA	[redacted]					
		Pairwise	[redacted]					

This table draws on information presented in CS Table 29 and the company's written response to clarification question A10, Table 11.

Ana, anastrozole; L95%, lower limit of the 95% confidence interval; NMA, network meta-analysis; Ref, reference; U95%, upper limit of the 95% confidence interval

The ERG and NICE also asked the company to provide the results obtained from a network that omits study PO25 (clarification question A13). The company provided results of the fixed -effects meta-analyses obtained when excluding the PO25 trial (again using the ganeralised gamma model) and these results were almost identical to those reported with PO25 included in the NMA (see Clarification Responses to question A13).

3.3.2 Summary of OS (secondary outcome)

OS was a secondary outcome of the FALCON trial planned at trial inception. In contrast, in the FIRST trial, OS was added in a protocol amendment as a secondary outcome to see whether the improvement observed in TTP would translate into an improvement in OS.

The OS data for the FALCON trial are immature and consequently, at the time of data analysis, it was not possible to calculate a median OS (Table 16). Although the proportion of deaths in the fulvestrant arm is slightly lower than in the anastrozole arm (29% vs 32% respectively[8]), the 95% CI for the HR spans 1 and there is no statistically significant difference between survival in the fulvestrant and anastrozole arms at the time of the primary efficacy analysis for PFS (Figure 6), which occurred approximately 36 months after the start of randomisation. The CS states that the next survival analysis will be performed when approximately 50% of patients have died (CS p. 59). [redacted]

The analysis of OS in the FIRST trial was performed when approximately 65% of deaths had occurred. At the data cut-off, the proportion of patients who had died was lower in the fulvestrant arm than in the anastrozole arm (61.8% versus 71.8% respectively,[7] Table 16). The improvement in the fulvestrant group was statistically significant (HR 0.70, 95% CI 0.50 to 0.98, p=0.04) with a median survival of 54.1 months in comparison to 48.4 months in the anastrozole arm (Figure 7). A limitation of this result is that it comes from an analysis that was not originally specified and some patients (n=35) did not contribute data to this outcome.

Table 16 OS results for the FALCON and FIRST trials

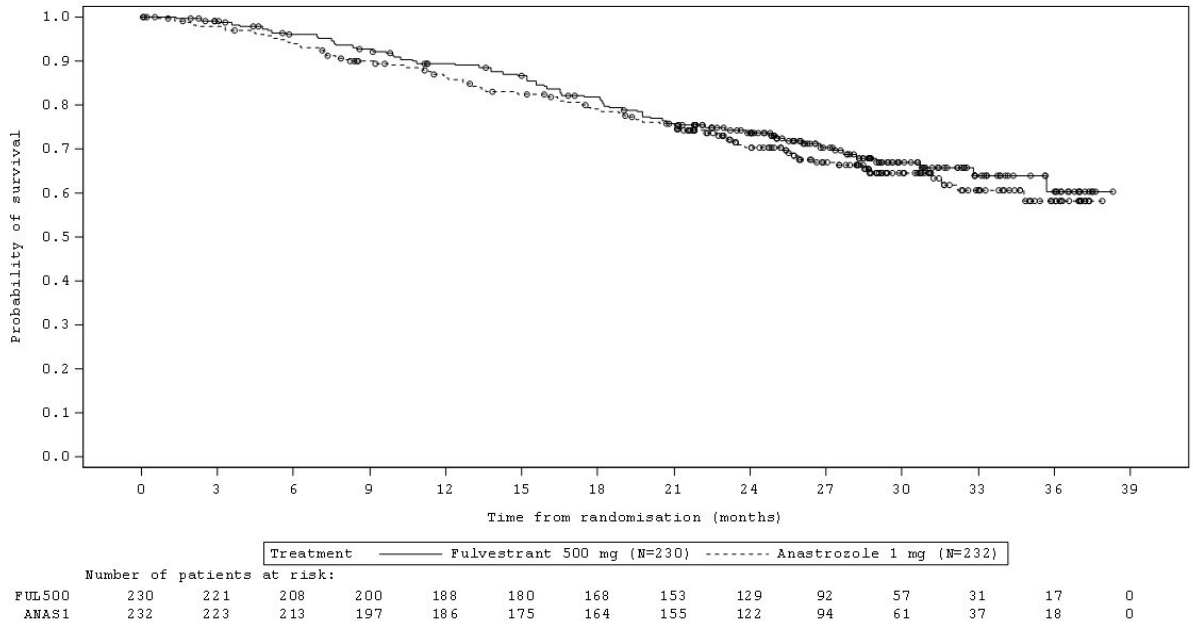
	FALCON		FIRST	
	Fulvestrant (n=230)	**Anastrozole (n=232)**	**Fulvestrant (n=102)**	**Anastrozole (n=103)**
Events, n (%)	67 (29%)[a]	75 (32%)[a]	63 (61.8%)[b]	74 (71.8%)[b]
HR, (95% CI)	0.875 (0.629 to 1.217)		0.70 (0.5 to 0.98)	
p-value	0.4277		0.04	
Median OS (95% CI), months	NC[c]	NC[c]	54.1 (NR)	48.4 (NR)

CI, confidence interval; HR, hazard ratio; n, number; NC, Not calculated; NR, Not reported; OS, Overall survival;

[a] These data were not presented in the CS so have been obtained from the published paper.[8]

[b] The CS presents rounded percentage values only so these data come from the published paper[7]

[c] Median overall survival could not be calculated because of insufficient follow-up time (31% maturity).



Note: A circle represents a censored observation
ANAS1: Anastrozole 1 mg; FUL500: fulvestrant 500 mg; OS: Overall survival; PFS: Progression-free survival

Figure 6 Kaplan-Meier plot of OS in the FALCON trial at the time of the PFS analysis (CS Figure 15, p. 72)

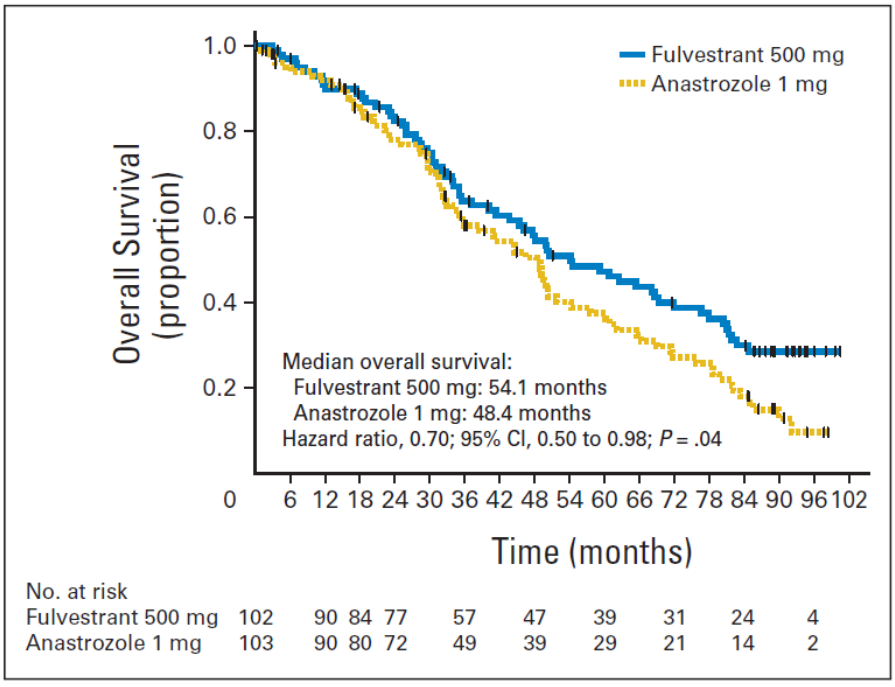


Figure 7 Kaplan-Meier plot of OS in the FIRST population (CS Figure 13 p. 69)

Using the same methodology described above for PFS, an indication of the effectiveness of fulvestrant and anastrozole in comparison to the other comparators letrozole and tamoxifen

was obtained by NMA using data from clinical trials identified by the company's systematic review and matched (where IPD were available) to include ER+/PgR+ participants plus endocrine treatment naive participants. KM plots of OS for FALCON, the matched populations for FIRST and NorthAmTarget, and the full population for PO25 (for which only aggregate data were available) are presented in CS Figure 24. The ERG notes that HRs for the matched OS data were not presented. Furthermore it should also be borne in mind that i) data from the FALCON trial are immature (and extend to approximately 36 months after baseline) and ii) OS analysis for the matched FIRST data extend to approximately 96 months, therefore the majority of the long-term OS data for fulvestrant come from the FIRST trial. OS is included in the economic model (ERG report section 4.3.5.1).

As shown in Table 17, there was a statistically significant difference in the scale parameter for letrozole when compared against anastrozole for the Gompertz distribution and in the shape parameter for four of the five distributions (Weibull, Gompertz, loglogistic and generalised gamma) when compared against anastrozole. The ERG has added grey shading to Table 17 to indicate where the statistically significant differences are. Grey shading indicates where letrozole is statistically significantly different in comparison to anastrozole.

Table 17 Fixed effect network meta-analysis OS results: baseline parametric distribution parameters and difference from baseline for treatment alternatives versus (FALCON) anastrozole (CS Table 30, p. 105)

Weibull	Scale			Shape		
	Estimate	L95%	U95%	Estimate	L95%	U95%
Anastrozole (reference)						
	Estimate	L95%	U95%	Estimate	L95%	U95%
Fulvestrant						
Letrozole						
Tamoxifen						
Gompertz	Scale			Shape		
	Estimate	L95%	U95%	Estimate	L95%	U95%
Anastrozole (reference)						
	Estimate	L95%	U95%	Estimate	L95%	U95%
Fulvestrant						
Letrozole						
Tamoxifen						

Log-logistic	Scale			Shape		
	Estimate	L95%	U95%	Estimate	L95%	U95%
Anastrozole (reference)	[redacted]					
	Estimate	L95%	U95%	Estimate	L95%	U95%
Fulvestrant	[redacted]					
Letrozole	[redacted]					
Tamoxifen	[redacted]					
Lognormal	Scale			Shape		
	Estimate	L95%	U95%	Estimate	L95%	U95%
Anastrozole (reference)	[redacted]					
	Estimate	L95%	U95%	Estimate	L95%	U95%
Fulvestrant	[redacted]					
Letrozole	[redacted]					
Tamoxifen	[redacted]					
Generalised gamma	Scale			Shape		
	Estimate	L95%	U95%	Estimate	L95%	U95%
Anastrozole (reference)	[redacted]					
	Estimate	L95%	U95%	Estimate	L95%	U95%
Fulvestrant	[redacted]					
Letrozole	[redacted]					
Tamoxifen	[redacted]					
Common parameter	Estimate	L95%	U95%	-	-	-
Q	[redacted]					

Abbreviations: L95%, lower limit of the 95% confidence interval; OS, overall survival; U95%, upper limit of the 95% confidence interval

The company provided results of pairwise comparisons for OS in response to ERG and NICE clarification question A10. The comparison between the fixed-effect NMA and fixed-effect pairwise comparisons (Weibull model) indicates the results are almost identical (Table 18).

Table 18 Comparison of OS results obtained from NMA and pairwise meta-analysis

Intervention	**Ref**		**Scale**	**L95%**	**U95%**	**Shape**	**L95%**	**U95%**
Fulvestrant	Ana	NMA	[redacted]					
		Pairwise	[redacted]					
Letrozole	Ana	NMA	[redacted]					
		Pairwise	[redacted]					
Tamoxifen	Ana	NMA	[redacted]					
		Pairwise	[redacted]					

This table draws on information presented in CS Table 30 and the company's written response to clarification question A10, Table 12.
Ana, anastrozole; L95%, lower limit of the 95% confidence interval; NMA, network meta-analysis; Ref, reference; U95%, upper limit of the 95% confidence interval

As stated earlier the ERG and NICE also asked the company to provide the results obtained from a network that omits study PO25 (clarification question A13). The company provided results of the fixed -effects meta-analyses obtained when excluding the PO25 trial (using the Weibull model) and these results were almost identical to those reported with PO25 included in the NMA (see Clarification Responses to question A13).

3.3.3 Summary of clinical benefit rate for the FALCON trial (secondary outcome) and the FIRST trial (primary outcome)

CBR was a secondary outcome of the FALCON trial (Table 19). Although a higher proportion of participants in the fulvestrant arm had a clinical benefit compared to the anastrozole arm (78% versus 74% respectively) the comparison of the CBR between the arms was not statistically significant (OR 1.25, 95% CI 0.82 to 1.93, p = 0.3045). There was an increase in the EDoCB (21.9 months for fulvestrant versus 17.5 months for anastrozole), but again the comparison between arms did not reach statistical significance (EDoCB ratio 1.26, 95% CI 0.99 to 1.59, p-value = 0.0561).

CBR was the primary outcome of the FIRST trial. However, it must be remembered that this trial was designed as a non-inferiority trial and was therefore not powered to detect a statistically significant difference in CBR. A higher proportion of participants in the fulvestrant arm had a clinical benefit compared to the anastrozole arm (72.5% versus 67% respectively) the comparison of the CBR between the arms was not statistically significant (OR 1.30, 95% CI 0.72 to 2.38, p = 0.386) (Table 19). A blinded independent review of the response data was carried out on 190 records (95 from each trial arm) and concordance was above 86% in both arms (88.4% concordance in the fulvestrant arm and 86.3% in the anastrozole arm). The CBRs calculated after blinded independent review (CBR 69.5% fulvestrant versus 66.3% anastrozole) were similar to those obtained from the investigator's evaluation (Table 19).

Table 19 CBR for the FALCON and FIRST trials (Full analysis sets)

	FALCON		FIRST	
	Fulvestrant (n=230)	**Anastrozole (n=232)**	**Fulvestrant (n=102)**	**Anastrozole (n=103)**
Clinical benefit, n (%)				
CR	7 (3%)[a]	8 (3%)[a]	0	1 (1.0)[b]
Partial response	86 (37%)[a]	82 (35%)[a]	32 (31.4)	32 (31.1)
Stable disease ≥24 weeks	87 (38%)[a]	82 (35%)[a]	42 (41.2)[b]	36 (35)
Total with clinical benefit	180 (78)	172 (74)	74 (72.5)	69 (67)
CBR odds ratio (95% CI), p-value	1.25 (0.82 to 1.93), 0.3045		1.30 (0.72 to 2.38), 0.386	
Absolute difference (95% CI)	NR		5.6% (-7.8 to 15.8%)[b]	
EDoCB, months	21.9	17.5	NR	NR
EDoCB ratio (95% CI), p-value	1.26 (0.99 to 1.59), 0.0561		NR	
No clinical benefit, n (%)				
Stable disease <24 weeks	9 (4)	22 (9)	15 (14.7)	12 (11.7)
Progression RECIST progression Death	30 (13) 27 (12) 3 (1)	33 (14) 28 (12) 5 (2)	10 (9.8)[b]	20 (19.4)[b]
Not evaluable	11 (5)	5 (2)	3 (2.9)	2 (1.9)
Total with no clinical benefit	50 (22)	60 (26)	28 (27.5)	34 (33)

This table draws on information presented in CS Table 18, p. 67; ; Table 20, p. 76.

CI, confidence interval; EDoCB, Expected duration of clinical benefit; n, number; RECIST, Response Evaluation Criteria in Solid Tumors

[a] These data are not presented in the CS but were available in the published paper by Robertson et al. 2016[8]

[b] There appeared to be several typographical errors in CS Table 18 p. 67 and some data had been rounded in the CS. These data are taken from Table 2 in the published paper by Robertson et al. 2009.[6]

3.3.4 Summary of other secondary outcomes reported for the FALCON and FIRST trials

Secondary outcomes reported by both the FALCON[8] and FIRST[6] trials were based on response to treatment (Table 20). The ORR (defined as the best overall response of CR or PR) was broadly the same in the fulvestrant and anastrozole groups in both trials, as were proportions of participants with stable and progressive disease. In the FALCON trial, although the median time to onset of response was longer in the fulvestrant arm (8.1 months versus 5.6 months in the anastrozole arm), the median DoR was numerically longer. The proportions of patients responding to treatment and the mean DoR in responding patients was used to calculate the EDoR according to the method described by Ellis and colleagues.[25] The EDoR was numerically higher for the fulvestrant arm than the anastrozole arm and the EDoR ratio favoured fulvestrant . A similar effect was seen with the EDoCB.

Table 20 Additional secondary outcomes for the FALCON (for patients with measurable disease at baseline) and FIRST (evaluable for response analysis set) trials

	FALCON		FIRST	
	Fulvestrant (n=193)	**Anastrozole (n=196)**	**Fulvestrant (n=89)**	**Anastrozole (n=93)**
ORR, n (%)	89 (46)	88 (45)	32 (36)	33 (36)
OR (95% CI) & p-value for ORR	1.07 (95% CI: 0.72-1.61) p=0.7290		1.021 (0.556-1.87[a]) p=0.95	
CR, n (%)	[redacted]	[redacted]	0	1 (1)
Partial response, n (%)	[redacted]	[redacted]	32 (36)	32 (34)
Stable disease, n (%)	[redacted]	[redacted]	45 (51)	41 (44)
Progressive disease, n (%)	[redacted]	[redacted]	9 (10)	18 (19)
Median time to onset of response, months	8.1	5.6	NR	NR
Median DoR, months	20.0 (15.9 to 27.63)	13.2 (10.64 to 16.72)	NC	14.2[c]
Mean DoR (days)	752.14 (SE 0.138)	506.88 (SE 0.097)	NR	NR
Expected DoR (EDoR), months	11.4	7.5	NR	NR
EDoR ratio (95% CI), p-value	1.52 (95% CI 1.03 to 2.26) p=0.0367[c]		NR	
Mean DoCB (days)	853.48 (SE: 0.083)	717.64 (SE 0.068)	NR	NR

	FALCON		FIRST	
	Fulvestrant (n=230)	**Anastrozole (n=232)**		
EDoCB months	22.1 (18.46 to 24.87)	19.1 (16.53 to 20.47)	NR	NR
EDoCB ratio (95% CI), p-value	1.26 (0.99-1.59) p=0.0561[d]		NR	

This table draws on information presented in CS Table 19, p. 70 and Table 20 p. 76.

DoCB, duration of clinical benefit; DoR, duration of response; EDoCB, expected duration of clinical benefit; EDoR, expected duration of response; n, number of participants; NR, not reported; OR, odds ratio; ORR, objective response rate.

[a] the CS reports upper limit of the 95% CI as 1.687 but the paper published in 2009[6] reports 1.87.

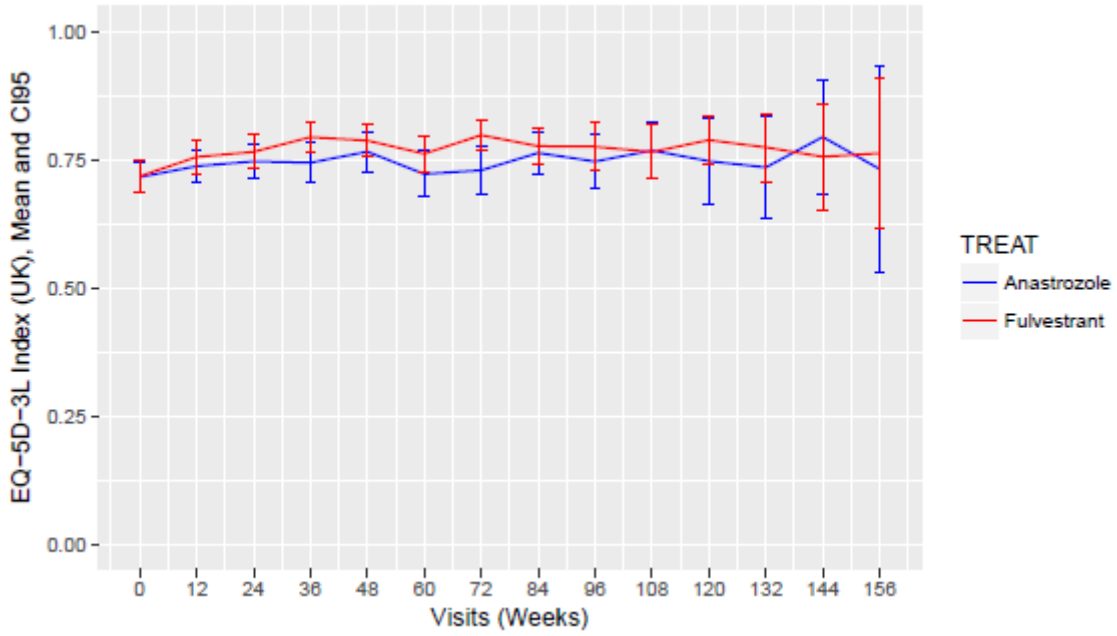
[b] these data were obtained from the FALCON CSR

[c] The CS reports 12 months but the paper published in 2009[6] reports 14.2 months

[d] The EDoR and EDoCB ratios presented in CS Table 20 (CS p. 76) match those presented in the response to clarification question A4, however the 95% CIs and p-values differ. The ERG has reported the values from the CS.

3.3.5 Summary of Health-related quality of life

As stated HRQoL was not measured in the FIRST trial.[6] In the FALCON trial[8] HRQoL was a secondary outcome. Two HRQoL questionnaires were utilised, the EQ-5D-3L and the FACT-B, and results from both are presented in the CS. Data collected using the EQ-5D-3L was used to inform HRQoL values, using the utility value set for the UK, in the economic model (see ERG report section 4.3.6). The CS reports that the results from the EQ-5D-3L questionnaire show that, over the 156 weeks of the study period, general health status was maintained across both treatment arms. These data are presented in a CS figure which is reproduced below (Figure 8). This figure did not indicate how many of the trial participants contributed data at each time point so NICE and the ERG requested clarification on this (clarification question A3). In response the company supplied a confidential reference which contains numerous tables and analyses. The ERG believes they have identified the correct patient numbers from this document and these have been added by the ERG to the figure.

