CONFIDENTIAL UNTIL PUBLISHED

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

Abemaciclib with an aromatase inhibitor for untreated advanced hormone receptor-positive, HER2-negative breast cancer

Produced by Southampton Health Technology Assessments Centre (SHTAC)

Authors
Dr Jonathan Shepherd, Principal Research Fellow
Mr Olu Onyimadu, Research Fellow (Health Economics)
Dr Jill Colquitt, Senior Researcher, Effective Evidence LLP
Ms Petra Harris, Research Fellow
Dr Emma Loveman, Senior Researcher, Effective Evidence LLP
Professor Joanne Lord, Professorial Fellow in Health Economics

Correspondence to
Dr Jonathan Shepherd
Southampton Health Technology Assessments Centre (SHTAC)
Wessex Institute
Alpha House
Enterprise Road, University of Southampton Science Park
Southampton SO16 7NS
www.southampton.ac.uk/shtac

Date completed 17th August 2018

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 14/12/01

Declared competing interests of the authors

None

Acknowledgements

We would like to thank Dr Eleni Karapanagiotou, Consultant Medical Oncologist, Guy's and St Thomas' NHS Foundation Trust, London and Dr Peter Simmonds, Consultant Medical Oncologist, Southampton University Hospitals NHS 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 Dr Geoff Frampton, Senior 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:

Shepherd, J., Onyimadu, O., Colquitt, J., Harris, P., Loveman, E., Lord, J. Abemaciclib with an aromatase inhibitor for untreated advanced hormone receptor-positive, HER2-negative breast cancer: A Single Technology Appraisal. Southampton Health Technology Assessments Centre (SHTAC), 2018.

Contributions of authors

Jonathan Shepherd critically appraised the clinical effectiveness review, drafted the report, project managed the report and is the project guarantor. Olu Onyimadu critically appraised the economic evaluation, and drafted the report. Jill Colquitt critically appraised the clinical effectiveness review and drafted the report. Petra Harris critically appraised the clinical effectiveness review and drafted the report. Emma Loveman critically appraised the clinical

effectiveness review and drafted the report. Joanne Lord critically appraised the economic evaluation, and drafted the report.

Word count: 46,125

[redacted]

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABC	Advanced breast cancer
ABE	Abemaciclib
ABE+NSAI	Abemaciclib and non-steroidal aromatase inhibitor
ABE+FUL	Abemaciclib and fulvestrant
AE	Adverse event
AFT	Accelerated Failure Time
AI	Aromatase inhibitor
AIC	Akaike information criterion
ANAS	Anastrozole
BIC	Bayesian inference criteria
BNF	British National Formulary
BOR	Best overall response
BSA	Body surface area
BSC	Best supportive care
CBR	Clinical benefit rate
CDK	Cyclin-dependent kinase
CHMP	Committee for Medicinal Products for Human Use
CrI	Credible interval
CR	Complete Response
CRD	Centre for Reviews and Dissemination
CS	Company submission
CSR	Clinical study report
CTCAE	Common Terminology Criteria for Adverse Events
DFI	Disease free interval
DIC	Deviance information criteria
DoR	Duration of response
DSU	Decision Support Unit
ECOG	Eastern Cooperative Oncology Group
EMA	European Medicines Agency
EORTC QLQ	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire
EPAR	European Public Assessment Report
ER(+)	oestrogen receptor (positive) (abbreviation for US spelling of estrogen)
ERG	Evidence Review Group
EQ-5D	EuroQol 5-Dimension
EXE	Exemestane
EVE+EXE	Everolimus + exemestane
FDA	Food and Drug Administration
FUL	Fulvestrant
HER2(+ or −)	Human epidermal growth factor receptor 2 (positive or negative)
HR	Hazard ratio
HR+	Hormone receptor-positive
HRQoL	Health related quality of life
IC	Interval censoring
ICER	Incremental cost effectiveness ratio
ITC	Indirect treatment comparison
ITT	Intention-to-treat