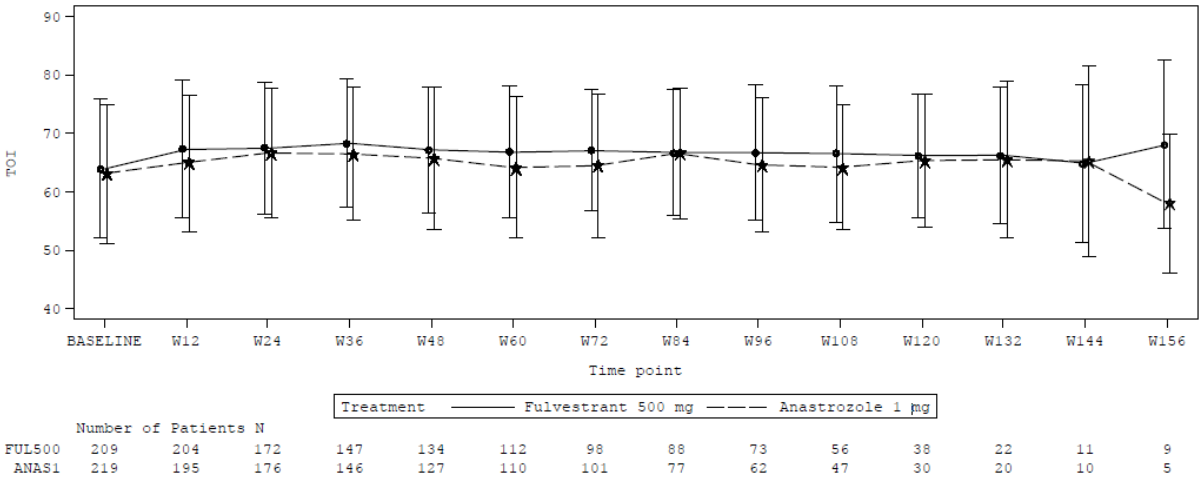


Number of observations per visit

Visits (weeks)	[redacted]													
Fulvestrant	[redacted]													
Anastrozole	[redacted]													

Figure 8 EQ-5D-3L Index (UK) per treatment and visit (CS Figure 17, p. 74)

The outcome measure from the FACT-B questionnaire was the TOI, which summarises three of the five subscales assessed by this questionnaire (physical well-being, functional well-being and breast cancer subscales). The CS reports (pp. 72 -73) that mean baseline TOI scores were high and comparable between the treatment arms and remained similar and high during treatment. The results are summarised in figure (CS Figure 16) which is reproduced below (Figure 9).



SOURCE: (AstraZeneca 2015a)
ANAS1, Anastrozole 1 mg; FUL500, fulvestrant 500 mg; TOI, Trial outcome index.

Figure 9 Mean TOI score across time points, by treatment group (CS Figure 16, p. 73)

3.3.6 Sub-group analyses results

As stated in Section 3.1.6 subgroup analyses (based on for the FIRST trial were not pre-planned. Results for these subgroups, which appear to be performed as defined by pre-defined covariates, are presented in forest plots (CS Figure 18 and Figure 19, pp. 77 to 78), which match the data in the published papers.[5,7] The CS states that the statistically significant difference in TTP reported for the FIRST trial population was maintained when adjusted for the pre-defined covariates (HR 0.64; 95% CI 0.46 to 0.90; p=0.01). The global interaction test was not statistically significant (p = 0.34). The treatment effect was consistent across the five subgroups (age, receptor status, visceral involvement, prior chemotherapy, measurable disease). The ERG notes that prior endocrine therapy was not included in the TTP subgroup analysis. Consistent results across six subgroups (age, receptor status, visceral involvement, prior chemotherapy, measurable disease, prior endocrine therapy) were also observed for OS. The FIRST trial's exploratory subgroup analysis suggesting that patients who had not received prior endocrine therapy received a greater OS benefit was the reason that the FALCON study focussed on endocrine therapy naive patients (although the CS p. 78 does caution that the endocrine therapy naive subgroup analysis was based on a very small sample size).

Subgroup analyses of PFS for the FALCON trial were pre-planned (albeit with some amendments prior to unblinding the data) and subgroup analysis for OS was added (prior to unblinding the data) as previously stated (Section3.1.6). PFS results for eight subgroups are presented in a forest plot (CS Figure 20, p. 79) and this matches the data in the published

paper.[8] For the subgroup analysis of OS, the geographic region subgroup is omitted (CS Figure 21, p. 80). The OS subgroup analysis is not presented in the published paper,[8] but the data match those reported in the FALCON CSR.[26] The subgroup analyses of PFS (breast cancer type, previous chemotherapy, geographic region, measurable disease, ER+ and PgR+, previous systemic ER containing HRT, bisphosphonate use, visceral disease) showed that the numerical improvement in PFS favouring fulvestrant was largely consistent across the subgroups. The largest numerical difference in the reported hazard ratios for PFS of subgroups was observed for the visceral disease versus no visceral disease at baseline. The analysis indicates that those with no visceral disease at baseline have a greater benefit than those with visceral disease at baseline. Consideration of subgroups of people with visceral disease and those with non-visceral disease (if evidence allows) was included in the company's decision problem. However, the company do not discuss this subgroup result in the CS. The published paper for the FALCON trial[8] points out that there is potential for an enhanced fulvestrant treatment effect in the non-visceral disease subgroup but indicates that the observation requires further study. Subgroup analysis was also conducted for the interim analysis of OS, in which the results were also consistent with the results for the whole study population. The CS states that a further analysis of OS will be conducted when approximately 50% of patients have died (CS p. 59). [redacted]

As already noted, for some subgroups the sample size is small and so caution is needed in interpreting these results.

3.3.7 Summary of adverse events

The CS presents an overview of the safety and tolerability of fulvestrant in CS section 4.12.1 (CS p. 119).

For the FALCON study, AEs are presented in the CS from the 11th April 2016 data cut off (the point of PFS analysis). At this point the median duration of exposure to fulvestrant was 14.7 months (range 0.9 to 37.7) and to anastrozole 13.9 months (range 0.2 to 36.0). The CS reports those events that occurred with a frequency of more than 5% in any treatment group (CS Table 39, p. 122).

Publications from the FIRST trial data provide safety data from three time points:

- At first data cut-off (10th January 2008)[6]
- At first follow-up (26th March 2010)[5]

- At the final assessment of OS (15th July 2014)[7]

The CS summarises data from the first two of these three time points and CS Table 38 (p. 121) summarises the data from the main study period and the follow-up period combined.

The combined data from the FIRST study and the data reported in the CS for the FALCON study have been used to populate Table 21.

The proportions of AEs and SAEs were similar between treatment groups. Due to differences in the length of follow-up and methods of recording AEs it is not possible to make comparisons between the two trials.

In the FALCON trial joint disorders and back pain were specified as AEs of special interest. These were reported by 26% of the fulvestrant group and 18% of the anastrozole group. In almost all cases the AEs of special interest were mild or moderate in severity (grade 1 or 2). The single exception was one patient (<1%) in the fulvestrant group who had grade 3 back pain.

Table 21 Summary of AEs reported in the CS for FALCON and FIRST

	FALCON		**FIRST**	
Parameter, n (%)	**Fulvestrant (n=228)**	**Anastrozole (n=232)**	**Fulvestrant (n=101)**	**Anastrozole (n=103)**
	At data cut-off (11/04/2016)		At final data cut-off (65% OS)	
Any AE	166 (73%)	173 (75%)		
Any SAE	30 (13%)	31 (13%)	24 (23.8%)	22 (21.4%)
Any SAE with outcome other than death			21 (20.8%)	18 (17.5%)
Any causally related SAE			2 (2.0)	0 (-)
Grade 3 or worse AEs	51 (22%)	41 (18%)		

Parameter, n (%)	AEs ≥5% in any treatment group[8]		Most commonly reported SAEs (≥2 patients)	
Alanine aminotransferase increased	16 (7%)	7 (3%)		
Anaemia	9 (4%)	20 (9%)		
Arthralgia	38 (17%)	24 (10%)		
Aspartate aminotransferase increased	12 (5%)	8 (3%)		

Atrial fibrillation			1 (1.0)	1 (1.0)
Back pain	21 (9%)	14 (6%)		
Cardiac failure			2 (2.0)	0 (-)
Constipation	13 (6%)	11 (5%)		
Cough	12 (5%)	8 (3%)		
Death			0 (-)	2 (1.9)
Decreased appetite			2 (2.0)	0 (-)
Dehydration			2 (2.0)	0 (-)
Diarrhoea	14 (6%)	13 (6%)		
Dyspnoea	9 (4%)	13 (6%)	2 (2.0)	0 (-)
Fatigue	26 (11%)	16 (7%)		
Femur fracture			1 (1.0)	2 (1.9)
Hot flush	26 (11%)	24 (10%)		
Hypertension	15 (7%)	21 (9%)		
Insomnia	15 (7%)	13 (6%)		
Myalgia	16 (7%)	8 (3%)		
Nausea	24 (11%)	24 (10%)		
Neuralgia			1 (1.0)	1 (1.0)
Oedema peripheral	9 (4%)	13 (6%)		
Pain in extremity	13 (6%)	10 (4%)		
Transient ischaemic attack			0 (-)	2 (1.9)

This table draws on information presented in CS Table 38, p. 121 and Table 39 p. 122.

AE, adverse event; n, number; OS, overall survival; SAE, serious adverse event

Discontinuations

In the FALCON study, 7% of the fulvestrant arm and 5% of the anastrozole arm discontinued because of AEs. The CS presents discontinuations by organ class in CS Table 40 (p. 123). For the FIRST study, information on discontinuation due to an AE is reported from the first data cut-off in the CS, with additional information being presented in the published paper from the first follow-up. In both studies the proportion of patients discontinuing due to an AE were similar in the fulvestrant and anastrozole groups (Table 22).

Table 22 Summary of study discontinuations due to an AE

	FALCON		FIRST		
Parameters, n (%)	**Fulvestrant (n=228)**	**Anastrozole (n=232)**	**Time-point**	**Fulvestrant (n=101)**	**Anastrozole (n=103)**
Discontinuation due to an AE	16 (7%)	11 (5%)	First data cut-off, 10/01/2008[6], CS p. 119	3 (3.0%)	3 (2.9%)
			First follow-up (26th March 2010)[5]	0	2 (1.9%)

AE, adverse event; CS, company submission; n, number

Deaths related to adverse events

In the FALCON trial 3% of deaths were considered to be related to AEs (6 in the fulvestrant group and 7 in the anastrozole group) at the 11th April 2016 data cut off (the point of PFS analysis). None of these deaths were considered to be causally related to study treatment. A similar proportion of deaths from the main study period and the follow-up period combined in the FIRST study were due to AEs [3 (3%) SAEs in the fulvestrant group and 5 (4.9) SAEs in the anastrozole group] (Table 23).

Table 23 Summary of deaths related to AEs

Parameters, n (%)	FALCON		FIRST	
	Fulvestrant (n=228)	**Anastrozole (n=232)**	**Fulvestrant (n=101)**	**Anastrozole (n=103)**
Deaths related to AEs	6 (3%)	7 (3%)	3 (3.0%)	5 (4.9%)

AE, adverse event; n, number

3.4 Summary

The systematic review of clinical effectiveness evidence in the CS identified two RCTs of fulvestrant as a treatment for people with untreated hormone-receptor positive locally advanced or metastatic breast cancer (FIRST and FALCON). Both trials compared fulvestrant to anastrozole.

The two RCTs were judged to be of good methodological quality although there was the potential for the FIRST trial to be at a high risk of bias due to the absence of blinding. Overall, both studies appear to have been well conducted. The main clinical efficacy outcomes reported in the CS are PFS, OS, CBR (response rates) and HRQoL. AE outcomes are also reported. Follow-up of participants from the FALCON study is continuing, particularly with regard to OS for which there are currently only interim results.

The company's SLR had broad inclusion criteria, enabling the identification of studies that could contribute to the wider evidence base where necessary. As there is no direct evidence comparing fulvestrant to either letrozole or tamoxifen it was necessary for the company to conduct an NMA. In addition to the two trials of fulvestrant versus anastrozole, the NMA included data from a further four trials: combined data from the North American and TARGET studies (these two trials were prospectively designed to allow for combined data

analysis) which compared anastrozole to tamoxifen, and the PO25 trial which compared letrozole to tamoxifen. A sixth trial (comparing anastrozole with 40 mg tamoxifen) was not included in the final network because it used the higher 40 mg tamoxifen dose which is not approved by the EMA and its inclusion caused heterogeneity in the NMA.

The additional RCTs contributing data to the NMA were judged to be of good methodological quality where judgements about the risk of bias could be made. However, the ERG found that for many items the risk of bias judgement was 'unclear' because the published papers did not report the necessary details.

The ERG found some evidence of heterogeneity in trial participants between the five trials that contributed to the final NMA. However, the company conducted a matching process, so that for the trials where IPD were available (FIRST and NorthAmTarget) only data for ER+/PgR+ patients plus endocrine treatment naive patients would be included in the NMA. Inevitably the matching process decreased the sample size of the FIRST and NorthAmTarget studies. The ERG is aware that a benefit of the matching process is that it allows for the exclusion of participants [redacted] creating a more homogeneous population for the NMA (except for study PO25 for which IPD were not available). Although, the ERG is uncertain about the potential disadvantages of this approach in terms of the effects on the original randomised trial arms (e.g. if it creates scope for bias as randomisation has been broken) the ERG has concluded that it is likely that the benefits outweigh potential disadvantages.

Two outcomes were analysed by NMA, PFS and OS. The PFS and OS data from the individual trials (after matching where applicable) were examined to determine whether the assumption of proportional hazards held. Visual inspection of both the KM plots and of log cumulative hazard plots suggested that a constant relative treatment effect in the studies was unlikely. Therefore the company concluded that methods of NMA reliant on the assumption of proportional hazards were inappropriate and instead used an alternative method developed by Ouwens et al.[23] In this method PFS and OS data can be both synthesized in the NMA and extrapolated beyond the available trial follow-up.

A fixed-effect NMA was undertaken because of the small number of studies and the difficulty in determining an appropriate informative prior for a random-effects analysis. The ERG accepts that the small number of trials available in the network is a limitation and is

concerned that the inability to conduct random-effects analyses means that the potential for uncertainty may be inadequately captured.

PFS was the primary outcome of the FALCON trial and TTP was a secondary outcome of the FIRST study (the FIRST study definition of TTP included deaths and hence is treated the same as PFS). In both studies a benefit was observed for the fulvestrant group: median PFS 2.8 months longer for the fulvestrant group in FALCON; median TTP 10.3 months longer for the fulvestrant group in FIRST. In both cases these improvements were statistically significant (FALCON HR = 0.797, 95% CI 0.637 to 0.999, p = 0.0486; FIRST HR = 0.66, 95% CI 0.47 to 0.92, p = 0.01). The fixed-effect NMA was conducted for five different parametric distributions. For four of these (Weibull, log-logistic, lognormal and generalised gamma) the difference in the scale parameter indicated that fulvestrant PFS is better than anastrozole and was statistically significant, whereas with tamoxifen PFS was statistically significantly worse than anastrozole. For the shape parameter a statistically significant difference was apparent only for the lognormal distribution indicating fulvestrant was better than anastrozole whereas tamoxifen was worse than anastrozole.

OS was a secondary outcome of both the FALCON and FIRST RCTs, in the case of the FIRST study the outcome was added in a protocol amendment after improvements in TTP had been observed. The OS data for FALCON are immature and median survival has not yet been reached. The slight difference in the proportion of deaths in favour of the fulvestrant arm (29% vs 32%) is not statistically significant. In the FIRST trial median survival in the fulvestrant arm was almost 6 months longer than that of the anastrozole arm (54.1 months versus 48.4 months) and this improvement with fulvestrant was statistically significant (HR 0.70, 95% CI 0.50 to 0.98, p=0.04). A fixed-effect NMA was conducted for five different parametric distributions. Although some statistically significant differences were observed for the comparison of letrozole versus anastrozole there were not statistically significant differences with any of the parametric distributions for the fulvestrant versus anastrozole comparison.

CBR was the primary outcome of the FIRST trial (powered for non-inferiority) and a secondary outcome of the FALCON trial. In both trials the CBR favoured fulvestrant (FIRST: fulvestrant 72.5% versus anastrozole 67%; FALCON: fulvestrant 78% versus anastrozole 74%).

Secondary outcomes reported by both trials were based on response to treatment. ORR (the best overall response of CR or PR) was broadly the same in the trial arms of both trials.

Other secondary outcomes based on response reported for the FALCON trial only (e.g. median DoR, EDoR, mean DoCB and EDoCB) were numerically in favour of fulvestrant and in the case of EDoR ratio, statistically significantly in favour of fulvestrant.

HRQoL was reported only from the FALCON trial using both the EQ-5D-3L and FACT-B questionnaires. Results from both questionnaires showed HRQoL was similar in the trial arms at the start of treatment and was maintained during treatment. Data from the EQ-5D-3L informed the economic model.

Subgroup analyses for both trials indicate that the TTP/PFS and OS results were consistent across the subgroups tested [FIRST: age, receptor status, visceral involvement, prior chemotherapy, measurable disease, prior endocrine therapy; FALCON: breast cancer type, previous chemotherapy, geographic region (not for interim OS subgroup analysis), measurable disease, receptor status, prior systemic ER containing HRT, bisphosphonate use, visceral disease]. Although the company decision problem includes provision for consideration of people with visceral disease and non-visceral disease the CS does not make any specific comments about this subgroup analysis.

The FIRST and FALCON trials reported similar proportions of AEs and SAEs between the study arms. Joint disorders and back pain were specified as AEs of special interest in the FALCON trial (reported by 26% of the fulvestrant group and 18% of the anastrozole group). Apart from one patient (<1%) in the fulvestrant group who had grade 3 back pain the AEs of special interest were mild or moderate in severity (grade 1 or 2). Discontinuations due to AEs were similar in the fulvestrant and anastrozole groups of the two trials. Some deaths due to adverse events were recorded but none were reported as being causally related to study treatment.

4 COST EFFECTIVENESS

4.1 Overview of company's economic evaluation

The company's submission to NICE includes:

i) a targeted review of published submissions made to national reimbursement and health technology assessment organisations of therapies for the treatment of locally advanced or metastatic breast cancer.

ii) a report of an economic evaluation undertaken for the NICE STA process. The cost effectiveness of fulvestrant is compared with anastrozole, letrozole and tamoxifen for post-menopausal women with HR+ locally advanced or metastatic breast cancer who have not previously been treated with any hormonal therapy.

4.2 Company's review of published economic evaluations

A targeted review was conducted by the company to identify HTAs of therapies for locally advanced or metastatic breast cancer. The review included a search of national reimbursement and technology assessment organisations. The websites were searched in May 2016 for any HTA in breast cancer and those related to advanced or metastatic breast cancer were included. The company did not include any additional inclusion or exclusion criteria. More details on the search strategies are provided in section 3.1.1.

The HTAs identified are shown in CS Appendix E. The company considered those published by NICE to be most relevant and therefore included these assessments in their review. They identified 10 NICE technology appraisals, relating to metastatic breast cancer. These are summarised in CS Table 43 and the methods and results of each submission are shown in CS Table 44. The company does not provide any discussion about the assessments identified, for example concerning their relevance to the current submission.

The ERG notes that five submissions relate to first-line therapy but none of the submissions relate to the same population as in the current submission. The ERG notes that the company has not searched for published cost-effectiveness literature. The ERG has therefore completed a search of published cost-effectiveness studies.

The ERG searched EMBASE and Pubmed database from 2010 (date of search in previous NICE appraisal for fulvestrant) for economic evaluations of anastrozole, letrozole, tamoxifen or fulvestrant in post-menopausal women with ER+ advanced or metastatic breast cancer. We excluded studies reported as abstracts or not in English. We identified two studies.[27,28]

Newman et al.[27] compared fulvestrant 250 mg with fulvestrant 500 mg for patients previously treated with an AI or antiestrogen therapy. Das et al.[28] compared fulvestrant 500 mg with nonsteroidal AIs (anastrozole and letrozole) in patients who had previously received hormonal therapy in the United Kingdom. The ERG notes that these studies are for a relevant population but are not for first-line treatment so may be of limited relevance to this appraisal.

4.3 Critical appraisal of the company's submitted economic evaluation

4.3.1 NICE reference case

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

Table 24 NICE reference case requirements

NICE reference case requirements:	Included in submission	Comment
Decision problem: As per the scope developed by NICE	Yes	CS Table 1
Comparator: As listed in the scope developed by NICE	Yes	CS Table 1
Perspective on costs: NHS and PSS	Yes	CS Table 46
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	CS Table 46
Perspective on outcomes: All direct health effects, whether for patients or, when relevant, carers	Yes	
Type of economic evaluation: Cost utility analysis with fully incremental analysis	Yes	
Synthesis of evidence on outcomes: Based on a systematic review	Yes	
Time horizon: Long enough to reflect all important differences in costs or outcomes between the technologies being compared	Yes	CS Table 46
Measuring and valuing health effects: Health effect should be expressed in QALYs. The EQ-5D is the preferred measure of health-related quality of life.	Yes	EQ-5D used for disease health states from the company's clinical trial
Source of data for measurement of health-related quality of life: Reported directly by patients and/or carers.	Yes	
Source of preference data: 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	
Discount rate: 3.5% pa for costs and health effects	Yes	CS Table 46

As shown in Table 24, the methods and data inputs used in the company's economic evaluation conform to NICE methodological guidance.

4.3.2 Model Structure

The company presented a cohort-based partitioned survival model with three mutually exclusive health states: PFS, PD and death. A schematic of the model was presented in CS Figure 39 which showed the proportion of patients in the three mutually exclusive health states over the model time-horizon. However, the company's model schema did not reflect the direction of the patients' flow across the three mutually exclusive health states. To address this, the ERG produced a diagram of the three-state model structure (shown in Figure 10) to illustrate the patient flow in a transparent and intuitive manner. A lifetime horizon of 30 years was applied in the base case model. The CS justified the time-frame by stating that at the end of this time horizon <1% of the population were alive. A cycle length of four weeks was used in the model, which, as stated in the CS, is the shortest time-period to observe any change in the disease symptoms and was consistent with the follow-up visit schedule in the FALCON trial. A half-cycle correction was applied correctly and costs and health effects were discounted at 3.5% per annum as outlined in the NICE reference case.[18] The perspective adopted was that of the NHS and PSS. The model was constructed in Microsoft Excel 2010.

The CS stated that this model structure was chosen as the health states are in line with the clinical pathway and the model structure is consistent with the approaches used in earlier NICE appraisals for ABC as well as other cancers. The model accurately represents the clinical pathway of patients' transition through the course of their treatment for advanced/metastatic breast cancer by assuming that patients with disease progression cannot transition back to progression-free health state.

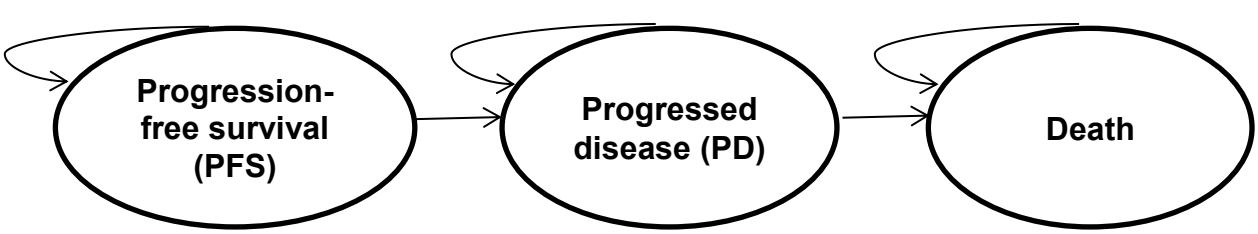


Figure 10 Model structure (illustration adapted by the ERG)

To inform the clinical parameters of PFS and OS within the economic model, the CS used the results from the NMA, as discussed earlier in section 3.3. Long term data for these parameters were extrapolated by fitting parametric survival curves (see more details in section 4.3.5). The model derived the proportion of patients in the PD state as the difference between the PFS and OS curves. Patients received treatment until disease progression. All patients were assumed to receive subsequent treatments and these subsequent treatments were only assumed to impact costs. AEs were included as a one-off event in the first treatment cycle within the company's analyses to account for the AEs associated costs and quality of life whilst on treatment. The model included costs associated with disease management, treatment acquisition, treatment administration, subsequent therapy and AEs and incorporated quality-adjusted life years (QALYs) by assigning utility values to the health states (further details are discussed in sections 4.3.6 and 4.3.7).

Overall, the ERG considers the model approach to be appropriate and consistent with the clinical pathway of patients with advanced or metastatic cancer. The CS presents sufficient justification for the company's methodological and structural choices in CS Section 5.2.

4.3.3 Population

The economic evaluation includes the population defined in the company's decision problem as postmenopausal people with locally advanced or metastatic HR+ breast cancer who have not received endocrine therapy. This corresponds with the final scope issued by NICE[1] [redacted] The patient population is also consistent with the patient population included in the FALCON trial.

4.3.4 Interventions and comparators

The cost-effectiveness model compares fulvestrant to anastrozole, letrozole and tamoxifen, as specified by the NICE scope[1] and the company's decision problem. Fulvestrant currently has a marketing authorisation for the treatment of patients on or after adjuvant anti-oestrogen therapy, or disease progression on therapy with an anti-oestrogen therapy. [redacted] The recommended dose is 500 mg administered IM into the buttocks as two injections, twice in the first month and then monthly thereafter. This dosage was used in the FIRST and FALCON trials. The ERG considers that the intervention in the decision problem reflects the anticipated use in UK clinical practice.

The comparators listed in the NICE scope[1] and the company's decision problem are aromatose inhibitors (anastrozole and letrozole) and tamoxifen. The use of tamoxifen is restricted to the instance where AIs are not tolerated or are contra-indicated. Anastrozole, letrozole and tamoxifen are available as oral medication taken daily by patients. The ERG's clinical advisors agreed that the comparators in the NICE scope were appropriate and were routinely used in the UK NHS.

4.3.5 Treatment effectiveness and extrapolation

As described in section 4.3.2, the economic model comprises three health states. In the company's base case analysis, patients were modelled to move through these mutually exclusive states over four-weekly cycles for a life time horizon. Patients were modelled to discontinue the first-line therapy and move to the PD state when the disease progressed.

The company identified six relevant studies of first-line pharmacological therapies for post-menopausal women with HR+, locally advanced or metastatic breast cancer through their SLR. The SLR has been summarised and critiqued in section 3.1 of the ERG report. The studies are: the FALCON[8] and FIRST[5] trials of fulvestrant versus anastrozole, the North American[9] and TARGET[10] trials of anastrozole versus tamoxifen (20 mg), the Milla-Santos et al.[11] trial of anastrozole versus tamoxifen (40 mg) and the PO25 trial[12] of letrozole versus tamoxifen (20 mg). These studies enabled indirect treatment comparison of fulvestrant and the other first line interventions tamoxifen and the AIs, anastrozole, and letrozole.

The company modelled clinical outcomes using a fixed-effect NMA. The key outcomes, PFS and OS were estimated from extrapolated survival curves. Incidence rates of AEs specific to each treatment group were used to estimate associated costs and disutilities for the corresponding cohort in the company's model. In this section, we summarise and discuss the methods used by the company to estimate the effectiveness outcomes of PFS, OS and time to treatment discontinuation (TTD) as well as AE rates for fulvestrant and the above-mentioned comparators.

4.3.5.1 Survival outcomes

As stated above the company identified six studies from its SLR for analysis in the model. Given the absence of head-to-head trials between fulvestrant and the comparators tamoxifen and letrozole, the company analysed these studies using indirect treatment comparison. The Milla-Santos[11] trial was ultimately dropped from the company's meta-analysis because it included comparison of a higher dose (40 mg) of tamoxifen than the

other pooled studies and its inclusion led to heterogeneity in the NMA of both the OS and PFS outcomes. The company combined individual patient level data from the North American and TARGET trials, and these two trials are jointly referred to as NorthAmTarget. A detailed description and critique of the company's approach to evidence synthesis can be found in section 3.1.7 of the ERG report.

The CS reports an attempt to match patients in each treatment arm of all relevant studies to the FALCON trial population. This was not possible for PO25 as patient-level data had to be reconstructed from a published article on the trial. The methodology of the matching process is discussed in detail in section 3.1.7 of this report. The KM plots of PFS and OS for the four studies (FALCON, FIRST, NorthAmTarget, and PO25) included in the economic model are reported in the CS (CS Figures 23 and 24, respectively) and reproduced below (Figure 11 and Figure 12). Plots for the FALCON, FIRST, and NorthAmTarget trials are for the matched data and not the full data set. In the absence of individual patient level data for PO25, the CS reports that KM data were digitised to permit the estimation of survival functions.

Figure 11 PFS KM plots from FALCON and studies identified in the SLR (CS Figure 23, p. 92)

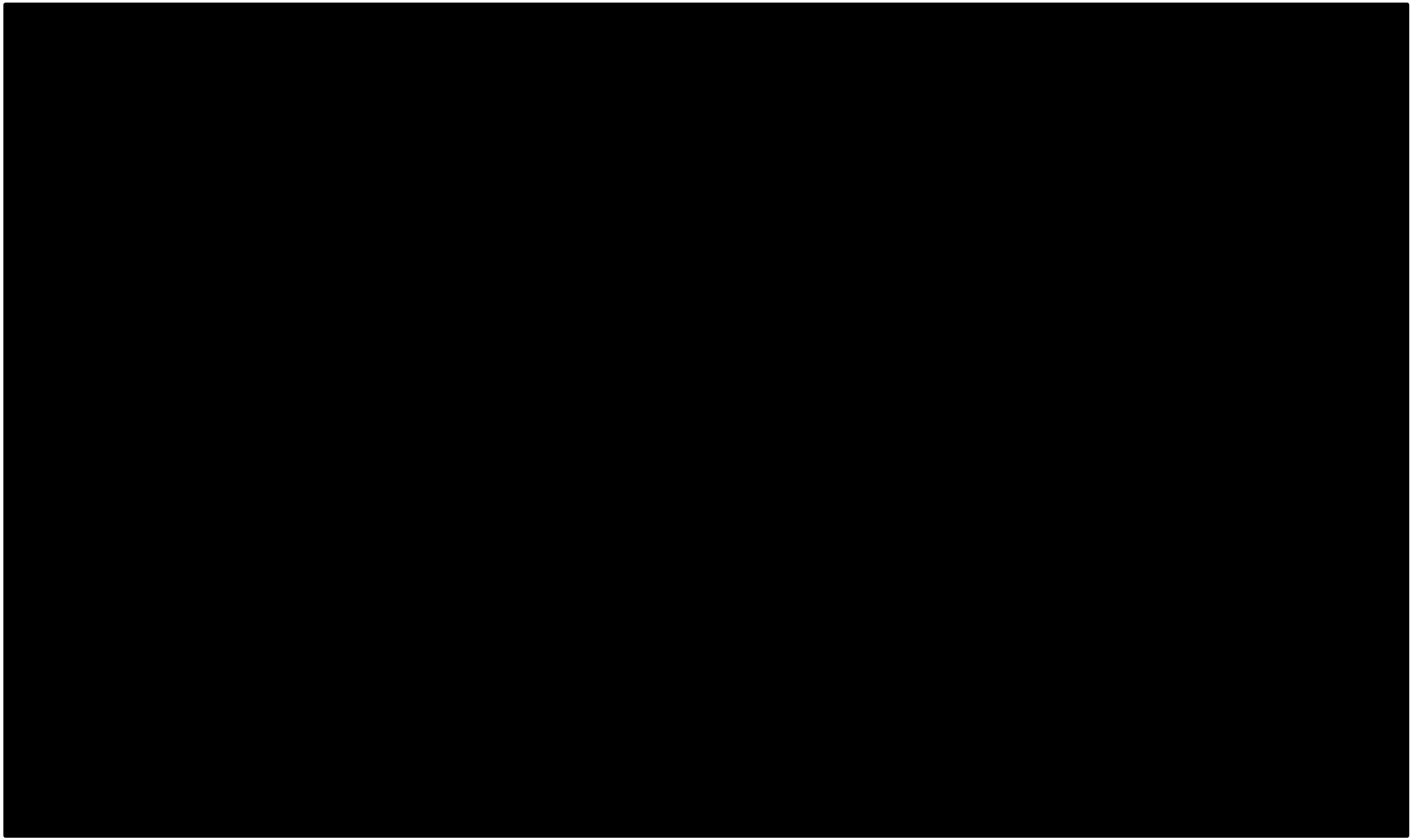
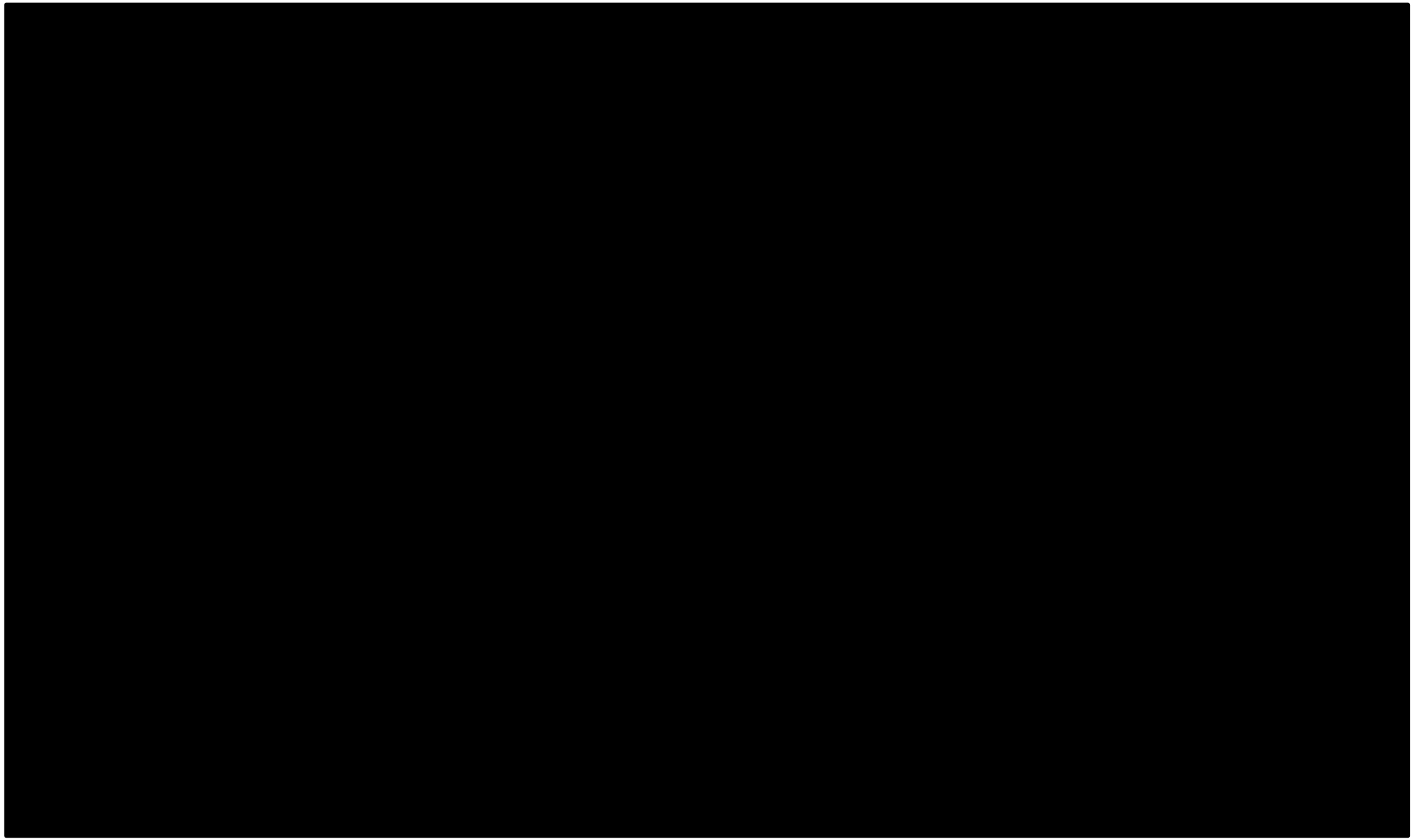


Figure 12 OS KM plots from FALCON and studies identified in the SLR (CS Figure 24, p. 93)



The company states that visual inspection of KM plots for PFS showed that treatment arms remained separated over the trial period. KM plots for OS depict late separation (21 months) for the FIRST trial and crossing plots for the PO25 trial and the NorthAmTarget trial. We agree with the company that, based on visual inspection, some of the treatment arms particularly for the KM plots of OS (NorthAmTarget and PO25) cross or separate beyond the median survival time. This suggests that NMA methods, which rely on the assumption of proportional hazards, may not be suitable for analysing the studies.

The CS further estimates the log cumulative hazard plots for PFS and OS for the four trials, to further investigate the violation of proportional hazards. These hazard functions are presented in CS Figures 25 and 26. Like the KM plots, visual inspection seems to suggest that the assumption of proportional hazards is violated: it can be observed that for OS, the treatment arms of the PO25 and NorthAm Target trial crossed. The log cumulative hazard arms in the FALCON, FIRST and NorthAmTarget trials cross for PFS, while for OS, arms cross for the NorthAmTarget trials. The CS further argues that using HRs as outcomes for the analysis places a restriction on the choice of distributions (such as log-normal and log-logistic distributions) that can be used to extrapolate PFS and OS. The company therefore sought alternative methods suitable for assessing NMA to extrapolate the treatment effect. The CS implements a method developed by Ouwens et al.[23] The Ouwens et al. method is premised on the fact that survival distributions, such as the Weibull or Gompertz, commonly used to extrapolate outcomes for cost-effectiveness analysis, can be described by two parameters (shape and scale). Further, applying a constant HR implies that treatment only affects the scale parameter. The Ouwens et al. method can be applied to both IPD and data derived from published KM curves such as the PO25 trial. A previous NICE appraisal for fulvestrant for previously treated patients with ABC reports the use of the Ouwens et al. method.[21] The ERG finds this method appropriate for implementing NMA, given the violation of the proportional hazards assumption.

The CS argues in favour of a fixed-effect NMA for PFS and OS rather than the random-effects model. The company's preference for a fixed-effect NMA has been critiqued in section 3.1.7 of this report. The CS describes two types of fixed-effect analyses. The first scenario is the 'All-shapes' model which permits the modelling of parametric survival distributions with the estimation of their shape and scale parameters, since it does not rely on the assumption of

proportional hazards. It forms the basis of the base case survival curves used in the cost-effectiveness model and tabulated results from the CS are reported in section 3.3.1 of the ERG report. The second scenario is the 'No shape arm' model, which assumes proportional treatment effects between treatment arms. The ERG believes the choice of the 'all shapes' model for the base case analysis is reasonable. The ERG queried the inclusion of the PO25 trial in the analysis (see section 3.1.3 of this report) and the company has provided cost-effectiveness results excluding this trial in its clarification response (Question A13, Table 25). The ERG has conducted a scenario analysis that excludes the PO25 trial data and the results are shown in this report section 4.4.

PFS extrapolation

The company extrapolated KM curves for all the selected parametric distributions (Weibull, Gompertz, log-logistic, lognormal and generalised gamma). The ERG verified that the extrapolated curves reported in the CS corresponded to those used in the economic model. Extrapolated curves for all distributions were simultaneously plotted along with observed data from each of the meta-analysed studies. See Figure 13 to Figure 16 below (CS Figures 29-32).

Figure 13 FALCON PFS study fit with fixed effects 'all shapes' network meta-analysis model (CS Figure 29, p. 110)

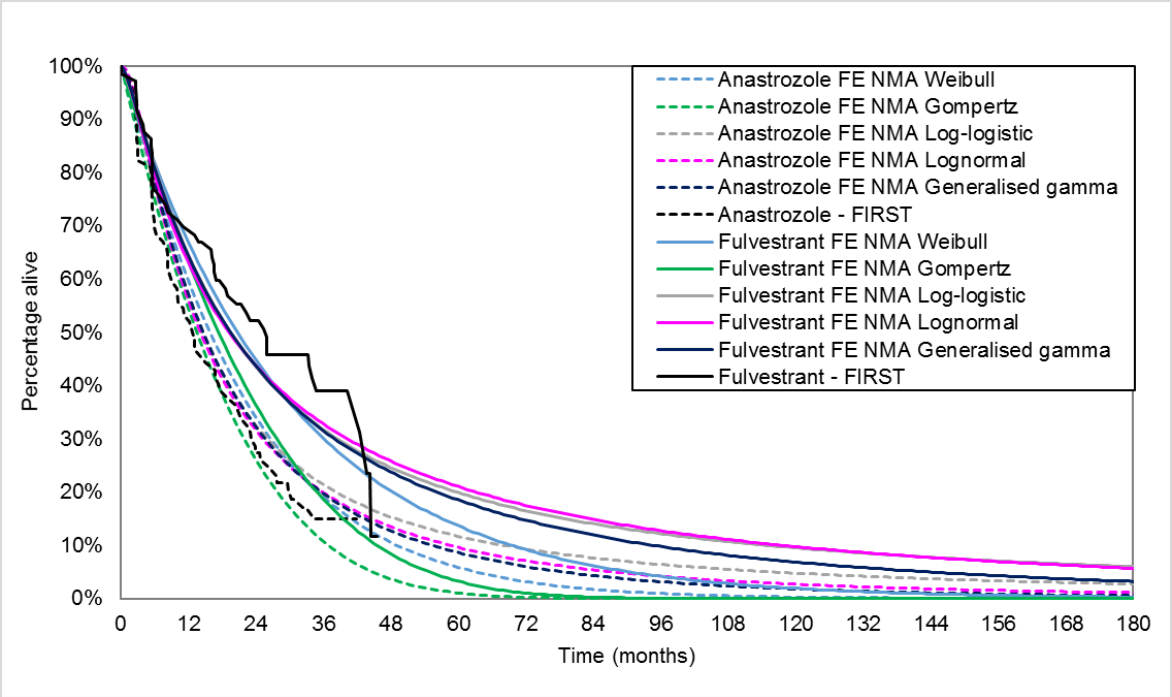


Figure 14 FIRST PFS study fit with fixed effects 'all shapes' network meta-analysis model adjusted for between-study differences (CS Figure 30, p. 110)