IV	Intravenous
KM	Kaplan-Meier
LTZ	Letrozole
N/A	Not applicable
NSAI	Non-steroidal aromatase inhibitor
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NMA	Network meta-analysis
NR	Not reported
NSAI	Non-steroidal aromatase inhibitor
OS	Overall survival
OS2	Overall survival from the start of second-line treatment
OR	Odds ratio
ORR	Objective response rate
PAL	Palbociclib
PAL+NSAI	Palbociclib and non-steroidal aromatase inhibitor
PAL+FUL	Palbociclib and fulvestrant
PAS	Patient access scheme
PD	Progressed disease
PF	Progression free
PFD1	Progression-free death from the start of first-line treatment
PFD2	Progression-free death from the start of second-line treatment
PFS	Progression free survival
PFS1	First-line progression-free survival
PFS2	Second-line progression-free survival
PgR	Progesterone receptor
PH	Proportional hazards
PR	Partial response
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRO	Patient reported outcomes
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit
RECIST	RECIST: Response Evaluation Criteria in Solid Tumours
RIBO	Ribociclib
RIBO+NSAI	Ribociclib and non-steroidal aromatase inhibitor
QALY	Quality adjusted life year
QoL	Quality of life
RCT	Randomised controlled trial
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SD	Stable disease
SLR	Systematic literature review
SmPC	Summary of Product Characteristics
STA	Single Technology Appraisal
TEAE	Treatment emergent adverse events
TMX	Tamoxifen
TTD	Time to treatment discontinuation
TTD1	Time to discontinuation of first-line treatment

TTD2	Time to discontinuation of second-line treatment
TTP	Time to progression
TTP1	Time to progression from first-line treatment
TTP2	Time to progression from second-line treatment

SUMMARY

Scope of the company submission

The company submission (CS) assesses the clinical effectiveness and cost effectiveness of abemaciclib (ABE) in combination with a non-steroidal aromatase inhibitor (NSAI) in women with hormone-receptor positive (HR+), human epithelial growth factor receptor 2-negative (HER2-) advanced breast cancer. The comparators are palbociclib with an NSAI and ribociclib with an NSAI.

The decision problem generally meets the NICE scope, however, there are some differences in the population presented. The population in the decision problem is narrower by concentrating on locoregionally recurrent or metastatic breast cancer in post-menopausal women. The scope specified people with advanced breast cancer.

Summary of submitted clinical effectiveness evidence

A good quality systematic literature review of clinical effectiveness identified one randomised controlled trial (RCT) of abemaciclib relevant to the decision problem. The MONARCH 3 trial was a double blind, phase III RCT of abemaciclib (150 mg taken orally twice daily) and NSAI (ABE+NSAI) versus (vs) placebo+NSAI (n=493 patients randomised). The NSAIs used were either letrozole or anastrozole (investigator choice). A small number of patients from the UK ([redacted]) were enrolled in the trial. MONARCH 3 was judged by the ERG to be of reasonable methodological quality, though the possibility of unblinding, imbalance in drop-outs and selective reporting of outcomes increasing the risk of bias. The ERG believes that the company has identified all the relevant available RCTs of abemaciclib.

The CS presents interim results from MONARCH 3 (pre-specified and previously published) at a median follow-up of 17.8 months (data cut-off 31st January 2017), and results at the final progression free survival (PFS) assessment (from a confidential clinical study report) at a median follow-up of [redacted] months (data cut-off 3rd November 2017). Analyses were from an intention-to-treat (ITT) population for the majority of outcomes. The primary outcome of PFS (defined as the date of randomisation to objective progression or death) was investigator-assessed at the interim and final analysis. An independent review of PFS was also undertaken at both assessments.