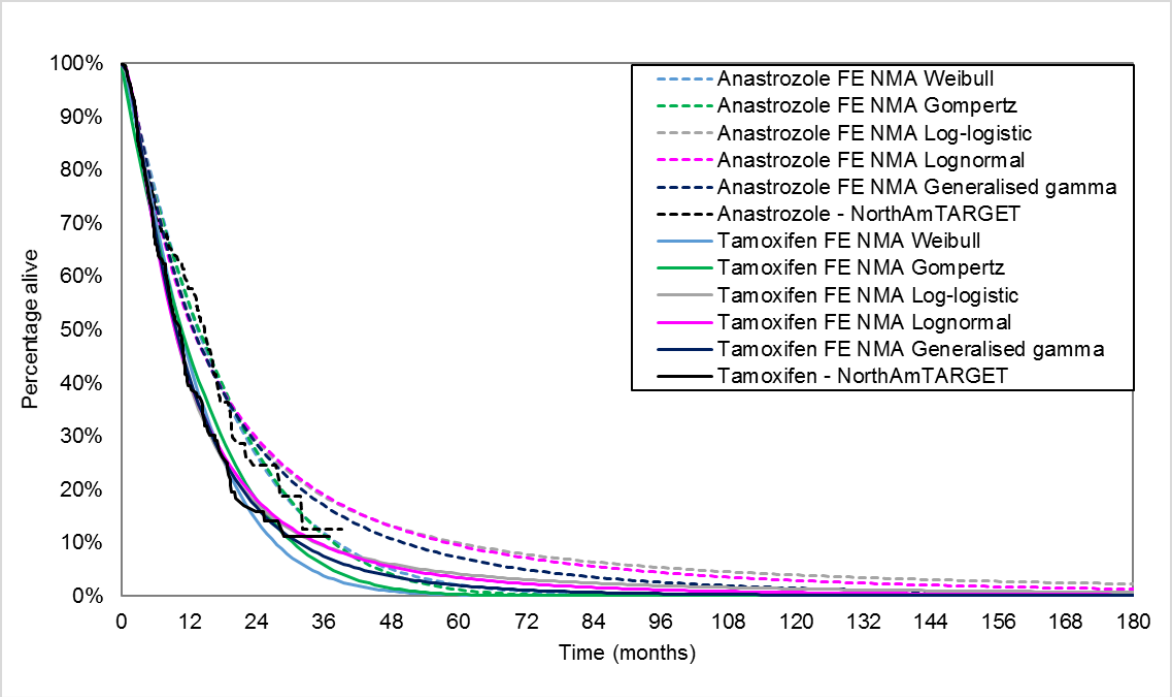


Figure 15 NorthAmTarget PFS study fit with fixed effects 'all shapes' network meta-analysis model adjusted for between-study differences (CS Figure 31, p. 111)

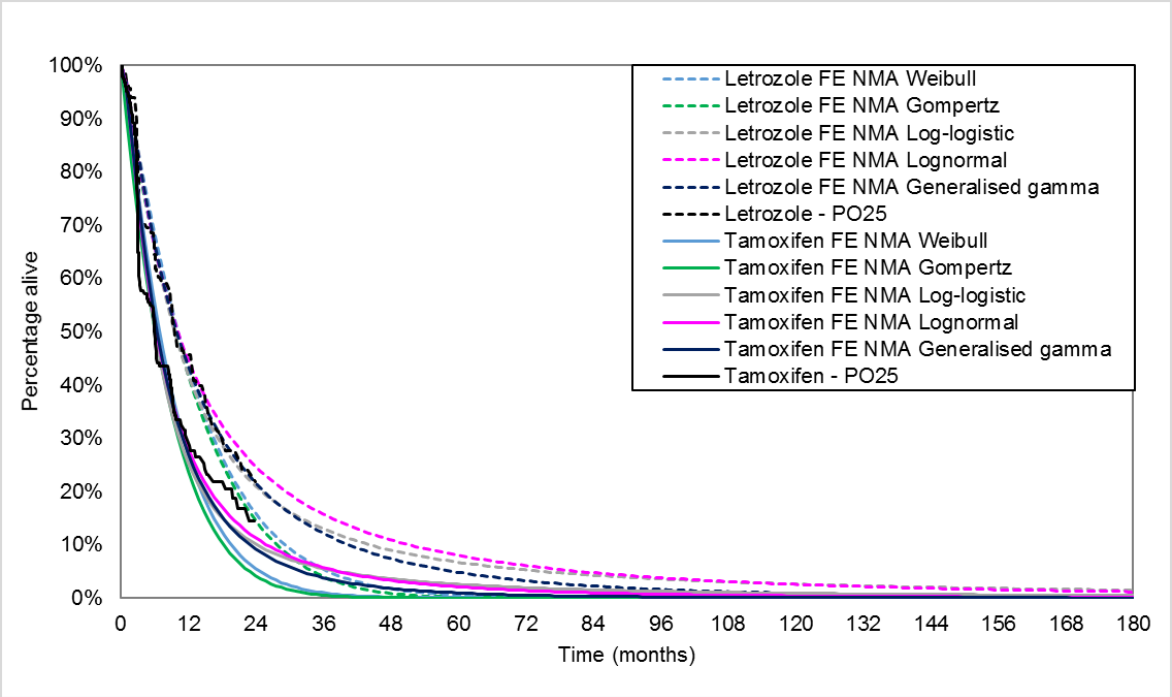


Figure 16 PO25 PFS study fit with fixed effects 'all shapes' network meta-analysis model adjusted for between-study differences (CS Figure 32, p. 111)

The ERG notes that the PFS curves in figures 12 to 15 are different for the same intervention. The ERG understands that the company has fitted the four curves to the observed data from the trials separately to give outputs for their Akaike and Bayesian Information Criteria statistics, but this is not stated explicitly in the CS.

The CS reports Akaike and Bayesian Information Criteria statistics for PFS (Table 25). The Bayesian information criterion (BIC) and the Akaike information criterion (AIC) are closely related statistics commonly used for model or distribution selection. The distribution with the lowest AIC or BIC value represents the best fit to the observed survival data. One limitation of the AIC and BIC is that they cannot be extended to make predictions of fitness beyond the observed data. The CS reports that visual inspection and expert opinion have been used to assess the different extrapolations of the survival data. Based on the company's clinical experts, 1-5% of patients treated with anastrozole are estimated to still be progression-free after 10 years (see Table 26). The company chose the generalised gamma distribution as the most appropriate fit, based on visual inspection and the opinion of the company's clinical experts[29], although AIC and BIC (Table 25) placed the distribution at second best after the log-logistic distribution. Other distributions (log-logistic, lognormal, Weibull and Gompertz) were tested in

sensitivity analyses (Table 99 of the CS) which shows that the choice of distribution did not impact significantly on the ICER.

We note that the results from the company's AIC / BIC statistics seem inconclusive as certain distributions perform better for one trial and worse for others. For instance, the Weibull distribution which was one of the least favourable according to AIC and BIC criteria had a better fit to the tail of the FALCON KM plot for fulvestrant than the log-logistic, generalised gamma and log-normal. Advice from our clinical expert confirmed the PFS estimates in Table 26. We note that the fit in Figure 13 against the FALCON study is reasonable. The ERG, therefore, considered the company's choice of the generalised gamma distribution was reasonable for modelling PFS.

Table 25 AIC and BIC statistics for PFS based on fixed effects NMA model (CS Table 31, p. 109)

Distribution	AIC	AIC rank	BIC	BIC rank
Log-logistic	8624.747	1	8703.403	1
Generalised gamma	8627.055	2	8711.329	2
Lognormal	8636.065	3	8714.721	3
Weibull	8687.484	4	8766.140	4
Gompertz	8720.786	5	8799.441	5

Abbreviations: AIC, Akaike's Information Criterion; BIC, Bayesian Information Criterion; NMA, network meta-analysis; PFS, progression-free survival.

Table 26 KOL opinion on PFS at 1, 2, 5 and 10 years (CS Table 32, p. 112)

	1 year	2 years	5 years	10 years
KOL estimate	50-60%	30-40%	5-10%	1-5%

Abbreviation: KOL, key opinion leader.

OS extrapolation

KM curves were also extrapolated for OS for the same parametric distributions as for PFS. Extrapolated curves for all the distributions were simultaneously plotted along with the observed data (using the matched subgroup as stated earlier) from each of the meta-analysed studies and are shown in Figure 17 to Figure 20 (CS Figures 34 -37). Similarly, we ascertained that extrapolated curves reported in the CS corresponded to those used in the economic model.

Figure 17 FALCON OS study fit with fixed effects 'all shapes' network meta-analysis model (CS Figure 34, p. 115)

[redacted]

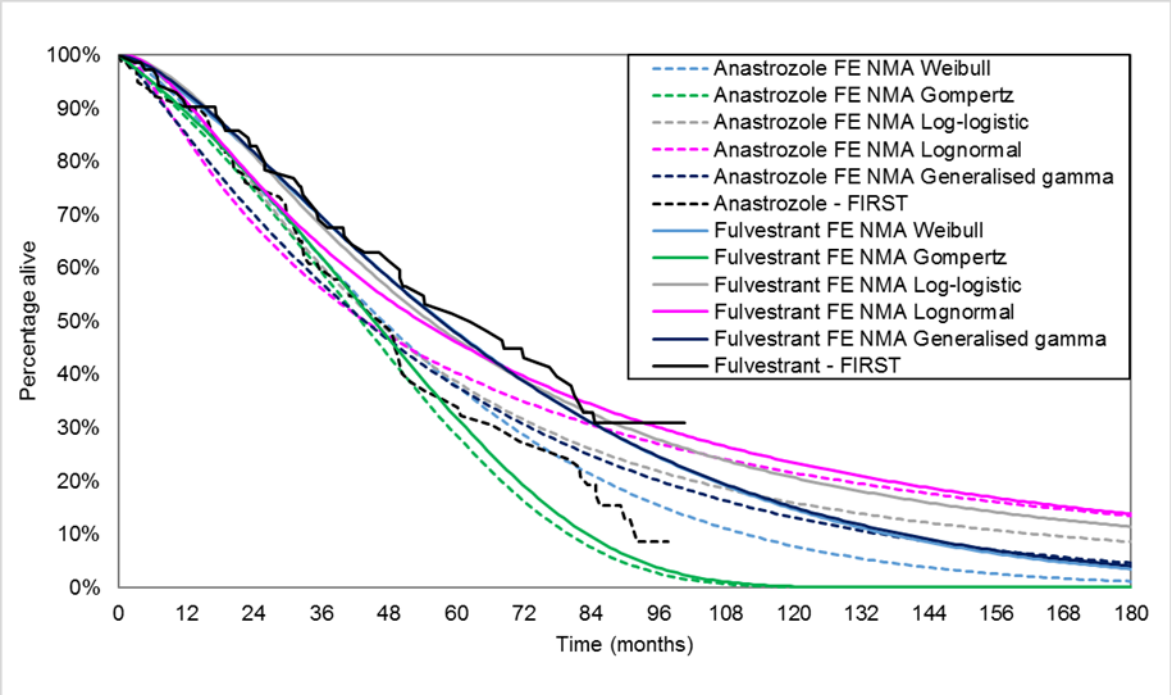


Figure 18 FIRST OS study fit with fixed effects 'all shapes' network meta-analysis model (CS Figure 35, p. 115)

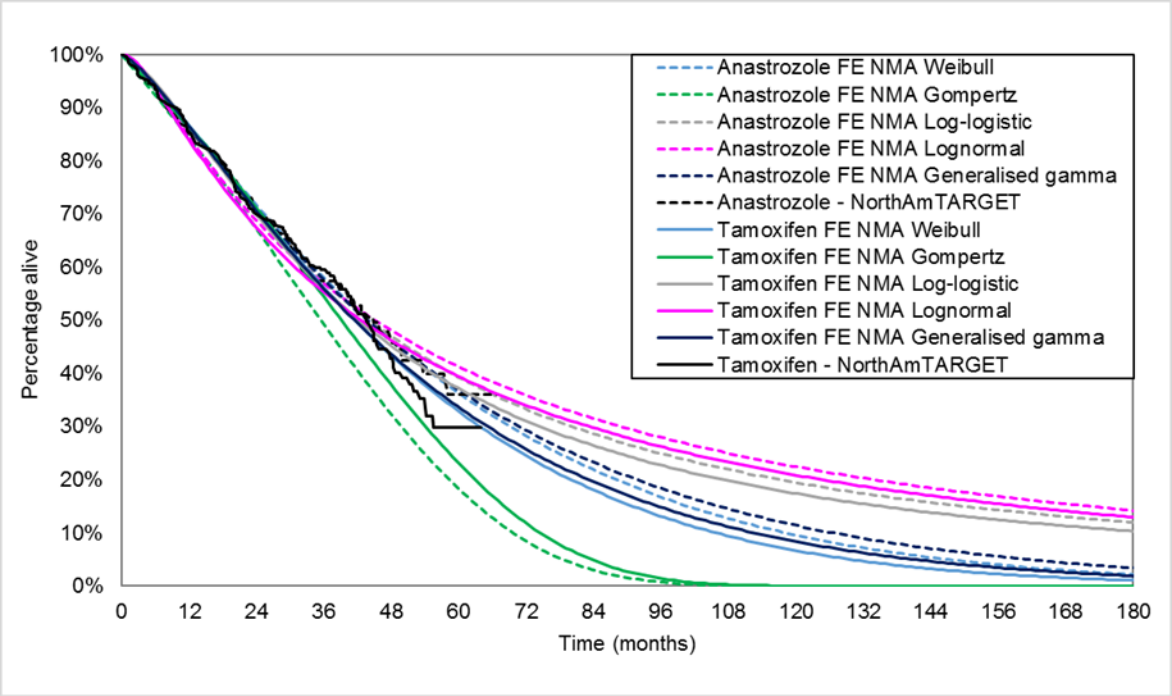


Figure 19 NorthAmTarget OS study fit with fixed effects 'all shapes' network meta-analysis model (CS Figure 36, p. 116)

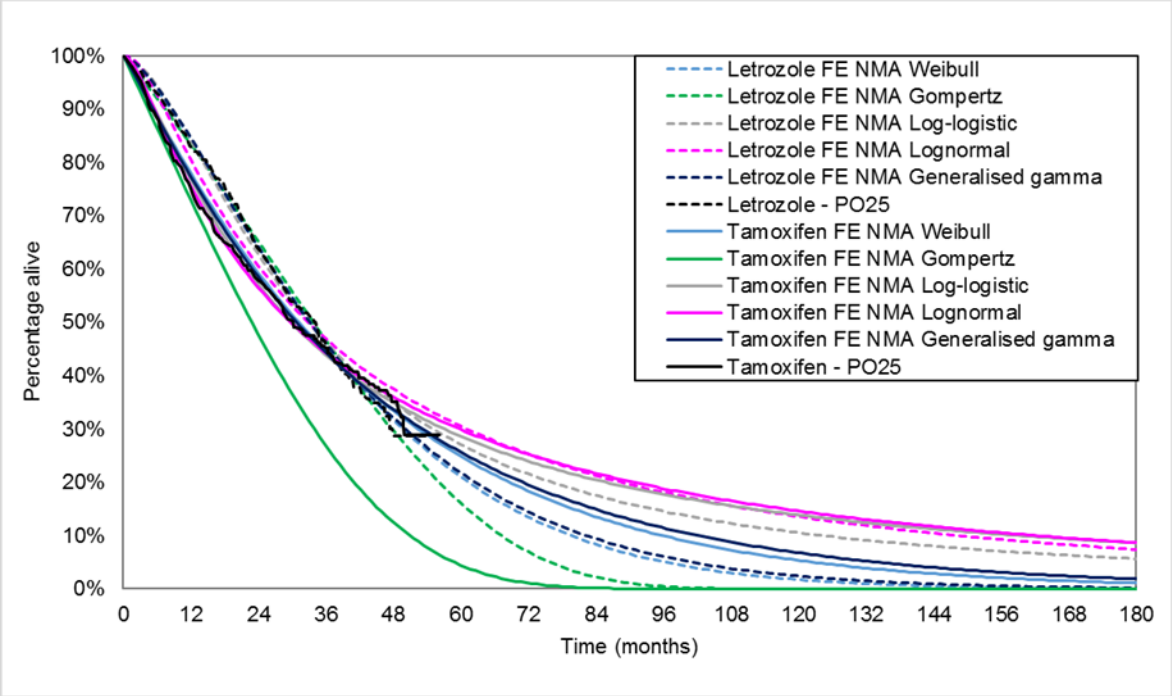


Figure 20 PO25 OS study fit with fixed effects 'all shapes' network meta-analysis model (CS Figure 37, p. 116)

The CS based its choice of the Weibull distribution as the best fit on visual inspection, AIC and BIC (Table 27) and clinical expert opinion[29] (Table 28). Advice from our clinical expert confirmed the estimates of OS for anastrozole (Table 28). The OS in the FALCON study (Figure 17) is immature as median survival had not been reached by the time of this submission; therefore the OS in the company model for fulvestrant is largely based on the FIRST trial. Therefore Figure 18 provides better insight regarding the suitability of the distributions explored.

The company carried out a sensitivity analysis using the gamma distribution (CS Table 98). The company considered that only the generalised gamma or the Weibull distribution was appropriate based upon the long-term extrapolations for these distributions compared to expert clinical opinion for anastrozole. The ERG carried out further analysis to explore the Gompertz, log-logistic and lognormal distributions (see section 4.4.1 for details). Based on our additional analysis, we consider that the choice of distribution does not have a significant effect on the ICER, except for the Gompertz distribution which does not provide a good fit to the FIRST study. We consider that the company's choice of distribution is reasonable based on the explanation in the CS.

Table 27 AIC and BIC statistics for OS based on NMA (CS Table 34, p. 114)

Distribution	AIC	AIC rank	BIC	BIC rank
Weibull	10499.131	1	10577.848	1
Generalised gamma	10500.300	2	10584.640	2
Gompertz	10508.995	3	10587.713	3
Log-logistic	10513.882	4	10592.599	4
Lognormal	10552.618	5	10631.335	5

AIC, Akaike's Information Criterion; BIC, Bayesian Information Criterion; NMA, network meta-analysis; PFS, progression-free survival.

Table 28 KOL opinion on OS at 1, 2, 5 and 10 years (CS Table 35, p. 117)

	1 year	2 years	5 years	10 years
KOL estimate	75-85%	55-70%	20-30%	5-10%

KOL, key opinion leader.

Response rate and other outcomes

The CBR, EDoR, ORR and other outcomes are discussed in section 3.3.3 of this report. These outcomes have no direct implication on progression or survival in the company's economic model. This seems to be because any impact on patients' survival is implicitly built into survival data

Time to treatment discontinuation

The model structure assumes that treatment duration is until objective disease progression, when patients switch over to second line treatment. CS Figures 43 and 44 compare the time to treatment discontinuation (TTD) and PFS curves for anastrozole and fulvestrant respectively in the FALCON trial. The curves are reasonably similar for anastrozole but there is a separation on the curves for fulvestrant. The company's approach in modelling was to use PFS as a proxy for TTD. We note that the company did not attempt to extrapolate TTD beyond the trial period. This limits the conclusions that can be drawn from CS Figures 43 and 44 and PFS may not be a good proxy for TTD with fulvestrant.

Adverse events

Only the FIRST and FALCON trials provided comprehensive individual patient level data on AEs. The North American trial reported certain AEs such as diarrhea, fatigue and nausea (see Table 41 of the CS). The CS reports that the incidence rates for AEs for fulvestrant and anastrozole were sourced from the FALCON trial, while rates for letrozole and tamoxifen were sourced from the literature. The company has given the paucity of data as the reason for not performing an indirect comparison of AE data (see page 146 of the CS). The company acknowledges that, as a result, the analysis may suffer some bias due to difference in follow-up periods and patient characteristics across the treatment groups.

AE rates for all four pharmacological agents are reported in CS Table 49 which is reproduced below (Table 29). The company model applies these event rates on a one-off basis, rather than as monthly rates applied throughout the time horizon of the model. The CS provides justification for this approach (CS section 5.3.2). The ERG is of the view that this approach is acceptable, given that the AEs are not expected to last beyond one year. While the CS states that events of grade ≥ 3 and experienced by 2% or more of patients in the treatment groups of interest are to be modelled for costs and utility impacts, we found that some AEs outside this definition were included in the model (e.g. Bilirubin increased). In the company's clarification response

(Question B2), it admits that such events should have been excluded. The ERG considered that this error was not likely to have a significant impact on the model outcomes. The ERG also spotted discrepancies regarding the rates reported in the CS and those used in the model (AST increased for tamoxifen), as well as differences in the rate reported in the CS and that reported in the literature and the CS/company model (dyspnoea for letrozole). The company acknowledged these errors in its clarification responses (Question B3). The ERG's view is that these errors have only a minor impact on the model results.

Table 29 Incidence rates of adverse events used in the model (CS Table 49, p. 147)

Adverse event	Fulvestrant	Anastrozole	Letrozole	Tamoxifen
Source:	FALCON[26]	FALCON[26]	Finn 2016[30]	Paridaens 2008[31]
Sample size (n)	228	232	222	189
ALT increased	1.3%	0.0%	0.0%	4.2%
AST increased	1.3%	0.4%	0.0%	1.6%
Hypertension	1.8%	1.7%	0.0%	3.2%
Pleural effusion	2.2%	0.4%	0.0%	0.0%
Pain, bone	0.4%	0.4%	0.0%	5.8%
Pain, other	1.3%	0.9%	1.4%	3.2%
Dyspnoea	0.0%	0.9%	0.5%	2.6%
Bilirubin increased	0.0%	0.4%	0.0%	1.6%

ALT, Alanine aminotransferase; AST, Aspartate aminotransferase.

Implications of survival parameters for cost-effectiveness

To predict the proportion of patients flowing from state to state per cycle throughout the modelled time horizon, the company estimated shape and scale parameters (and an additional Q parameter for the generalised gamma distribution) of PFS and OS from the fixed-effects NMA model. An 'All shapes' version of these parameters was then used in the company's health economic model for their base case cost-effectiveness results. The CS reports these parameter values in CS Table 73. We pointed out to the company that the values in CS Table 73 for PFS were not used in the model and this was acknowledged as a transcript error in the company's clarification response (Question B7). The actual model values are shown in CS Table 29 and in this report in Table 17.

One-way deterministic sensitivity analysis helped identify the top 10 parameters which cause the most significant change in the incremental cost-effectiveness ratio (ICER) for pair-wise comparisons between fulvestrant and the comparators. These analyses are discussed in section 4.3.10 of the ERG report. Choice of parameter inputs for OS are key model drivers of the cost effectiveness results.