There are no known trials of ABE+NSAI compared with the scoped comparators palbociclib (PAL) and ribociclib (RIBO). The CS present a Bayesian network meta-analysis (NMA) using published methods to perform indirect comparisons with these (and other) comparators (we refer to this as the 'first-line treatment NMA' in this report). A broad range of (non-scoped) comparator treatments were eligible from the SLR informing the NMA to allow a fully connected network. The NMA included a total of 18 RCTs, though only four of these were directly relevant to the decision problem: The MONARCH 3 trial of abemaciclib; the MONALEESA-2 trial of ribociclib; the PALOMA-1/TRIO-18 and PALOMA-2 trials of palbociclib (all with respective NSAIs). The ERG believes the SLR has identified all relevant RCTs. OS and PFS results from this NMA are used to inform the economic model: PFS results inform the time to first progression estimate and OS results inform the estimate of deaths before first progression (see below for a description of the economic model).

The company also briefly presents an additional NMA (in an appendix) to provide comparative evidence of abemaciclib as a second-line treatment in advanced breast cancer. The phase III MONARCH 2 RCT, which compares abemaciclib and fulvestrant to placebo and fulvestrant, is indirectly compared with trials of other endocrine therapies for patients who have progressed following first-line treatment for advanced breast cancer. This NMA (referred to in this report as the 'second-line treatment' NMA) was necessary as the OS data from the MONARCH 3 trial are immature and the economic model therefore includes a PFS2 health state to estimate OS from abemaciclib indirectly via the effects of second-line and subsequent treatment lines.

In the MONARCH 3 trial at the final PFS analysis:
Investigator assessed median PFS was [redacted] months in the ABE+NSAI group compared with [redacted] in the placebo+NSAI group; HR [redacted] (95% CI [redacted], 2-sided [redacted]), giving a reduction in the risk of progression of disease or death of 46%. Expert clinical advice to the ERG is that these results are clinically meaningful.

Median OS was [redacted], HR [redacted] (95% CI [redacted] 2-sided stratified log-rank [redacted]). [redacted] the OS rate at 24 months (ABE+NSAI [redacted] vs placebo+NSAI [redacted] OS data are currently immature ([redacted] events recorded, with final OS analysis to be done after 315 events).

The investigator assessed objective response rate (defined as the proportion of patients with best response of complete response (CR) or partial response (PR)), was [redacted] with ABE+NSAI compared with placebo+NSAI ([redacted]).

Among patients with an investigator assessed response (ABE+NSAI n=163, placebo+NSAI n=61), the median duration of response was [redacted] months (95% CI [redacted]) in the ABE+NSAI arm compared with [redacted] months (95% CI [redacted]) in the placebo+NSAI arm.

Health related quality of life (HRQoL), assessed on the global health status of the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30), showed a [redacted]. On the specific symptom scale of [redacted]) was observed, with [redacted] There was [redacted] in change from baseline in the EuroQol 5-Dimension 5-level (EQ-5D-5L) index score or visual analogue scale.

Proportions of participants with adverse events were higher in the ABE+NSAI arm for at least one adverse event judged as related to treatment ([redacted] ABE+NSAI vs [redacted] placebo+NSAI); grade ≥3 adverse events (ABE+NSAI arm [redacted] vs placebo+NSAI arm [redacted]; at least one serious adverse event (ABE+NSAI arm [redacted] vs placebo+NSAI arm [redacted]); serious adverse events judged to be related to study treatment (ABE+NSAI group [redacted]; placebo+NSAI group [redacted] and discontinuations of all study treatments (ABE+NSAI arm [redacted] vs placebo+NSAI arm [redacted]).

All treatment emergent adverse events, with the exception of arthralgia and back pain, occurred more frequently in the ABE+NSAI arm. Specific grade ≥3 adverse events of interest were diarrhoea (ABE+NSAI [redacted]; placebo+NSAI [redacted]); neutropenia (ABE+NSAI [redacted]; placebo+NSAI [redacted]) and leukopenia (ABE+NSAI [redacted]; placebo+NSAI [redacted]).