4.3.5.2 Summary of ERG views on treatment effectiveness and extrapolation

One limitation with the comparison between anastrozole and fulvestrant stems from the immature OS data. Overall, the FIRST and FALCON trials seem to have been well conducted. The lack of individual patient level data from the PO25 trial makes it disparate in comparison with the FIRST and FALCON trials and not best suited for inclusion in the NMA. There are minor errors in the estimation of AE costs but these are unlikely to affect the conclusion from the results in a significant way. In general, we consider that the company's choice of base case distributions for extrapolating PFS and OS, are reasonable. As will be shown later, the results of cost-effectiveness are sensitive to survival outcomes, particularly for the OS scale and shape parameters, suggesting that longer term data from the FALCON could potential have a significant effect on the model results and more analysis might be needed to draw firm conclusions on the cost-effectiveness of anastrozole.

4.3.6 Health-related quality of life

Review of health-related quality of life

The company conducted a structured review to identify health state utility values for the economic evaluation. The EMBASE database was searched using the search strategy shown in CS Table 52. The search was for studies published between October 2013 and June 2016. The company chose this start date on the basis that this was the date of the search in the latest NICE Technology Appraisal for breast cancer. In their letter of clarification (Question B4), the company stated that this refers to the submission for trastuzumab emtansine for treating HER2-positive, unresectable locally advanced or metastatic breast cancer after treatment with trastuzumab and a taxane (TA371).[32] Other more recent appraisals for breast cancer such as TA424 (pertuzumab for neoadjuvant treatment of HER2-positive breast cancer),[33] TA421 (everolimus with exemestane for advanced breast cancer after endocrine therapy),[34] TA423 (erbulin for locally advanced or metastatic breast cancer after 2 or more chemotherapy regimens)[35] and the ongoing technology appraisal for palbociclib in combination with an aromatase inhibitor[36] did not identify any further utility studies.

The inclusion criteria included quality of life studies in patients with advanced or metastatic breast cancer. The search identified 354 studies. Titles and abstracts were screened and studies were excluded if they did not mention EQ-5D or QALYs, were not about breast cancer, were not written in English or were posters or conference abstracts that did not provide utility values. Thirteen studies that contained primary sources of utility data were identified. Of these studies eleven studies were excluded for not using EQ-5D to measure utilities or were published prior to 2013. A PRISMA flow diagram that shows the exclusion process is shown in CS Table 42.

Two studies were included in the review: Fukuda et al.[37] and Eyles et al.[38] A summary of these studies is shown in Table 30 (CS Table 53). Of these studies, the company suggests that the study by Fukuda et al.[37] is the most relevant as it was a randomised trial with a large sample of comparable patients with HER2 negative metastatic breast cancer receiving first-line therapy. EQ-5D values were calculated for up to 36 months and also post-progression. However, the CS notes that 57% of patients had received prior endocrine therapy.

Table 30 Metastatic cancer utility studies

Study [country]	Population / disease area (sample size)	Study design /intervention	Population, method of elicitation and valuation technique / tariff	Health states and/or treatment description	Mean
Fukuda et al. (2015)[37] Takashima et al. (2016)[39] [Japan]	HER2 negative metastatic breast cancer, resistant to endocrine therapy (57% previous endocrine treatment after recurrence)	Randomised open-label phase III trial 1L taxane (docetaxel or paclitaxel) vs. S-1	Patients EQ-5D-3L (pre-treatment, 3 months after randomisation, every 6 months thereafter) Tariff:Japanese	Mean EQ-5D scores up to 60 months (S-1)	0.748
				Mean EQ-5D scores up to 60 months (taxanes)	0.741
				During 1L – mean EQ-5D up to 36 months (S-1)	0.810
				During 1L – mean EQ-5D up to 36 months (taxanes)	0.781
				Post-progression period (S-1)	0.729

	Age (years), median (IQR): S-1 59.0 (53–65) and taxane 58.5 (51–65)			Post-progression period (taxanes)	0.703
Eyles et al. (2015) [England][38]	Metastatic breast cancer, stable disease Age (years): 37–65 Years since diagnosis, mean: 2.76 (0.5–7) Life expectancy: >6 months ECOG: 0–2 (n=19)	Feasibility study Mindfulness-based stress reduction for self-management of anxiety, depression, QoL, and fatigue	Patients EQ-5D-3L (baseline, during treatment [4 and 8 weeks] and follow-up [16 and 24 weeks]) Tariff: NR	Baseline	0.74
				End of follow-up	0.72
				End of follow-up (extreme outlier removed)	0.76

Abbreviation: 1L, first line; CI, confidence interval; ECOG, Eastern Cooperative Oncology Group; EQ-5D, EuroQol-5 Dimension; HER2: Human epidermal growth factor; IQR, interquartile range; NR, not reported; QoL, quality of life; SD, standard deviation; SG, standard gamble; TTO, time trade off; VAS, visual analogue scale.

The ERG considers the company's search to not be fully up-to-date and therefore we have updated the search until June 2017. We found four more potentially relevant primary studies that reported EQ-5D values in patients with advanced or metastatic breast cancer.[40-43] Three studies[40-42] were reported as conference abstracts and the details of these studies are shown in Table 31. The other study was a more detailed description of the study by Fukuda et al.,[43] shown in Table 30.

Table 31 Metastatic cancer utility studies identified in ERG update searches

Study [country]	Population / disease area (sample size)	Study design /intervention	Population, method of elicitation and valuation technique / tariff	Health states and/or treatment description	Mean
Lambert-Obry (2016)[42]	Postmenopausal women with ER+/HER2 negative locally advanced or metastatic breast cancer	Retrospective non-interventional study	EQ-5D 5L and WPAI questionnaire (at recruitment and at 3 and 6 months after recruitment)	First-line progression-free EQ-5D	0.73
				First-line progressed disease EQ-5D	0.62
Loibl (2016)[41]	HR+, HER2 negative metastatic breast cancer	Randomised trial (PALOMA-3) of palbociclib plus fulvestrant versus fulvestrant alone	EQ-5D 3L and EQ-5D VAS (at baseline and then on day 1 of each cycle until cycle 4 and then every alternate cycle until end of treatment)	Palbociclib plus fulvestrant EQ-5D	0.73
				Fulvestrant only EQ-5D	0.71
Mitra (2016)[40]	HR+/HER2 negative advanced or metastatic breast cancer	Multicenter real world study	EQ-5D 3L, EQ-5D VAS	EQ-5D all patients	0.73
				EQ-5D, 1st line	0.77
				EQ-5D 2nd line	0.69
				EQ-5D, 3rd and subsequent line	0.69

Abbreviation: EQ-5D, EuroQol-5 Dimension; HER2: Human epidermal growth factor; HR+: Hormone-receptor positive; IQR, interquartile range; VAS, visual analogue scale.

The company also reviewed utility values used in previous NICE breast cancer appraisals and details of these are shown in CS Table 54 - 55. The CS reports that four primary studies are used in the NICE Technology Appraisals for ABC and all these studies used the standard gamble to elicit utilities from either the general public or from medical personnel. These studies do not meet the NICE reference case criteria, as HRQoL have not been directly measured from patients. The most common utility study used in previous technology appraisals was by Lloyd et al.[44] but this study has been criticized by previous ERG reports for other STA appraisals for not

meeting the NICE reference case and that the utility values derived may not be reflective of patients with breast cancer.

Health-related quality of life from clinical trials

The FALCON trial[8] collected HRQoL data including the EQ-5D 3L, using the UK tariff (section 3.3.5). The questionnaire was administered at baseline and every 12 weeks until disease progression or treatment discontinuation. For patients whose disease had progressed, the questionnaire was administered three months after disease progression and then at 6-monthly intervals. Health state utility values for progression-free and progressed disease are shown in Table 32 (CS Table 50) for patients treated with fulvestrant, anastrozole and all patients. The CS states that the mean EQ-5D values were similar across treatments with overlapping 95% CIs. For this reason, the company used the same utility values for both treatments. The ERG agrees that it is reasonable to use the same utility values for patients receiving fulvestrant and anastrozole.

Table 32 Health state utility values from the FALCON trial

Treatment	Health state		ITT	
		n	Mean	95% CI
Overall	Progression-free	449	0.75	[0.73, 0.77]
	Progressed disease	232	0.69	[0.65, 0.72]
Fulvestrant	Progression-free	225	0.76	[0.73, 0.78]
	Progressed disease	104	0.69	[0.63, 0.74]
Anastrozole	Progression-free	224	0.74	[0.71, 0.76]
	Progressed disease	128	0.69	[0.63, 0.74]

Abbreviations: CI, confidence interval; ITT, intention-to-treat.

The company uses the utility values from the FALCON trial for all patients for the health state utility values in the economic model. However the company adjusted these values using repeated measures mixed effects regression models (MMRMs) "*in order to take account of the repeated measures per patient, and estimate the association between utilities and clinical events in the FALCON study"*. The company included two mixed models: MMRM (1) included only a coefficient for disease progression, while MMRM (2) included coefficients for patient characteristics. The company preferred MMRM (1) because the coefficients used in MMRM (2) were not statistically significant. The ERG agrees with this choice and presumes that the values

from MMRM (1) are equivalent to those for the mean unadjusted utility values shown in Table 32, as there is only one coefficient for progression. The utility value used for the progression-free health state in the model is 0.7511 and for the progressed state is 0.6913.

The company has also provided the utility values at different time points from baseline (CS Figure 65 and 67). The CS comments that the utility values collected in the FALCON trial are higher than those used in previous appraisals but they are preferred because they align with the NICE reference case in the use of EQ-5D data collected in a patient population as specified in the decision problem. Further the CS comments that utility values have face validity as the utility estimates are lower than the EQ-5D population norms for this age and sex group (Kind et al.[45]).

The ERG agrees with company's use of the health state utilities from the FALCON trial in the economic model and considers that the utility values collected are an improvement on the data used in previous technology appraisals for advanced and metastatic breast cancer. As noted above, the utility values have been collected in the same patient group as specified in the NICE scope[1] and the methodology used is consistent with the NICE reference case. Further, the ERG considers that the utility values are consistent with those collected by Fukuda et al.,[37] in terms of the difference between the utility values for progression-free and progressed health states.

Adverse event disutilities

The company includes disutility for AEs. These are applied to grade 3/4 AEs and applied for the duration of the AE. The disutility values were taken from previous NICE submissions, as shown in Table 33 (CS Table 57).

Table 33 Disutilities associated with AEs

Adverse event	Utility decrement per event	Duration (days)	Source
ALT increased	-0.050	28.0*	Boehringer Ingelheim Ltd. (2014) [46]
AST increased	0.000	0.000	
Hypertension	-0.153	8.0	Swinburn et al. (2010)[47]
Pleural effusion	-0.371	3.0	Swinburn et al. (2010)[47]
Pain, bone	-0.069	17.0	Doyle et al. (2008)[48]
Pain, other	-0.069	17.0	Doyle et al. (2008)[48]
Dyspnoea	-0.05[b]	12.7	Doyle et al. (2008)[48]
Bilirubin increased	0.000	0.000	

CS Table 57
Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase.
*Assumption
[b] Value corrected in company clarification response

In a clarification response (Question B5), the company noted that the disutility value for dyspnoea should be 0.05, rather than the initial value of -0.103, and that there is a mistake in the disutility values for pleural effusion, which had been confused with palmar-plantar erythrodysesthesia in the study by Swinburn and colleagues.[47] They have been unable to identify an alternative published value for pleural effusion.

The ERG considers the company's approach to including disutilities for AEs in the economic model is reasonable and notes that the effect of AE disutilities on the model results is negligible due to the low frequency of SAEs.

4.3.7 Resource use and costs

The economic evaluation includes costs for disease management, treatment acquisition, treatment administration, subsequent therapy and AEs. Unit costs for health care resources were taken from National Reference Costs[49] and PSSRU Unit costs of health and social care.[50] The company did not conduct a review of resource use in ABC.

Treatment cost and resource use

The dosing schedules for fulvestrant and its comparators are shown in Table 34 (CS Table 63). The dosing information is taken from the British National Formulary (BNF).[51] The recommended dose for fulvestrant is 500 mg, administered twice in the first two weeks and monthly thereafter. Fulvestrant is administered in the outpatient setting by IM injection into the buttocks. The dosing schedule is consistent with that used in the FALCON trial. The unit cost for fulvestrant is £522.41 per dose. Anastrozole, letrozole and tamoxifen are oral treatments and all cost less than £2 per 4 week treatment cycle. In the model, patients are treated until disease progression, on the basis that the treatment discontinuation and disease progression curves were similar.

Table 34 Treatment dosing, administration and drug acquisition costs

		Fulvestrant (first 4 weeks)	Fulvestrant (after first 4 weeks)	Anastrozole	Letrozole	Tamoxifen
Label information	**Administration method**	IV	IV	Oral	Oral	Oral
	Dose per administration (mg)	500	500	1.0	2.5	2.5
	Administration frequency	2 per 4 weeks	1 per 4 weeks	1 per day	1 per day	1 per day
Package information	**Formulation (mg)**	250	250	1.0	2.5	20
	Pack size	2	2	28	28	30
	Cost per pack (£)	£522.41	£522.41	£0.75	£1.52	£1.62
Dosing required in model	**Required dose (mg)**	500	500	1.0	2.5	20
	Vials/ capsules per administration	1	1	0.04	0.04	0.03
Relative dose intensity/ compliance		1.00	0.99	0.99	1.00*	1.00*
Drug cost per 4-week cycle		£1,044.82	£522.41	£0.75	£1.52	£1.51
Administration cost 1st cycle		£370.35	-	£196.64	£196.64	£196.64
Administration cost subsequent cycles		-	£73.74	£27.93	£27.93	£27.93

CS Table 63
Abbreviations: IV, intravenous. *Assumption (data not available)

The approach for calculating administration costs is based on the previous NICE appraisal for fulvestrant (TA239).[21] The administration costs differ between the first four-week cycle and subsequent cycles. The treatment-related administration costs for fulvestrant 500 mg in the first

month is £370.35, which includes an initial visit with the oncologist for the initial dose (£199.64), the administration of fulvestrant by a clinical nurse (£99.97), plus the average cost of administrating the dose two weeks later, assuming 32% are administered in the primary care setting and 68% are administered in the secondary care outpatient setting (£18.75). The cost for administering fulvestrant in subsequent cycles is £73.74. One of our clinical experts stated that all patients in their locality would be administered in the hospital setting and that it may be a challenge persuading primary care to take on treatment delivery. The ERG has run an analysis assuming that all patients receiving fulvestrant are treated in the outpatient setting and none receive the treatment in primary care (section 4.4).

Anastrozole, letrozole and tamoxifen are oral medications and the only administration costs are the cost of prescription each month (after the first month) and this was assumed to be a telephone consultation with general practitioner lasting 7.1 minutes (£27.93). The initial cycle includes a visit with the oncologist (£199.64).

Disease management costs

Disease management costs are included in the model for the progression-free and progressed health states and for a one-off cost of terminal care. The company did not collect health care resource use data for the FIRST or FALCON trials. Health-state costs are taken from the NICE clinical guidelines for ABC (CG81)[2] using the resources specified for 'Package 1' and 'Package 2'. Unit costs have been inflated to 2015-6 using the PSSRU Hospital and Community Health Services (HCHS) indices.[50] The resource use and unit costs are shown for the progression-free and progressed health states in Table 35 and Table 36 respectively (CS Table 60 - 61).

Table 35 Costs of progression-free health state

Items	Resource usage per 4 weeks	Frequency	Unit cost (£) inflated to 2015/16	Total cost per month	Source*
Community nurse (home visit - 20 minutes)	2	1 per 2 weeks	£14.67	£29.34	PSSRU 2015/16
GP contact (surgery visit – 11.7 minutes)	1	1 per month	£46.02	£46.02	PSSRU 2015/16
Clinical nurse specialist (1 hour)	1	1 hour every month	£108.00	£108.00	PSSRU 2015/16
Total progression-free cost per 4 weeks				£183.36	Calculation

Abbreviation: GP, General Practitioner.

*PSSRU 2015 used to provide duration of appointment time; PSSRU 2016 used to provide unit costs.

Table 36 Costs of progressed disease health state

Resource	Resource usage per 4 weeks	Frequency	Unit cost (£)	Total cost per 4 weeks (£)	Source*
Community nurse (home visit 20 minutes)	4	1 per week	£14.67	£58.67	PSSRU 2015/16
Consultation with a GP (home visit)	2	1 per 2 weeks	£65.00	£130.00	PSSRU 2015/16
Clinical nurse specialist (duration 1 hour)	4	1 per week	£108.00	£432.00	PSSRU 2015/16
NHS community occupational therapist	2	1 per 2 weeks	£42.00	£84.00	PSSRU 2015/16
Total progressed disease cost per 4 weeks				£704.67	Calculation

Abbreviations: GP, General Practitioner; NHS, National Health Service; PSSRU, Personal Social Services Research Unit.

*PSSRU 2015 used to provide duration of appointment time; PSSRU 2016 used to provide unit costs.

The ERG notes that the resources described in the NICE clinical guidelines refer to patients receiving chemotherapy, rather than patients receiving endocrine therapy and therefore do not appropriately estimate the resource used in this submission. For example, our clinical experts stated that patients treated with endocrine therapy would not receive home visits from a nurse. Furthermore they stated that patients would see a medical oncologist regularly (every three months) and this resource has not been included in the model. The ERG considers that the resources would be more consistent with previous clinical trials for AIs (such as Karnon et al.[52]) and provides an analysis with alternative resource use in section 4.4.

Terminal care costs

Terminal care costs are included in the model for patients with progressed disease for the end of patients' life and consist of time spent either in the hospital, hospice or at home. Based on NICE clinical guidance CG81,[2] the company assumes that 40% of patients died at the hospital, 10% at a hospice and 50% at home. The unit costs from CG81 were inflated to 2015/16 costs using the HCHS index.[50] The total terminal care cost per patient in the model is £4,379.03.

Subsequent therapy

The economic model includes subsequent lines of treatment for patients whose disease progresses. Second-line and third-line therapies include further endocrine therapy (fulvestrant, anastrozole, letrozole, exemestane, tamoxifen), targeted therapies (everolimus plus exemstane), chemotherapy (docetaxel, capecitabine, paclitaxel, erbulin) or no treatment. The proportions of patients receiving subsequent therapies and the treatment durations are based upon Kurosky et al.,[53] a retrospective cohort study of postmenopausal patients with metastatic ER+, HER2 negative breast cancer in the UK.

The proportions of patients receiving second-line and third-line therapy are shown in Table 37 (CS Table 66). It was assumed in the model that all patients that initiated first-line treatment received second-line treatment and 54.41% of patients who received second-line treatment received third-line treatment.

Table 37 Proportion of patients using subsequent treatments in the second- and third-line settings

From primary treatment to →→→ / Setting	Endocrine therapy (%)	Targeted therapy (%)	Chemotherapy (%)	No treatment (%)	Total (%)
Second-line	54.35%	8.08%	37.57%	0.00%	100.00%
Third-line	24.02%	0.00%	30.39%	45.59%	100.00%

Based on Kurosky et al.,[53] patients on endocrine therapy were assumed to receive treatment for 9.16 months for second-line and 6.17 months for third-line.

Dosing schedules, unit costs and administration costs for the chemotherapy treatment and the targeted therapies are shown in CS Table 68-69. A weighted cycle cost was calculated for the first and subsequent cycles for second-line and third-line treatment for the endocrine therapies, targeted therapies and chemotherapies (CS Table 70). It was assumed that patients starting on fulvestrant would not receive fulvestrant as a second-line or third-line therapy. For all other initial therapies subsequent treatment options would be the same. The weighted average costs of the subsequent therapies are shown in CS Table 71.

The ERG notes that in the population in the Kurosky retrospective study about a third of patients were initially diagnosed at early stage breast cancer and of these the majority received surgery and adjuvant endocrine therapy. Furthermore, only 49.3% of patients received endocrine therapy as first-line therapy. The ERG consulted their clinical advisors on the proportion of patients receiving endocrine therapy as subsequent therapy. Their view was that the proportion of patients receiving endocrine therapy as second-line treatment would be higher and in the region of 67-80% with fewer patients receiving chemotherapy. The ERG has conducted an analysis varying the proportions of patients receiving subsequent treatment in section 4.4.

Adverse event costs

The costs of treating treatment-related AEs are shown in CS Table 72. The costs are taken from National Reference costs 2015-16[49] and the cost codes are based upon those reported in previous NICE appraisals.

4.3.8 Model validation

In line with the recommendations developed by a task force of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and the Society for Medical Decision Making (SMDM)[54] for model quality assurance, the ERG checked the economic model for transparency and validity. These are discussed below.

Model transparency

The CS clearly described the model structure, parameter values and their sources, data identification methods, and assumptions used in the model. The model was technically transparent and the visual basic code used within the model was accessible. In general, the technical report and the model described the analyses clearly and provided adequate information to assess the model. The CS clearly presented the results of the NMA but did not present the WinBUGS code used to derive those results.

Model validation

To validate the economic model, the company stated that the model was reviewed by their internal health economists. They undertook an assessment of the face-validity of the model, and conducted a third-party validation of the model calculations and the data sources. Extreme value and log tests were also conducted by the company to examine if the model behaved as expected and that the results obtained were logical.

The ERG checked the model for internal as well as external validity. The step-by-step approach used for this purpose is discussed below.

- *Face validity*

The company conducted an extensive review of the existing NICE appraisals in advanced/metastatic breast cancer in May 2016 to inform their modelling approaches. They also conducted a structured review of utility studies in June 2016 to inform the quality of life parameters. The opinions of seven UK clinical experts were used to validate the extrapolation of PFS and OS within the company's analyses. The CS compared the long-term predicted model outcomes for PFS and OS with the corresponding clinical expert opinion as shown in Table 38. The modelled outcomes appeared comparable with the expert opinion.

Table 38 Comparison of predicted model outcomes with those of clinical opinions

Outcomes	Time-frame			
	1 year	2 years	5 years	10 years
PFS				
KOL opinion (anastrozole)	50-60%	30-40%	5-10%	1-5%
Modelled PFS (anastrozole)	52.2%	25.7%	4.6%	0.6%
Modelled PFS (letrozole)	59.3%	30.8%	5.8%	0.7%
OS				
KOL opinion (anastrozole)	75-85%	55-70%	20-30%	5-10%
Modelled OS (anastrozole)	86.0%	69.6%	30.7%	5.5%
Modelled OS (letrozole)	91.5%	74.5%	23.2%	0.7%

Source: CS Table 108 & 109. KOL: Key Opinion Leader; PFS: Progression-free survival; OS: overall survival.

The company did not provide any further details if the third-party constituted experts from clinical and/or health economic backgrounds. The CS also did not explicitly document the steps taken to validate the model calculations and the data sources. Therefore, the ERG is unable to comment on these. Further, no information was presented to ascertain if the model assumptions were validated by clinicians or experts.

- *Internal validity*

Internal validity checks consist of two main steps: checking the individual equations within the model; and verifying their accurate implementation in code.[54]

Although the company cited a number of internal validity checks, they did not present any formal checklist for quality assurance of the model used by their health economists. Below is a summary of the checks conducted by the ERG to assess the internal validity of the model:

i. Individual equations were checked for their mathematical correctness. However, due to time constraints, the ERG focused primarily on the equations defining survival functions, patient transition in different health states, costs, QALYs, and overall results. Within the costs calculations, the ERG identified errors in estimating the discounted costs. The company rectified these errors and submitted new sets of base case results in the clarification response (Clarification response Appendix Table 30). The ERG were able to reproduce the new sets of base case results of the CS.

ii. The visual basic programming code within the model was checked and appeared to be correct, except for a few minor errors in the model. These errors affected cosmetic features of the model and did not have any impact on the overall model calculations or results.

iii. The ERG checked for consistency of the parameters reported in the technical document and those utilised within the model. There were minor reporting errors in CS Table 77 which the company rectified in their clarification response (Clarification response A13.1(c) Table 26).

iv. The ERG conducted a range of extreme value and logic tests to check the plausibility of changes in results when parameters are changed. The list of the tests conducted is presented in Appendix 1.

Based on the checks conducted as stated above, the company's model had a few calculation errors, although the overall technicalities of the model appeared to be correct. Rectifying the calculation errors did not have significant impact on the overall base case model results.

- *External validity*

The company presented comparisons of the modelled outcomes for PFS and OS with the results of a systematic review and previous HTA assessments, shown in CS Table 79 and Table 80, respectively. These are reproduced below in Table 39. The results obtained from the model appeared to be comparable with the existing evidence.

Table 39 Comparison of the modelled outcomes with other sources

Treatment	Median PFS (months)		
	Model outcomes	**Systematic literature review**	**Previous HTA assessments**
Fulvestrant	16.56	Range: 16.6 – 25.9	NA
Anastrozole	11.96	Range: 12.9 – 14.8	NA
Letrozole	14.72	9.60	14.5
Tamoxifen	9.20	Range: 5.9 – 10.4	NA

Treatment	Median OS (months)		
	Model outcomes	Systematic literature review	Previous HTA assessments
Fulvestrant	47.84	62.5	NA
Anastrozole	39.56	Range: 44.9 – 46.5	NA
Letrozole	38.64	34	33.3
Tamoxifen	36.80	Range: 30.3 – 43.6	NA

Source: CS Table 79 & 80; HTA: Health Technology Assessment; NA: Not available; PFS: Progression-free survival; OS: Overall Survival

In addition to the above analyses, the ERG compared the predicted OS data for fulvestrant and anastrozole with the observed data from the FIRST trial (using the matched population alone) as shown in Figure 21. The graph shows that the predicted OS data provided a reasonable comparison of the observed data in the FIRST trial.

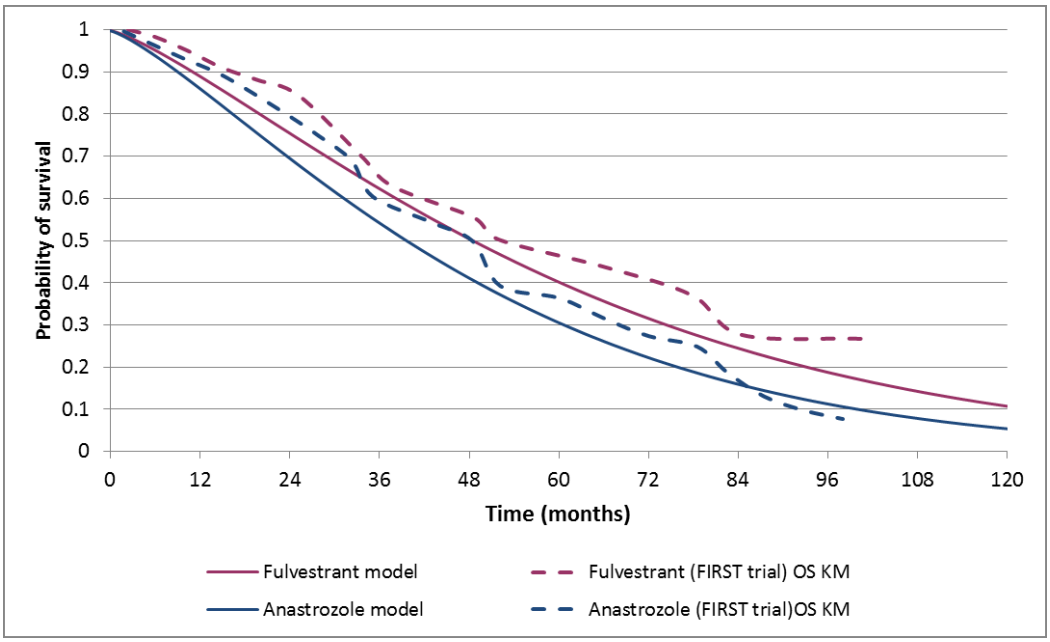


Figure 21 Comparison of predicted OS data against the observed data from the FIRST trial (using the matched patients only)

- *Cross validity and predictive validity*

Cross validation checks, which involve assessing different mathematical models addressing the same decision problem, were not relevant from the perspective of this technology appraisal as there are no existing models with the same decision problem for the same drug. Fulvestrant is a *de novo* intervention for post-menopausal people with locally advanced or metastatic HR+ breast cancer who had not received endocrine therapy. The ERG did not perform any checks on predictive validity of the economic model.

4.3.9 Cost effectiveness results

The company presented base case results in terms of total costs, life years gained, QALYs and incremental cost per QALY. Results were presented as pair-wise comparisons of fulvestrant versus anastrozole, letrozole and tamoxifen (CS Table 74 – 76) along with an incremental analysis of fulvestrant versus AIs (CS Table 77). As mentioned in section 4.3.8, the company rectified a few calculation errors for the costs in the economic model and submitted new sets of results for the base case analyses with the clarification response. The results presented in the following sections of this appraisal are based on the corrected economic model.

Results of the incremental analysis of fulvestrant versus comparators are summarised below in Table 40. The results are presented in order of increasing costs. Letrozole was associated with lowest overall costs. Tamoxifen was dominated as it was associated with comparatively higher costs and lower QALYs when compared against anastrozole in the incremental analysis; thereby resulting in an incremental ICER of £34,099 for fulvestrant versus anastrozole.

Table 40 Results of incremental analysis (based on corrected economic model)

Treatments	Total costs	Total QALYs	Incremental costs	Incremental QALYs	Incremental ICER (cost per QALY)
Letrozole	£26,221	2.46	--	--	--
Anastrozole	£30,572	2.68	£4,351	0.22	£19,702
Tamoxifen	£32,328	2.47	£1,756	-0.21	Dominated
Fulvestrant	£49,431	3.23	£18,859	0.55	£34,099

This table draws on information presented in the Table 30 in the appendix to the company's written response to clarification questions.

No sub-group analysis was conducted as part of the submission. This was considered appropriate and aligned with the final NICE scope.

4.3.10 Assessment of uncertainty

In accordance with the NICE final scope,[1] the company assessed methodological, structural and parameter uncertainties associated with the base-case analyses by conducting a range of deterministic sensitivity-, probabilistic sensitivity- and scenario- analyses, details of which are discussed below.

Deterministic sensitivity analysis

Deterministic sensitivity analysis (DSA) was conducted on a number of key parameter groups. The parameters and their ranges are shown in Table 41. In general, the choice of parameters included and the ranges for variation appeared to be reasonable, although the ERG viewed that it would have been more appropriate to use a range of 95% confidence intervals for the health state utilities.

Table 41 Parameters and their ranges used for deterministic sensitivity analyses

Parameters	Range
Parametric survival distribution parameters	95% confidence interval
Disease management costs	20% of the mean values
Terminal care/ end of life costs	20% of the mean values
Treatment acquisition and administration (per 4 weeks)	20% of the mean values
Health state utilities	10% of the mean values
Discount rates	0% to 6%

The company produced tornado plots for the 10 most sensitive parameters for each of the comparisons. The ERG observed that, unlike those presented in the CS, the tornado plots programmed in the model excluded the parameters for parametric survival distributions. We were, therefore, unable to reproduce the same sets of top 10 sensitive parameters by running the 'Update DSA' button within the economic model as reported in the CS Figure 51-53. Owing to this limitation, we reproduced the results of the DSA for fulvestrant versus comparators in Table 42, Table 43 and Table 44 based on the corrected model for the parameters that were reported in CS Tables 94-96.

Table 42 Results of DSA - fulvestrant versus anastrozole (based on corrected model)

Parameter	Base case (ICER)	Lower value (ICER)	Upper value (ICER)
(OS) fulvestrant: Weibull scale parameter	£34,099	£338,729	£23,236
Health state utilities: PF	£34,099	£42,187	£28,613
Discount rate - Outcomes	£34,099	£27,193	£39,387
Treatment acquisition costs per 4 weeks: fulvestrant	£34,099	£28,371	£39,827
Discount rate - Costs	£34,099	£38,592	£31,660
(OS) anastrozole: Weibull scale parameter	£34,099	£36,757	£31,584
(PFS) anastrozole: gamma scale parameter	£34,099	£31,560	£36,791
(OS) fulvestrant: Weibull shape parameter	£34,099	£31,031	£35,450

Abbreviations: ICER, incremental cost-effectiveness ratio; PF, progression-free; PFS, progression-free survival; OS, overall survival.

Table 43 Results of DSA - fulvestrant versus letrozole (based on corrected model)

Parameter	Base case (ICER)	Lower value (ICER)	Upper value (ICER)
(OS) letrozole: Weibull scale parameter	£29,991	£23,917	£94,487
(OS) fulvestrant: Weibull scale parameter	£29,991	£63,332	£22,677
(PFS) letrozole: gamma scale parameter	£29,991	£24,832	£37,963
Discount rate - Outcomes	£29,991	£23,213	£35,521
Treatment acquisition costs per 4 weeks: fulvestrant	£29,991	£25,897	£34,084
Discount rate - Costs	£29,991	£34,864	£27,352
Health state utilities: PD	£29,991	£31,608	£28,531
Health state utilities: PF	£29,991	£31,531	£28,594

Abbreviations: ICER, incremental cost-effectiveness ratio; PF, progression-free; PFS, progression-free survival; OS, overall survival.

Table 44 Results of DSA - fulvestrant versus tamoxifen (based on corrected model)

Parameter	Base case (ICER)	Lower value (ICER)	Upper value (ICER)
(OS) tamoxifen: Weibull scale parameter	£22,498	£19,408	£40,262
Health state utilities: PF	£22,498	£25,502	£20,495
Treatment acquisition costs per 4 weeks: fulvestrant	£22,498	£18,330	£26,665
Discount rate - Outcomes	£22,498	£17,981	£25,976
Discount rate - Costs	£22,498	£26,239	£20,495
(PFS) tamoxifen: gamma scale parameter	£22,498	£19,975	£25,710
(PFS) tamoxifen: gamma shape parameter	£22,498	£21,151	£24,158
(OS) fulvestrant: Weibull scale parameter	£22,498	£41,586	£18,470

Abbreviations: ICER, incremental cost-effectiveness ratio; PF, progression-free; PFS, progression-free survival; OS, overall survival.

None of the sensitivity analyses reduced the ICER for fulvestrant compared to the AIs (i.e. analstrozole and letrozole) to below £20,000 per QALY. When fulvestrant was compared to the AIs, the model results were most sensitive to the OS parameters. For example, the ICERs ranged from £23,236 per QALY for the lower value of the OS scale parameter to £338,729 per QALY for the upper value of the parameter for fulvestrant versus anastrozole. The ICERs were also sensitive to the PFS parameters and moderately sensitive to the health state utilities, discount rates and treatment acquisition costs for fulvestrant.

The results of the DSA show that the OS parameters had the most influence on the base case model results with a wide range in the ICERs obtained from using the upper and lower values for this parameter. This indicated a considerable amount of uncertainty in the model results.

Scenario analysis

The company analysed structural and methodology uncertainties by performing a range of scenario analyses. These analyses and their justifications, reproduced from CS Table 97, are presented in Table 45.

Table 45 List of scenario analyses conducted by the company

Variables	Base case	Scenario	Rationale
OS extrapolations	'All shapes' NMA model OS - Weibull	'All shapes' NMA model plausible extrapolations: OS - generalised gamma	To assess the impact of a range of survival estimates
PFS extrapolations	'All shapes' NMA model PFS - generalised gamma	'All shapes' NMA model: PFS - log-logistic PFS - lognormal PFS - Weibull PFS - Gompertz	To assess the impact of a range of survival estimates
OS and PFS extrapolations	'All shapes' NMA model: OS - Weibull PFS - generalised gamma	'No shape arm' NMA model: OS - Weibull PFS - Weibull PFS - Gompertz	To assess the impact of not adjusted for differences in shapes between treatment arms
OS and PFS extrapolations	'All shapes' NMA model: OS - Weibull PFS - generalised gamma	Assume equivalent efficacy between AIs 'All shapes' NMA model (anastrozole curves used for letrozole): OS - Weibull PFS - generalised gamma	To assess the impact of commonly held clinical opinion that AIs have equal efficacy
Utility values	FALCON MMRM (1)	FALCON summary statistics; FALCON MMRM (1) and Lloyd (2006); Lloyd (2006)	To assess the impact of using alternative data sources for health state utility values
Time horizon	30	5; 10; 15; 20; 25; 35	To assess the impact of varying the time horizon.
Discount rate	3.5% for both costs and outcomes	1.5% for both costs and outcomes	NICE guidelines

Variables	Base case	Scenario	Rationale
AEs	AE costs and disutilities	No AE costs and disutilities	To assess the impact of inclusion of AE costs and disutilities on cost-effectiveness results
Treatment administration costs	Inclusion of administration costs for oral treatments	Exclusion of administration costs for all comparator therapies	To assess the impact that oral treatments are self-administered by the patient
Subsequent treatment costs and end of life care	Exclusion of fulvestrant as a subsequent treatment option for patients on first-line fulvestrant	Same subsequent treatment costs for all patients Exclusion of subsequent treatment costs altogether	To assess the impact of subsequent treatment overall and whether patients initially treated with fulvestrant will receive it again as a subsequent therapy

Source: CS Table 97

Results of the scenario analyses are presented in CS Table 98 – 99, 102 – 107. The ERG re-ran all the scenarios with the corrected model and have updated the results below in Table 46.

Table 46 Summary of the scenario analyses (based on the corrected economic model)

Parameters		Base case ICER	Scenario ICER
Scenario 1: OS generalised gamma; PFS: generalised gamma			
Fulvestrant vs	Letrozole	£29,991	£28,665
	Anastrozole	£34,099	£33,387
	Tamoxifen	£22,498	£22,183
Scenario 2: OS Weibull; PFS: various distributions			
OS Weibull; PFS Weibull			
Fulvestrant vs	Letrozole	£29,991	£28,488
	Anastrozole	£34,099	£33,079
	Tamoxifen	£22,498	£23,050
OS Weibull; PFS Gompertz			
Fulvestrant vs	Letrozole	£29,991	£30,267
	Anastrozole	£34,099	£33,551
	Tamoxifen	£22,498	£24,442
OS Weibull; PFS log-logistic			
Fulvestrant vs	Letrozole	£29,991	£31,458
	Anastrozole	£34,099	£35,252
	Tamoxifen	£22,498	£22,625
OS Weibull; PFS lognormal			
Fulvestrant vs	Letrozole	£29,991	£32,048
	Anastrozole	£34,099	£33,986
	Tamoxifen	£22,498	£22,233

Scenario 3: 'No shape arm' with OS			
		Base case (all shape model)	Base case (no shape model)
Fulvestrant vs	Letrozole	£29,991	--
	Anastrozole	£34,099	--
	Tamoxifen	£22,498	--
OS: Weibull; PFS: Weibull			
Fulvestrant vs	Letrozole	£29,991	£37,358
	Anastrozole	£34,099	£33,710
	Tamoxifen	£22,498	£25,036
OS: Weibull; PFS: Gompertz			
Fulvestrant vs	Letrozole	£29,991	£36,293
	Anastrozole	£34,099	£33,687
	Tamoxifen	£22,498	£25,210
OS: Weibull; PFS: Log-logistic			
Fulvestrant vs	Letrozole	£29,991	£46,189
	Anastrozole	£34,099	£39,664
	Tamoxifen	£22,498	£29,001
OS: Weibull; PFS: Lognormal			
Fulvestrant vs	Letrozole	£29,991	£45,356
	Anastrozole	£34,099	£38,753
	Tamoxifen	£22,498	£29,308
Scenario 4: Assuming equal efficacy between AIs using the anastrozole parametric survival models			
OS: Weibull; PFS: generalised gamma			
Fulvestrant vs	Letrozole	£29,991	£34,140
	Anastrozole	£34,099	£34,099
	Tamoxifen	£22,498	£22,498
OS: Weibull; PFS: Weibull			
Fulvestrant vs	Letrozole	£29,991	£33,123
	Anastrozole	£34,099	£33,079
	Tamoxifen	£22,498	£23,050
OS: Weibull; PFS: Gompertz			
Fulvestrant vs	Letrozole	£29,991	£33,597
	Anastrozole	£34,099	£33,551
	Tamoxifen	£22,498	£24,442
OS: Weibull; PFS: Log-logistic			
Fulvestrant vs	Letrozole	£29,991	£35,284
	Anastrozole	£34,099	£35,252
	Tamoxifen	£22,498	£22,625
OS: Weibull; PFS: Log-normal			
Fulvestrant vs	Letrozole	£29,991	£34,022
	Anastrozole	£34,099	£33,986
	Tamoxifen	£22,498	£22,233

Scenario 5: Utility values			
FALCON summary statistics			
Fulvestrant vs	Letrozole	£29,991	£30,042
	Anastrozole	£34,099	£34,151
	Tamoxifen	£22,498	£22,530
FALCON study MMRM model (1) and Lloyd (2006)			
Fulvestrant vs	Letrozole	£29,991	£35,211
	Anastrozole	£34,099	£34,516
	Tamoxifen	£22,498	£21,390
Lloyd (2006)			
Fulvestrant vs	Letrozole	£29,991	£34,921
	Anastrozole	£34,099	£34,281
	Tamoxifen	£22,498	£21,256
Scenario 6: Different time horizons			
5 years			
Fulvestrant vs	Letrozole	£29,991	£80,244
	Anastrozole	£34,099	£61,423
	Tamoxifen	£22,498	£35,472
10 years			
Fulvestrant vs	Letrozole	£29,991	£33,750
	Anastrozole	£34,099	£38,457
	Tamoxifen	£22,498	£24,245
15 years			
Fulvestrant vs	Letrozole	£29,991	£30,575
	Anastrozole	£34,099	£34,986
	Tamoxifen	£22,498	£22,815
20 years			
Fulvestrant vs	Letrozole	£29,991	£30,132
	Anastrozole	£34,099	£34,344
	Tamoxifen	£22,498	£22,580
25 years			
Fulvestrant vs	Letrozole	£29,991	£30,032
	Anastrozole	£34,099	£34,171
	Tamoxifen	£22,498	£22,521
35 years			
Fulvestrant vs	Letrozole	£29,991	£29,973
	Anastrozole	£34,099	£34,067
	Tamoxifen	£22,498	£22,487
Scenario 7: Discount rates of 1.5% for both costs and outcomes			
Fulvestrant vs	Letrozole	£29,991	£28,223
	Anastrozole	£34,099	£32,179
	Tamoxifen	£22,498	£21,609
Scenario 8: Exclusion of AE costs and dis-utilities			
Fulvestrant vs	Letrozole	£29,991	£29,861
	Anastrozole	£34,099	£33,990
	Tamoxifen	£22,498	£22,756

Scenario 9: Zero administration costs for comparator (oral) treatments			
Fulvestrant vs	Letrozole	£29,991	£31,039
	Anastrozole	£34,099	£35,424
	Tamoxifen	£22,498	£23,235
Scenario 10: Different assumptions regarding subsequent treatment costs			
Same subsequent treatment costs for all treatments			
Fulvestrant vs	Letrozole	£29,991	£30,377
	Anastrozole	£34,099	£34,639
	Tamoxifen	£22,498	£22,890
Exclusion of subsequent treatment costs			
Fulvestrant vs	Letrozole	£29,991	£30,799
	Anastrozole	£34,099	£35,232
	Tamoxifen	£22,498	£24,217

In scenario 1, the company changed the OS distribution for the generalised gamma distribution alone. The ERG felt that, for completeness, the company should have also presented results for all the other distributions (Gompertz, log-logistic, lognormal). This is explored in the ERG additional analyses in section 4.4.

In scenario 2, assigning various distributions to PFS resulted in the ICER of fulvestrant vs anastrozole to vary between £33,079 and £35,252 per QALY compared to the base case ICER of £34,099 per QALY. The ICER of fulvestrant vs letrozole ranged between £28,488 and £32,048 per QALY, and that of fulvestrant vs tamoxifen was between £22,233 and £24,442 per QALY respectively.