First-line treatment NMA results for abemaciclib, ribociclib and palbociclib:
For PFS (fixed effect model) all three treatments showed similar and statistically significant hazard ratios improving PFS relative to NSAI (ABE+NSAI [redacted]; PAL+NSAI [redacted]; RIBO+NSAI [redacted]). The random effects model resulted in similar hazard ratios but much wider credible intervals, and statistically nonsignificant differences relative to NSAI for each of the three treatments. There were no significant differences for the indirect comparisons between ABE+NSAI and PAL+NSAI or ABE+NSAI and RIBO+NSAI using either fixed or random effects models.

There were no statistically significant differences in OS for any of the three treatments relative to NSAI. Data for OS are immature and results are therefore uncertain. Similarly, there were no significant differences in OS for the indirect comparisons between ABE+NSAI and PAL+NSAI or ABE+NSAI and RIBO+NSAI using either fixed or random effects models. There were no statistically significant differences in measures of tumour response for any of the three treatments relative to NSAI.

Summary of submitted cost effectiveness evidence

Model structure and assumptions

The submission includes a three-state Markov model that estimates time spent progression-free on first-line treatment (PFS1) and time post-progression, for a cohort of people with HR+ HER2- advanced breast cancer. Costs and QALYs accumulated in the PFS1 health state are estimated in this model, but costs and QALYs after progression are estimated in a separate 'fixed pay-off' sub-model. The latter uses a 'partitioned survival' approach, using progression-free survival and overall survival outcomes for second-line treatments. Thus, PFS and OS evidence for second-line treatments contributes to estimating first line survival. Calibration enables exploration of uncertainty over the relationship between PFS and OS. The model adopts a 'partial surrogacy' approach, similar to that in the NICE appraisal of ribociclib TA496, with an assumption that the median gain in OS is a fixed proportion (27.5% in the base case) of the median gain in PFS for the first-line treatments.

Key model parameters are:

- *Clinical effectiveness*: Time to progression and deaths before progression for the first-line model, and PFS and OS for the second-line sub-model. These parameters are estimated individual data from the MONARCH 3 trial (first-line for ABE+NSAI and NSAI) and the MONARCH 2 trial (second-line fulvestrant), and from relative treatment effects from the first and second-line NMAs. This entails a series of assumptions that we critique in Chapter 4, and we highlight particular uncertainties below. In addition, as noted above, the company assumed a 'partial surrogacy' rate for calibration of the OS/PFS relationship of 27.5% (with 100% in scenario analysis).

- *Health Related Quality of Life*: Health state utility are derived from EQ-5D-5L data from patients in the MONARCH 3 and MONARCH 2 trials and from literature cited in other related NICE appraisals. MONARCH 3 was used in the company base case for the progression-free period at first-line ([redacted] for all comparators). For post-progression utilities, the company used the same estimates as the company in TA496, based on a formula reported by Lloyd et al. (2006): 0.774 for progression-free on second-line treatment, with an additional decrement of -0.113 for chemotherapy; and 0.505 for post second-line progression. Disutilities for adverse drug reactions are included in the model, but as the size and duration of the effects assumed are low, these have a negligible impact on cost-effectiveness results.

- *Use of second and third line treatments*

  The company assumes a mix of treatments at second and third line, based on the submission for the NICE appraisal of Fulvestrant at first line (TA503). This includes fulvestrant, exemestane, tamoxifen, everolimus with exemestane and chemotherapy.

- *Duration of treatment*

  Discontinuation rates for first and second-line treatments are modelled using survival curves, also estimated from MONARCH 3 (abemaciclib and NSAI) and MONARCH 2 (second-line fulvestrant). For other drugs, discontinuation is modelled relative to these curves, with hazard ratios estimated from median time to discontinuation reported in the key trials. Costs for third line treatments are included in the model, with an assumption that patients spend 37% of post-progression time on treatment.

- *Resource use and health care costs*

In addition to drug acquisition and administration costs (first, second and third line), the model includes costs for follow up care and monitoring, treatment of adverse drug reactions, hospital admissions, best supportive care and end of life care. Resource use was estimated using records from the MONARCH 3 and MONARCH 2 clinical trials, and recommendations in the NICE clinical guideline for advanced breast cancer (CG81). Average monthly non-drug costs were estimated at around £730 to £830 and end of life care at £4,400.