In scenario 3, using the 'no shape arm' model with the generalised-gamma distribution for PFS extrapolation provided implausible results as in this scenario; all the patients started the model in the PD health state and PFS was equal to zero. This distribution was excluded in the 'no shape arm' model because of the *"complexity in the interpretation of setting two of the three-parameter generalised gamma model equal".* (CS Section 4.10.1, Page 96; and clarification response to question B9). Assigning other distributions to the PFS (Weibull, Gompertz, log-logistic and lognormal) resulted in the ICER of fulvestrant vs anastrozole varying between £33,687 and £39,664 per QALY.

When anastrozole and letrozole were assumed to have equal efficacy (scenario 4), the ICER of fulvestrant vs letrozole was similar to fulvestrant vs. anastrozole and ranged between £33,123 and £35,284 per QALY for different distributions for PFS.

In scenario 5, the CS explored the impact of using three sets of utility values on the base case results. Of these sets, the values obtained from the combination of the FALCON study MMRM model (1) and Lloyd 2006 had the most influence on the base case results, particularly on the ICER of fulvestrant vs letrozole which increased by £5,220 from the base case value. The ICER for fulvestrant vs anastrozole increased slightly by £417, whilst the ICER for fulvestrant vs tamoxifen decreased by £1,108 compared to the base case results.

Using a lower time-horizon increased the ICERs relative to the base case values and vice-versa (scenario 6); using lower discount rates (scenario 7) and excluding AE costs and dis-utilities (scenario 8) lowered the ICERs of fulvestrant vs the comparators compared to the base case results.

Including zero administration costs (scenario 9), the same subsequent treatment costs for all comparators and excluding subsequent costs (scenario 10) increased the ICERs of fulvestrant vs comparators marginally, compared to the base case results.

In summary, the results from the above analyses indicate that alternative scenarios provided broadly similar results to the base case.

Probabilistic sensitivity analysis

The company conducted a probabilistic sensitivity analysis (PSA) on their base case analysis to assess parametric uncertainty (CS section 5.8.1). The PSA was well-conducted and accounted for uncertainty around most of the input parameters. The parameters, together with the chosen distribution alongside their rationale, are reproduced from CS Table 87 in Table 47 below.

Table 47 List of parameters and associated distributions included in the PSA

Parameter	Distribution	Comment
Survival distributions	Cholesky decomposition	Decomposition of a Hermitian, positive-definite matrix into the product of a lower triangular matrix and its conjugate transpose
Survival curve (shape, scale, and covariate parameters)	Multinomial normal	Incorporates the covariance between parameters estimated in a survival regression analysis
Costs	Gamma	Likely skewed nature of health care costs, and their constraint to positive values
AE rates (incidence)	Beta	Bounded between 0 and 1
Distribution of subsequent treatments	Dirichlet distribution	Normalised sum of independent gamma variables
Duration of subsequent treatment	Gamma	Bounded between 0 and infinity, and skewed
Utilities	Beta	Constrained to values between minus infinity and 1. Modelled as a disutility
AE disutilities	Lognormal	Bounded between 0 and infinity, and skewed

Source: CS Table 87; AE: Adverse Event

The CS presented the results of the PSA for 10,000 simulations; the ERG ran these simulations in the corrected model which took approximately 40 minutes to run. We considered the distributions assigned to the parameters along with the justifications provided to be appropriate. Patient age, discount rates, model time horizon and acquisition costs of fulvestrant and the comparator drugs were not varied in these analyses.

The results of the PSA were tabulated in CS Table 88 – 91 and diagrammatically presented as scatter-plots (CS Figure 46 - 48) and cost-effectiveness acceptability curves (CEACs) (CS Figure 49 - 50). The point estimates from the average PSA results from the corrected model were close to the results obtained from the deterministic analysis as summarized in Table 48.

Table 48 Comparison of the point estimates obtained from the deterministic and PSA analyses (based on corrected model)

Intervention vs comparator	Deterministic ICER (£/QALY)	Probabilistic ICER (£/QALY)
Fulvestrant vs Anastrozole	£34,099	£33,762
Fulvestrant vs Letrozole	£29,991	£31,264
Fulvestrant vs Tamoxifen	£22,498	£22,815

The probability of the treatments being cost-effective at different willingness-to-pay thresholds (WTP) are tabulated in Table 49 and the CEACs are reproduced from the company's model in Figure 22. At a WTP threshold of £30,000 per QALY, the probability of fulvestrant being cost-effective is 26.5%; whereas the probabilities are 37.4% for anastrozole; and 33.3% and 2.9% for letrozole and tamoxifen, respectively.

Table 49 Probability of the treatments being cost-effective at different WTP thresholds (based on corrected model)

WTP threshold (per QALY)	Probability of being cost-effective (%)			
	Fulvestrant	Anastrozole	Letrozole	Tamoxifen
£20,000	1.1%	46.4%	51.5%	1.0%
£30,000	**26.5%**	**37.4%**	**33.3%**	**2.9%**
£50,000	67.8%	14.5%	14.0%	3.8%

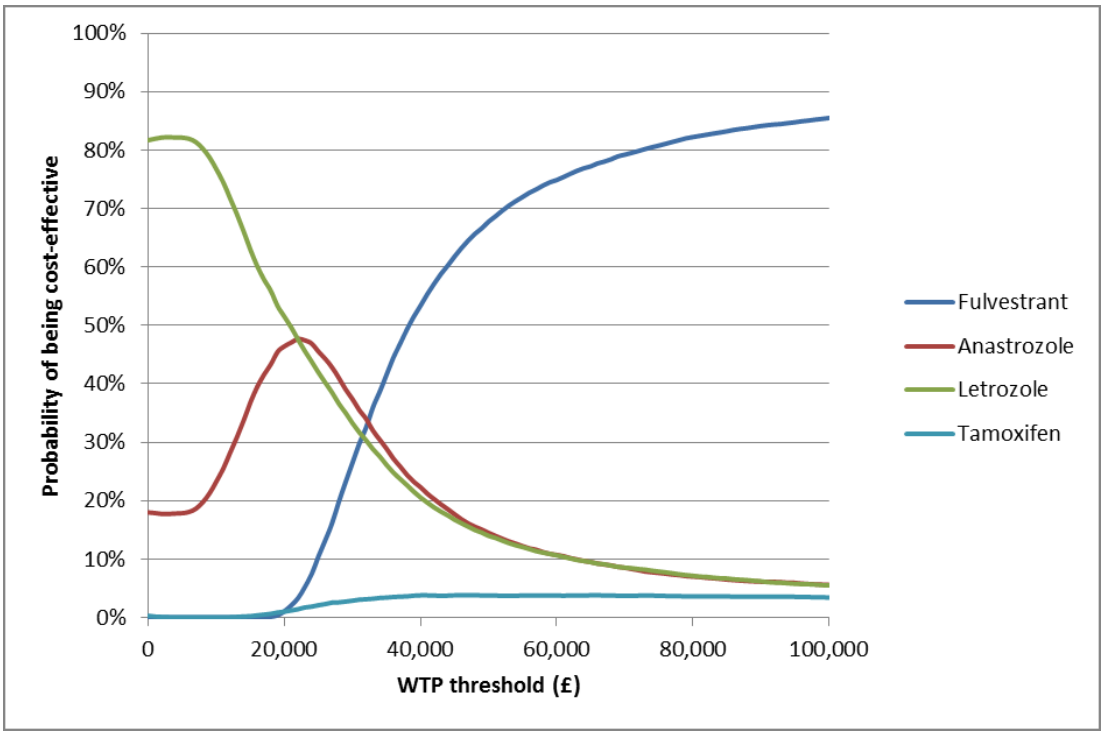


Figure 22 Cost-effectiveness acceptability curves for all the treatments (based on corrected model)

4.4 Additional work undertaken by the ERG

There were a few areas where the ERG considered the CS base case to be limited. In this section, we detail the ERG's further exploration of these issues and uncertainties which have been highlighted in the review and critique of the CS base case analyses, in the earlier sections of this report. A summary of the ERG's exploratory analyses are presented in Table 50, along with their justifications, and how these analyses changed the parameters from the CS base case. We then combine some of these analyses to form the ERG base case, which we regard as the most representative analysis for the cost-effectiveness of fulvestrant compared to anastrozole, letrozole and tamoxifen for treating advanced or metastatic breast cancer.

Table 50 Summary of the ERG's exploratory analyses

ERG scenario	Analysis description in the CS base case	ERG's analysis	Justification
1.	Clinical efficacy: OS extrapolation using Weibull distribution	OS extrapolation using Gompertz, log-logistic and log-normal distribution	The CS explored all the distributions for PFS but not for OS extrapolation. Hence, the ERG extrapolated OS with the remaining distributions for completeness.
2.	Deterministic sensitivity analyses for OS shape parameter	More detailed analysis of variation of ICER with changes in OS scale parameter	There remains uncertainty around the OS parameters, given the immature FALCON OS data
3.	Resource use associated with PFS and PD health states per cycle within disease management costs were derived from NICE clinical guidance -81 and PSSRU	Using resource use of PFS and PD from the study by Karnon et al..[52]	The resources described in the NICE CG refer to patients receiving chemotherapy, not patients receiving endocrine therapy.
4.	The proportions of patients receiving subsequent 2nd treatment are: Endocrine therapy: 54.35%; Chemotherapy: 37.57% ; Targeted treatment: 8.08%	The proportions of patients receiving 2nd line treatment are: Endocrine therapy: 67%; Chemotherapy: 25.92%; Targeted therapy: 8.08%	Based on ERG clinical expert opinion and ERG assumption.
5.	Inclusion of PO25 trial from the NMA to obtain PFS and OS estimates	Exclusion of PO25 trial from the NMA to obtain PFS and OS estimates	PO25 trial population differs from the other trial in the NMA and letrozole is widely accepted to be of equal efficacy as that of anastrozole.
6	Administration of fulvestrant in the outpatient setting for 67% of patients and in the primary care setting for 33%	Fulvestrant administered to all patients in the outpatient setting	Based on clinical expert opinion
7	ERG base case	Combining ERG scenarios 3, 4, 5 and 6	As stated above

Further discussion and results of all the above exploratory analyses are presented in the following sub-sections.

4.4.1 ERG Scenario 1: Extrapolation of OS curve: assigning different distributions for the 'all shapes model' (based on the corrected model)

The results obtained from assigning different distributions to extrapolate the OS curve are presented in Table 51.

Table 51 ERG scenario 1: OS extrapolation using different distributions for the 'all shapes model'

Parameters		Base case ICER (OS: Weibull)	Scenario ICER
Scenario 1: OS Gompertz; PFS: generalised gamma			
Fulvestrant vs	Letrozole	£29,991	fulvestrant dominated
	Anastrozole	£34,099	£59,953
	Tamoxifen	£22,498	£75,229
Scenario 2: OS log-logistic; PFS: generalised gamma			
Fulvestrant vs	Letrozole	£29,991	£29,628
	Anastrozole	£34,099	£35,128
	Tamoxifen	£22,498	£22,677
Scenario 2: OS log-normal; PFS: generalised gamma			
Fulvestrant vs	Letrozole	£29,991	£33,834
	Anastrozole	£34,099	£34,896
	Tamoxifen	£22,498	£22,976

Using the Gompertz distribution to extrapolate the OS curve changes the direction of the base case results. Fulvestrant is dominated when compared with letrozole, as letrozole is less expensive and more effective with higher QALYs, thereby resulting in a negative ICER in the south-west quadrant of the cost-effectiveness plane. Further, the ICERs increase significantly when fulvestrant is compared against anastrozole and tamoxifen. However, the ERG notes that the Gompertz distribution provides a poor fit to the observed data so results from this distribution should be treated with caution. Extrapolating the OS curve by assigning log-logistic and log-normal distributions has minimal impact on the ICERs for fulvestrant vs comparators, compared to the base case ICERs.

4.4.2 ERG Scenario 2: Changes to the OS scale parameter for fulvestrant

The company model results were most sensitive to changes in treatment effectiveness by varying the OS scale parameter (see section 4.3.10). As shown in Table 42, varying the OS scale parameter between the lower and upper 95% confidence intervals resulted in the ICER for fulvestrant vs. anastrozole varying between £23,236 and £338,729 per QALY (Incremental scale parameter for fulvestrant varied between [redacted] and [redacted]).

The ERG considers that there remains uncertainty around the OS scale parameters due to the immature OS data from the FALCON trial. The long-term OS data from this trial was not available to be included in the NMA. The ERG considers that the survival benefit for fulvestrant compared to anastrozole is likely to be lower from the FALCON trial than observed in the FIRST trial. Therefore, when data from the FALCON trial becomes available, the treatment benefit for fulvestrant may be lower than estimated in the NMA. The ERG varies the OS scale parameter in scenario 2 between its mean value and the upper 95% confidence interval to illustrate the effect of changes to the treatment benefit. The results are shown for four scale parameters (incremental values from [redacted]) in Table 52.

Table 52 ERG scenario 2: Effect of changes of the fulvestrant OS scale parameter

Parameters		Base case ICER (OS: Weibull)	Scenario ICER
Scenario 2: Fulvestrant Incremental scale parameter [redacted]			
Fulvestrant vs	Letrozole	£29,991	£33,475
	Anastrozole	£34,099	£40,761
	Tamoxifen	£22,498	£24,432
Scenario 2: Fulvestrant Incremental scale parameter [redacted]			
Fulvestrant vs	Letrozole	£29,991	£38,326
	Anastrozole	£34,099	£52,405
	Tamoxifen	£22,498	£27,146
Scenario 2: Fulvestrant Incremental scale parameter [redacted]			
Fulvestrant vs	Letrozole	£29,991	£45,842
	Anastrozole	£34,099	£79,337
	Tamoxifen	£22,498	£31,404
Scenario 2: Fulvestrant Incremental scale parameter [redacted]			
Fulvestrant vs	Letrozole	£29,991	£59,000
	Anastrozole	£34,099	£208,231
	Tamoxifen	£22,498	£39,027

These illustrative results indicate that even with a relatively small change to the OS scale parameter to [redacted], produces an ICER of £40,761 per QALY for fulvestrant vs. anastrozole.

4.4.3 ERG Scenario 3: Change in resource use for disease management costs (based on the corrected model)

To address the ERG's concerns in relation to the estimation of disease management costs for PFS and PD states (as outlined in section 4.3.7), the ERG calculated the base case results by estimating resource use for these health states from the study by Karnon et al.[52] This study conducted a trial-based cost-effectiveness analysis of letrozole (first-line) with the option of second-line tamoxifen vs tamoxifen (first-line) with the option of second-line letrozole in postmenopausal advanced breast cancer patients. The proportion of patients receiving interventions in each health state was estimated based on a three month period. For the purpose of this appraisal, we converted these proportions for a four week period as shown in Table 53 and updated the unit costs for these resources.

The revised estimated cost for both PFS and PD health states is £90.91 per cycle. The incremental results from this scenario analysis are presented in Table 54.

Table 53 Resource use and unit costs for PFS and PD health states based on Karnon et al.

	Proportion of patients per 4 weeks	Unit cost (£)	Total cost (£)	Source
Outpatient visits				
Oncologist	0.29	162.84	47.22	NHS reference costs 2015-16 (Non-admitted face to face attendance Follow-up Medical oncology code 370)[49]
GP	0.18	46.02	8.28	PSSRU 2015/16[50]
Radiographer	0.08	46	3.68	PSSRU 2015/16[50]
Lab tests				
Biochemical	0.28	1.18	0.33	NHS reference costs 2015-16 DAPS04[49]
Blood tests	0.27	3.1	0.84	NHS reference costs 2015-16 DAPS05[49]
Bone scintography	0.18	75.89	13.66	Cost updated from Karnon et al.[52] using PSSRU HCHS indices[50]
Ultrasound	0.06	53.45	3.21	NHS reference costs 2015-16 (Imaging codes Ultrasound scan RD40Z-FD43Z)[49]
Chest x-ray	0.14	15.10	2.11	Cost updated from Karnon et al.[52] using PSSRU HCHS indices[50]

Bone x-ray	0.08	24.58	1.97	Cost updated from Karnon et al.[52] using PSSRU HCHS indices[50]
Hospitalisation				
General medicine	0.01	246.58	2.47	Cost updated from Karnon et al.[52] using PSSRU HCHS indices[50]
Oncology	0.01	713.81	7.14	NHS reference costs 2015-16 (Non elective short stay codes Malignancy of bone or connective tissue HD40D – HD40H) [49]

Table 54 Incremental results of the ERG's scenario 3 for revised health state costs

Treatments	Total costs	Total QALYs	Incremental costs	Incremental QALYs	Incremental ICER (cost per QALY)
Letrozole	£11,098	2.46	--	--	--
Anastrozole	£11,388	2.68	£290	0.22	£1,314
Tamoxifen	£11,895	2.47	£507	-0.21	Dominated*
Fulvestrant	£29,133	3.23	£17,745	0.55	£32,084

For the incremental analyses, the treatment strategies were placed in order of increasing total costs. Letrozole was used as the baseline comparator as it is associated with the lowest total costs. The decrease in total costs for all the treatments compared to that of the base case analyses occurred due to the lower disease management costs associated with the PFS and PD states. The ICER of fulvestrant vs anastrozole was £32,084 per QALY compared to the base case ICER of £34,099 per QALY. Tamoxifen was dominated when compared with anastrozole as it was associated with higher incremental costs and lower incremental QALYs.

4.4.4 ERG Scenario 4: Change in the proportion of patients receiving subsequent treatments for second-line

In this scenario analysis the ERG changed the proportion of patients receiving second line treatments to address the views of our clinical experts, as previously outlined in section 4.3.7. Our clinical experts suggested that 67-80% of the patients would receive endocrine therapy as second-line treatment. One of our experts considered that 20% of the patients would have targeted therapy in combination with endocrine therapy or chemotherapy, whilst the other considered that fewer would have chemotherapy than estimated by the CS. Owing to limited information on the proportion of patients receiving these combination therapies, the ERG pragmatically assumed the proportions of patients for the subsequent therapies as shown in Table 55. The results of this analysis are presented in Table 56.

Table 55 ERG's assumptions related to the proportion of patients receiving second-line treatments

Proportion of patients (%)	Endocrine therapy	Chemotherapy	Targeted therapy
Baseline	54.35%	37.57%	8.08%
Scenario	67.00%	24.92%	8.08%

Table 56 Incremental results from ERG scenario 4 for changing proportion of patients receiving second-line endocrine therapy

Treatments	Total costs	Total QALYs	Incremental costs	Incremental QALYs	Incremental ICER (cost per QALY)
Letrozole	£26,188	2.46	--	--	--
Anastrozole	£30,539	2.68	£4,351	0.22	£19,702
Tamoxifen	£32,286	2.47	£1,737	-0.21	Dominated*
Fulvestrant	£49,369	3.23	£18,830	0.55	£34,046

*Tamoxifen is more expensive and less effective compared to anastrozole.

The change in the proportion of patients receiving second-line treatments results in an ICER for fulvestrant vs anastrozole of £34,046 per QALY, a decrement of £53 compared to the company's base case ICER of £34,099 per QALY. This indicates that varying the proportion of patients receiving subsequent therapies as second-line does not influence the base case results.

4.4.5 ERG Scenario 5: Excluding PO25 trial from the NMA network to obtain PFS and OS estimates for 'all shapes model'

This scenario uses the fixed-effects NMA results without the PO25 trial for both PFS and OS for the 'all shapes model' and assumes that anastrozole and letrozole have similar efficacy. The incremental results are presented in Table 57. The parameters used for the parametric distributions for this scenario are described in the company's clarification response (Clarification Question A13).

Table 57 Incremental results from ERG Scenario 5 obtained from excluding PO25 trial from the fixed-effects NMA for both PFS and OS for the 'all shape model'

Treatments	Total costs	Total QALYs	Incremental costs	Incremental QALYs	Incremental ICER (cost per QALY)
Letrozole	£30,541	2.68	--	--	--
Anastrozole	£30,561	2.68	--	--	--
Tamoxifen	£32,323	2.47	£1,762	-0.21	Dominated
Fulvestrant	£49,435	3.23	£18,874	0.55	£34,113

As letrozole and anastrozole are assumed to be of equal efficacy, incremental results are obtained by using anastrozole as the base-case.

When compared with anastrozole, tamoxifen is associated with an additional cost of £1,762 but lower QALYs of -0.21, thereby making tamoxifen a dominated strategy. The incremental ICER of fulvestrant vs anastrozole is £34,113, thereby indicating that exclusion of PO25 trial had almost no impact on the base case ICER of £34,099 per QALY.

4.4.6 ERG Scenario 6: Change in administration cost for fulvestrant

One of our clinical experts stated that all patients in his locality would be treated in the outpatients setting and none in the primary care setting. Furthermore, this expert considered that it would be difficult to persuade primary care to administer fulvestrant. We therefore included a scenario where all patients received fulvestrant in the outpatient setting.
The administration cost for fulvestrant for treatments in the first four weeks are £399.58 and in subsequent months are £99.97 using this assumption. The incremental results are presented in Table 58.

Table 58 Incremental results from ERG Scenario 6 with a change in the administration cost for fulvestrant

Treatments	Total costs	Total QALYs	Incremental costs	Incremental QALYs	Incremental ICER (cost per QALY)
Letrozole	£26,221	2.46	--	--	--
Anastrozole	£30,572	2.68	£4,351	0.22	£19,702
Tamoxifen	£32,328	2.47	£1,756	-0.21	Dominated
Fulvestrant	£50,203	3.23	£19,632	0.54	£35,496

When compared with anastrozole, tamoxifen is a dominated strategy. The incremental ICER of fulvestrant vs anastrozole is £35,496 per QALY, i.e. an increase of £1,397 on the base case ICER of £34,099 per QALY.

4.4.7 ERG base case

The assumptions for the ERG base case are listed below and results of this analysis are presented in Table 59. We consider this scenario to be most representative analysis of the available evidence for the cost-effectiveness of fulvestrant.

- Resource use for PFS and PS health states are based on the study by Karnon et al[52] as shown in ERG scenario 3
- Revised proportion of patients receiving second-line treatment, shown in ERG scenario 4
- Exclusion of PO25 trial from the NMA network and assuming similar efficacy for letrozole and anastrozole, shown in ERG scenario 5.
- All patients receiving fulvestrant administered in an outpatient setting, shown in ERG scenario 6.