The company's base case results are shown in the table below – calculated at list price for abemaciclib and all comparators and subsequent treatments. Based on this analysis, the company concluded that ABE+NSAI is more effective, accruing more life years and QALYs, and less expensive than the comparators PAL+NSAI and RIBO+NSAI. They note that the lower costs are driven by a shorter time on treatment with ABE+NSAI.

Treatment	**Total**		**Incremental analysis ICER (£/QALY)**	**Pairwise ICERs ABE+NSAI vs. comparator (£/QALY)**
	Costs (£)	**QALYs**		
NSAI	£56,449	2.997	Referent	£250,065
PAL+NSAI	£145,266	3.225	Dominated	ABE+NSAI Dominant
RIBO+NSAI	£148,170	3.222	Dominated	ABE+NSAI Dominant
ABE+NSAI	£129,803	3.291	£250,065	-

The company presented similar results from a probabilistic sensitivity analysis (PSA), but we note that the probabilistic analysis did not reflect correlations between NMA parameters. The company present 29 deterministic scenario analyses, testing the impact of selected changes in model assumptions and parameters. These produced similar ICERs for ABE+NSAI compared with NSAI (in the range of £160,000 to £572,000 per QALY gained) and fewer QALYs at higher costs for the comparators PAL+NSAI and RIBO+NSAI compared with NSAI. The company did not present a one-way sensitivity analysis for model parameters, or tornado diagram.

Commentary on the robustness of submitted evidence

Strengths

- The company conducted a good quality systematic review to identify relevant clinical effectiveness trials. All relevant trials are believed to have been included.
- The clinical effectiveness evidence for abemaciclib comes from a relatively large (n=493 patients) phase III double-blind multinational RCT (MONARCH 3). The ERG judged this trial to be of reasonable methodological quality, though with some potential risks of bias (see below).
- The company's indirect comparison of abemaciclib with palbociclib and ribociclib (the first-line treatment NMA), used standard statistical methods, though there are some methodological limitations (see below).
- The economic model structure is appropriate, given the immaturity of overall survival data for abemaciclib and for the comparators ribociclib and palbociclib. There is considerable uncertainty over the assumptions and parameters of the second-line model and over the partial surrogacy assumption. However, a standard partitioned-survival approach would likely be even more uncertain.
- Methods used to estimate survival functions are generally appropriate, though parameters are not provided for some functions and the reporting is rather sparse.
- MONARCH 3 was used for the estimate of utility in the progression-free period at first-line ([redacted] for all comparators). This complies with the NICE reference case (assuming that crosswalk values are used as stated); uses EQ-5D-5L data collected directly from participants in the pivotal trial; and the methods of analysis are appropriate, although we do have some reservations related to lack of detail in reporting.
- We have a general preference for the treatment-specific utility estimates from MONARCH 3, because they reflect benefits and harms of treatments directly assessed by patients. But equivalent treatment-specific utilities are not available for all comparators. We therefore agree with the company's decision to use the overall PFS1 utility for all comparators in their base case.

Weaknesses and areas of uncertainty

- A high frequency of adverse events such as diarrhoea in the ABE+NSAI arm of the MONARCH 3 trial could have lead to unblinding, thus increasing the potential risk of detection and performance bias.
- The OS data from the MONARCH 3 trial are immature. The estimated study completion date is February 2020.
- There are some uncertainties associated with the first-line treatment NMA:
    - There is variation between the included trials in the proportion of patients with visceral metastases (affecting internal organs including the liver, lungs or brain), and the effect of this on the results is uncertain.
    - The NMA method used assumes the proportional hazards assumption holds for survival outcomes. However, this assumption could not be supported by available data for some trials.
    - Due to the immaturity of the OS data in the scoped treatment trials the ERG considers the results of the first-line OS NMA to be highly uncertain.
    - Despite the limitations listed above, the results were considered by clinical experts advising the ERG to be clinically plausible.
- There are likewise uncertainties associated with the second-line treatment NMA, namely:
    - Apparent clinical heterogeneity between the included trials in terms of percentage of patients with visceral metastases, the number of prior treatments for advanced breast cancer received and HER2 status. The comparability of the MONARCH 2 trial to the comparator trials is questionable.
    - Proportional hazards do not appear to hold for all the trials included, for both OS and PFS.
    - OS data are immature in some of the trials, including the MONARCH 2 trial. The results of the OS network should therefore be interpreted with caution.
- Exploration of uncertainty around the model results is limited. The PSA does not include correlations between NMA parameters and one-way deterministic sensitivity analysis is