Table 59 Incremental results of the ERG base case

Treatments	Total costs	Total QALYs	Incremental costs	Incremental QALYs	Incremental ICER (cost per QALY)
Letrozole	£11,336	2.68	--	--	--
Anastrozole	£11,356	2.68	--	--	--
Tamoxifen	£11,852	2.47	£496	-0.21	Dominated
Fulvestrant	£29,866	3.23	£18,510	0.54	£33,455

The ERG base case incremental ICER for fulvestrant vs anstrozole is £33,455 per QALY gained Letrozole and anastrozole were assumed to be of equal efficacy, so the incremental analysis was estimated using anastrozole as the baseline comparator. Tamoxifen is dominated when compared with anastrozole as it is more expensive and less effective.

The ERG also conducted a PSA for 10,000 simulations of our base case. A comparison of the results obtained from the deterministic base case and the point estimates from the average PSA are presented in Table 60.

Table 60 Comparison of the point estimates obtained from the deterministic and PSA analyses of the ERG base case

Intervention vs comparator	Deterministic ICER (£/QALY)	Probabilistic ICER (£/QALY)
Fulvestrant vs Anastrozole	£33,455	£32,956
Fulvestrant vs Letrozole	£33,495	£32,983
Fulvestrant vs Tamoxifen	£23,687	£23,999

The ERG set anastrozole to have equal efficacy to that of letrozole, however we note that this produces PSA results for anastrozole and letrozole that have the same QALYs for each simulation, rather than simulating anastrozole and letrozole independently. We were unclear how to change the PSA calculations in the model to simulate anastrozole and letrozole independently as these calculations are not intuitive and have not been clearly explained.

4.5 Conclusions of cost effectiveness

The company used a model structure commonly used for economic models of cancer treatments with health states for progression-free survival, progression and death. The ERG considers the model structure to appropriate for the decision problem and the clinical pathway of advanced or metastatic breast cancer. The company used methods for the economic evaluation that are consistent with NICE methodological guidelines. The population, intervention and comparators used in the economic evaluation are consistent with the NICE scope.[1]

The company compares fulvestrant with anastrozole, letrozole and tamoxifen using an NMA that produces output in the form of parametric distributions of survival curves for PFS and OS. These curves are used directly in the economic model. The ERG considers that the distributions chosen by the company for PFS and OS are appropriate and provide a reasonable fit to the observed data. The ERG notes that using alternative parametric distributions for PFS and OS do not have significant impact on the model results, with the exception of the Gompertz distribution for OS (which the ERG considers to provide a poor fit to the observed data).

The ERG notes that the OS data from the FALCON trial are immature. Therefore OS for fulvestrant vs, anastrozole is largely based upon the FIRST trial. The ERG notes that the gain in PFS for fulvestrant compared to anastrozole was significantly lower in the FALCON trial than in

the FIRST trial and therefore suggests that it is likely that the OS benefit will also be lower in the FALCON trial than in the FIRST trial. Given the sensitivity of the model results to changes in OS, the ERG therefore considers there is some uncertainty in the cost-effectiveness estimates and the ICERs are likely to be higher when the full results of the FALCON trial become available.

5 End of life

The company do not consider fulvestrant to be an 'End of Life medicine' in this indication (CS p. 128).

6 Innovation

The CS highlights the innovative nature of fulvestrant based on its unique mechanism of action to block oestrogen by targeting and degrading the ER (CS section 2.5 p. 25). The CS states that this unique mechanism of action could potentially delay acquired resistance and increase OS. The ERG notes that evidence regarding resistance is not presented in the CS and whilst a significant improvement in OS was observed in the FIRST trial (where 72% of patients were endocrine therapy-naive [redacted]), median OS has not yet been reached in the FALCON trial.

In comparison to the AIs and tamoxifen which are oral therapies, the IM administration route for fulvestrant may improve compliance. The CS points out that a therapy with an IM route of administration may benefit patients who have difficulty swallowing and those whose compliance with oral therapy may be limited (e.g. the elderly or those with psychiatric illness). The ERG agrees that this would be the case. The ERG sought clinical advice regarding whether the IM administration would be unsuitable for any patients. The advice received was that for very thin women with little muscle in the gluteal area the injections would be very painful.

7 DISCUSSION

7.1 Summary of clinical effectiveness issues

The company identified one phase II RCT (the FIRST trial) and one phase II RCT (the FALCON trial) that are relevant to the decision problem. The two trials provide evidence on a total of 667 postmenopausal patients with hormone-receptor positive advanced breast cancer who were randomised to treatment with either fulvestrant or anastrozole. All participants in the FALCON

trial and 74.6% of participants in the FIRST trial were endocrine therapy naive. No head to head trials were identified comparing fulvestrant to either letrozole or tamoxifen.

The extent to which the benefits of fulvestrant in terms of TTP/PFS and OS exceed those of anastrozole are uncertain. For PFS the uncertainty is because the degree of PFS benefit seen with fulvestrant in the FIRST trial (median TTP 10.3 months longer with fulvestrant) is greater than that observed in the FALCON trial (median PFS 2.8 months longer with fulvestrant). For OS the uncertainty is because OS was added as an outcome after the TTP analysis for the FIRST trial. So although median survival in the fulvestrant arm of the FIRST trial was almost 6 months longer than that of the anastrozole arm, confirmation of this result is required from the FALCON trial, but median OS has not yet been reached in the FALCON trial so these results are not available.

To obtain an estimate of the comparative effectiveness of fulvestrant in comparison to anastrozole, letrozole and tamoxifen an NMA was conducted. The ERG found some evidence of heterogeneity between the trials. However, for the trials where there was IPD (FALCON, FIRST, NorthAmTarget) the company matched participants to inclusion criteria of the FALCON trial such that only ER+/PgR+ patients plus endocrine treatment naive patients would be included in the NMA. This created a more homogeneous population for the NMA (except for study P025 which could not be matched because the company did not have access to IPD for this study). The company undertook a fixed-effect NMA because the number of studies in final network was small (five studies, two of which were designed to allow for combined data analysis) and the methodological difficulties of conducting a random-effects NMA but this may mean that the results do not fully capture uncertainty.

7.2 Summary of cost effectiveness issues

The CS includes evidence on the cost effectiveness of fulvestrant compared to anastrozole, letrozole and tamoxifen in post-menopausal women with untreated hormone-receptor positive locally advanced or metastatic breast cancer. The model structure adopted for the economic evaluation is appropriate and consistent with the clinical disease pathway. The model contains health states of progression-free, progressed disease and death. Parametric survival curves are used for PFS and OS based upon the clinical evidence. The clinical evidence consists of an NMA of trials. The ERG considered that the parametric distributions chosen by the company to

model PFS and OS were appropriate and a reasonable fit to the observed data. However, the OS data for the FALCON trial are immature and so long-term OS data were not available to be included within the NMA.

The ERG considers that it is more appropriate to exclude the PO25 trial that compares letrozole and tamoxifen. When this trial is excluded, there is no clinical evidence to include in the NMA to compare letrozole with the other treatments. Based on clinical advice, we assume that the efficacy of letrozole is equal to that of anastrozole.

The CS models produce an ICER of £34,099 per QALY compared to anastrozole. The model results were particularly sensitive to changes in the OS parameter values. The company's probabilistic sensitivity analyses showed there is a probability of 1.1% and 26.5% of fulvestrant being cost-effective at a willingness to pay threshold of £20,000 and £30,000 per QALY respectively.

The ERG's base case analysis includes changes to the health state resources, the proportion of patients receiving second-line endocrine therapy, excluding the PO25 trial from the NMA and for all patients to receive fulvestrant administered in an outpatient care setting. The ERG's base case analysis produces an ICER of £33,455 per QALY compared to anastrozole.

8 REFERENCES

1. National Institute for Health and Care Excellence. *Fulvestrant for untreated hormone-receptor positive locally advanced or metastatic breast cancer. Final scope.*, 2017.
2. National Collaborating Centre for Cancer. Advanced breast cancer: diagnosis and treatment [CG81]. 2009.
3. National Institute for Health and Care Excellence. Everolimus with exemestane for treating advanced breast cancer after endocrine therapy. Technology appraisal guidance [TA421], 2016.
4. European Medicines Agency. Summary of Product Characteristics. Faslodex, 2009.
5. Robertson JF, Lindemann JP, Llombart-Cussac A, Rolski J, Feltl D, Dewar J, et al. Fulvestrant 500 mg versus anastrozole 1 mg for the first-line treatment of advanced breast cancer: follow-up analysis from the randomized 'FIRST' study. *Breast cancer research and treatment* 2012;136(2):503-11.

6. Robertson JF, Llombart-Cussac A, Rolski J, Feltl D, Dewar J, Macpherson E, et al. Activity of fulvestrant 500 mg versus anastrozole 1 mg as first-line treatment for advanced breast cancer: results from the FIRST study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2009;27(27):4530-5.
7. Ellis MJ, Llombart-Cussac A, Feltl D, Dewar JA, Jasiowka M, Hewson N, et al. Fulvestrant 500 mg Versus Anastrozole 1 mg for the First-Line Treatment of Advanced Breast Cancer: Overall Survival Analysis From the Phase II FIRST Study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2015;33(32):3781-7.
8. Robertson JF, Bondarenko IM, Trishkina E, Dvorkin M, Panasci L, Manikhas A, et al. Fulvestrant 500 mg versus anastrozole 1 mg for hormone receptor-positive advanced breast cancer (FALCON): an international, randomised, double-blind, phase 3 trial. *Lancet (London, England)* 2016;388(10063):2997-3005.
9. Nabholtz JM, Buzdar A, Pollak M, Harwin W, Burton G, Mangalik A, et al. Anastrozole is superior to tamoxifen as first-line therapy for advanced breast cancer in postmenopausal women: results of a North American multicenter randomized trial. Arimidex Study Group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2000;18(22):3758-67.
10. Bonneterre J, Thurlimann B, Robertson JF, Krzakowski M, Mauriac L, Koralewski P, et al. Anastrozole versus tamoxifen as first-line therapy for advanced breast cancer in 668 postmenopausal women: results of the Tamoxifen or Arimidex Randomized Group Efficacy and Tolerability study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2000;18(22):3748-57.
11. Milla-Santos A, Milla L, Portella J, Rallo L, Pons M, Rodes E, et al. Anastrozole Versus Tamoxifen as First-Line Therapy in Postmenopausal Patients With Hormone-Dependent Advanced Breast Cancer: A Prospective, Randomized, Phase III Study. *American Journal of Clinical Oncology* 2003;26(3):317-22.
12. Mouridsen H, Gershanovich M, Sun Y, Perez-Carrion R, Boni C, Monnier A, et al. Superior efficacy of letrozole versus tamoxifen as first-line therapy for postmenopausal women with advanced breast cancer: results of a phase III study of the International Letrozole Breast Cancer Group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2001;19(10):2596-606.
13. Mouridsen H, Gershanovich M, Sun Y, Perez-Carrion R, Boni C, Monnier A, et al. Phase III study of letrozole versus tamoxifen as first-line therapy of advanced breast cancer in

postmenopausal women: analysis of survival and update of efficacy from the International Letrozole Breast Cancer Group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2003;21(11):2101-9.

14. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]: The Cochrane Collaboration, 2011.
15. Centre for Reviews and Dissemination. Systematic reviews: CRD's guidance for undertaking reviews in health care. York Publishing Services Ltd.: CRD, 2009.
16. AstraZeneca. A Randomised, Double-blind, Parallel-group, Multicentre, Phase III Study to Compare the Efficacy and Tolerability of Fulvestrant (FASLODEX™)500 mg with Anastrozole (ARIMIDEX™) 1 mg as Hormonal Treatment for Postmenopausal Women with Hormone Receptor-Positive Locally Advanced or Metastatic Breast Cancer Who Have Not Previously Been Treated With Any Hormonal Therapy (FALCON), 2013.
17. U.S. Department of Health and Human Services NIoH, National Cancer Institude. Common Terminology Criteria for Adverse Events (CTCAE) Version 4.0. 2009. (accessed 06/06/2017).
18. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. *NICE, London* 2013.
19. Garcia SF, Hahn EA. Measure validation is an ongoing process: the Functional Assessment of Cancer Therapy-Breast Symptom Index as a case example. *Annals of Palliative Medicine* 2012;1(3):207-10.
20. Saad ED, Katz A. Progression-free survival and time to progression as primary end points in advanced breast cancer: often used, sometimes loosely defined. *Annals of oncology : official journal of the European Society for Medical Oncology* 2009;20(3):460-4.
21. National Institute for Health and Care Excellence. TA239 Fulvestrant for the treatment of locally advanced or metastatic cancer 2011.
22. Guyot P, Ades AE, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. *BMC medical research methodology* 2012;12:9.
23. Ouwens MJ, Philips Z, Jansen JP. Network meta-analysis of parametric survival curves. *Research synthesis methods* 2010;1(3-4):258-71.
24. Friede T, Rover C, Wandel S, Neuenschwander B. Meta-analysis of two studies in the presence of heterogeneity with applications in rare diseases. *Biometrical journal. Biometrische Zeitschrift* 2016.

25. Ellis S, Carroll KJ, Pemberton K. Analysis of duration of response in oncology trials. *Contemporary clinical trials* 2008;29(4):456-65.
26. AstraZeneca. [data on file]. A Randomised, double-blind, parallel-group, multicentre, phase III study to compare the efficacy and tolerability of fulvestrant (FASLODEX) 500 mg with anastrozole (ARIMADEX) 1 mg as hormonal treatment for postmenopausal women with hormone receptor-positive locally advanced or metastatic breast cancer who have not previously been treated with any hormonal therapy (FALCON). Clinical Study Report. , 2016.
27. Newman MJ, Jones LT, Kraft JM, Lee AR, Lech GM, Farrell NM, et al. Cost-effectiveness of fulvestrant 250 mg versus 500 mg in postmenopausal women with estrogen receptor-positive metastatic breast cancer and disease progression after antiestrogen therapy. *J Oncol Pharm Pract* 2012;18(4):394-401.
28. Das R, Cope S, Ouwens M, Turner P, Howlett M. Economic evaluation of fulvestrant 500 mg versus generic nonsteroidal aromatase inhibitors in patients with advanced breast cancer in the United Kingdom. *Clin Ther* 2013;35(3):246-60 e5.
29. AstraZeneca. Data on File: Minutes from the Breast cancer HTA advisory board (23 Feb 2017). , 2017.
30. Finn RS, Martin M, Rugo HS, Jones S, Im S-A, Gelmon K, et al. Palbociclib and letrozole in advanced breast cancer. *New England Journal of Medicine* 2016;375(20):1925-36.
31. Paridaens RJ, Dirix LY, Beex LV, Nooij M, Cameron DA, Cufer T, et al. Phase III study comparing exemestane with tamoxifen as first-line hormonal treatment of metastatic breast cancer in postmenopausal women: the European Organisation for Research and Treatment of Cancer Breast Cancer Cooperative Group. *Journal of clinical oncology* 2008;26(30):4883-90.
32. National Institute for Health and Care Excellence. TA371 Trastuzumab emtansine for treating HER2-positive, unresectable locally advanced or metastatic breast cancer after treatment with trastuzumab and a taxane 2015.
33. National Institute for Health and Care Excellence. TA424 Pertuzumab for neoadjuvant treatment of HER2-positive breast cancer. 2016.
34. National Institute for Health and Care Excellence. TA421 Everolimus with exemestane for advanced breast cancer after endocrine therapy. 2016.
35. National Institute for Health and Care Excellence. TA423 Erbulin for locally advanced or metastatic breast cancer after 2 or more chemotherapy regimens. 2016.

36. National Institute for Health and Care Excellence. Palbociclib in combination with an aromatase inhibitor 2017.
37. Fukuda T, Shirowas T, Shimozuma K, Mouri M, Dolhara H, Akabane H. Long-term EQ-5d score for patients with metastatic breast cancer; comparison of first-line oral S-1 and taxane therapies in the randomized 'Select' trial: Value in Health; 2015.
38. Eyles C, Leydon GM, Hoffman CJ, Copson ER, Prescott P, Chorozoglou M, et al. Mindfulness for the self-management of fatigue, anxiety, and depression in women with metastatic breast cancer: a mixed methods feasibility study. *Integr Cancer Ther* 2015;14(1):42-56.
39. Takashima T, Mukai H, Hara F, Matsubara N, Saito T, Takano T, et al. Taxanes versus S-1 as the first-line chemotherapy for metastatic breast cancer (SELECT BC): an open-label, non-inferiority, randomised phase 3 trial. *Lancet Oncol* 2016;17(1):90-8.
40. Mitra D, Wood R, De Courcy J, Iyer S. Patient reported health utility in HR+/HER2-advanced/metastatic breast cancer. *Value in Health* 2016;19 (7):A749.
41. Loibl S, Demichele A, Turner NM, Cristofanilli M, Loi S, Verma S, et al. Impact of palbociclib plus fulvestrant on patient reported general health status compared with fulvestrant alone in HR + , HER2- metastatic breast cancer. *Annals of Oncology. Conference: 41st European Society for Medical Oncology Congress, ESMO* 2016;27(no pagination).
42. Lambert-Obry V, Gouault Laliberte A, Castonguay A, Zanotti G, Banerjee R, Mates M, et al. A real-world outcome research study on quality of life, work productivity and healthcare resource utilization in metastatic breast cancer for patients and caregivers. *Value in Health* 2016;19 (7):A616.
43. Shiroiwa T, Fukuda T, Shimozuma K, Mouri M, Hagiwara Y, Doihara H, et al. Long-term health status as measured by EQ-5D among patients with metastatic breast cancer: comparison of first-line oral S-1 and taxane therapies in the randomized phase III SELECT BC trial. *Qual Life Res* 2017;26(2):445-53.
44. Lloyd A, Nafees B, Narewska J, Dewilde S, Watkins J. Health state utilities for metastatic breast cancer. *British journal of cancer* 2006;95(6):683-90.
45. Kind P, Hardman G, S. M. Population norms for EQ-5D. 1999.
46. Boehringer Ingelheim Ltd. STA Nintedanib in combination with docetaxel for the treatment of adult patients with locally advanced, metastatic or recurrent non-small lung cancer (NSCLC) of adenocarcinoma tumour histology after first-line chemotherapy, 2014.
47. Swinburn P, Lloyd A, Nathan P, Choueiri TK, Cella D, Neary MP. Elicitation of health state utilities in metastatic renal cell carcinoma. *Curr Med Res Opin* 2010;26(5):1091-6.

48. Doyle S, Lloyd A, Walker M. Health state utility scores in advanced non-small cell lung cancer. *Lung Cancer* 2008;62(3):374-80.
49. Department of Health. NHS reference costs 2015-2016. *http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_122803* 2016.
50. Curtis L. Unit Costs of Health and Social Care 2016. *http://www.pssru.ac.uk/pdf/uc/uc2010/uc2010.pdf* 2016.
51. Joint Formulary Committee. *British National Formulary* London: British Medical Association and Royal Pharmaceutical Society of Great Britain; 2017.
52. Karnon J. A trial-based cost-effectiveness analysis of letrozole followed by tamoxifen versus tamoxifen followed by letrozole for postmenopausal advanced breast cancer. *Annals of Oncology* 2003;14(11):1629-33.
53. Kurosky S, Mitra D, Zanotti G, Kaye J. Patient characteristics and treatment patterns in ER+/HER2- metastatic breast cancer in the United Kingdom: Results from a retrospective medical record review *ISPOR 18th Annual European Congress*; 2015.
54. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB. Model Transparency and Validation. *Medical Decision Making* 2012;32(5):733-43.

9 APPENDICES

Appendix 1: List of verification checks conducted by the ERG

Checks conducted	Model outcome
Does the model provide a brief background on the model structure and design?	Yes
Are the different components of the model well presented?	Yes
Is it possible to navigate through the model easily?	Yes
Are the inputs used in the model clearly referenced?	Yes
Is the model is transparent with respect to its layout and technicalities?	The model is easy to navigate through. However, use of array functions in calculations have made it quite laborious and time-consuming to tease out the model calculations
Are there any of the key model outputs missing from the analysis?	No
Can the model results be reproduced (including any scenario analyses) as presented in the CS?	Yes- for all the scenario analyses- except for “no arms model” in CS Table 96
Set all the values to "0" and check if the results still pull through some figures	1. Cohort size = 0; no results are pulled through
	2. inputs for "safety", "utility" and "costs"= 0, model pulls through results of only LYs
Does the sum total of the number of patients in each of the health states at any given point (dead or alive) in time (time t+ n) equate to the total number of patients entering the model ?	Yes- except in the last cell of sheet "Pat_flow"
Set the same setting (including the drug) in both the intevention and comparator arm. Are the results for both the arms are similar?	Yes- checked for fulvestrant vs anastrozole
Was an exhaustive list of parameters included within the DSA and PSA?	yes
Are appropriate distributions used for the parameters included in the sensitivity analyses?	yes
Is the deterministic mean ICER approximately equal/close to the probabilistic mean ICER?	yes
Set difference in efficacy for all drugs to 0 ' equal health outcomes in all model arms	Yes
Set adverse event rate to 0%. No adverse events should occur	Yes
Set medical resource use to 0	Yes- get disease management costs as 0
Set unit cost for drugs and administration to 0. Total costs of drugs should be zero.	Yes- model behaves as expected; administration costs as 0
Use different discount rates (e.g. 0%, 3%, 7%)	Used in sensitivity analyses- model behaves as expected
For costs, total costs should decrease with increasing discount rates	
For health benefits, total number of events should decrease with increasing discount rates	
Set utility values to 0, utility adjusted health outcomes should be zero	Yes
Set utility values to 1, utility adjusted health outcomes should be equal to unadjusted life years	Yes- utility adjusted health outcomes = life years