not presented. The use of calibration within the model made it slow to run, so use of the PSA and other sensitivity analysis is difficult.

- On the basis of fit to observed data and clinical judgement on the plausibility of extrapolations, we agree with the choice of exponential survival curve for the time to progression on first-line treatment. We note, however, that the first line NMA indicated similar treatment effects for abemaciclib, ribociclib and palbociclib. This conflicts with the larger advantage predicted for abemaciclib when estimated directly from MONARCH 3 data. A similar issue arises when estimating the first-line pre-progression death rate, but in the opposite direction: direct estimation from MONARCH 3 for ABE+NSAI (jointly estimated with NSAI) gives a higher mortality rate than when this parameter is estimated from the NMA relative effects. Given that the decision problem is focussed on comparison between abemaciclib, ribociclib and palbociclib, it is important that comparators are modelled in a consistent way, and the NMAs are best source of evidence to judge relative treatment effects.
- At second-line, the company use data from the MONARCH 2 trial to estimate PFS and OS for second-line fulvestrant, with other drugs modelled relative to this curve using NMA results. As noted above, we have concerns over heterogeneity of the second-line trials and hence over the robustness of the NMA.
- The company choose to model second-line OS with an exponential curve fitted to the fulvestrant arm of MONARCH 2, and long-term extrapolation based on the CONFIRM trial. We disagree with this approach. Firstly, because the exponential curve had a poor fit to the MONARCH 2 data. Secondly, because very little information is provided to justify the fitting of the Weibull survival curve to the CONFIRM trial data. Based on evidence of goodness-of-fit and consideration of the plausibility of extrapolations, we consider the Gompertz or Log-logistic curves fitted to MONARCH 2 data are likely to be more reliable.
- Regarding the company's utility estimates in the base case, we suggest that the value used for second-line progression-free survival (0.69) in the final version of the TA496 appraisal looks more realistic than the original estimate, which is higher than the company's estimated for first-line utility.
- Our main concern over resource use assumptions: that the estimated use of second and third-line treatments does not reflect current NHS practice. In particular, the company includes fulvestrant which is not recommended by NICE in this context.

Summary of additional work undertaken by the ERG

We identified four minor errors in the coding of the model, which we correct. These made very little difference to the company's results. We also ran a range of scenario analyses to test uncertainties around model assumptions and parameters. Our preferred version of the model included the following changes to the company's base case:

- Estimation of time to progression and pre-progression deaths for ABE+NSAI estimated relative to fitted curves for NSAI using hazard ratios from the first-line NMA (as for other comparators.
- A Gompertz OS curve from second-line treatment. This was more pessimistic than the company's assumption of exponential with CONFIRM trial extrapolation.
- A utility of 0.69 for people free of progression at second line – as per the assumption suggested by the Decision Support unit in the NICE appraisal of ribociclib (TA496).
- And an alternative set of assumptions about use of second and third line treatments. This include the assumption that no patients would have fulvestrant, lower rates of exemestane monotherapy and higher rates of everolimus with exemestane at second line, higher rates of chemotherapy and fewer patients receiving no treatment at third line.

This version of the model (with list prices for all drugs) gave similar results to the company base case: an ICER of just under £200,000 per QALY gained for abemaciclib + NSAI compared with NSAI alone, compared with about £250,000 in the company's base case. For most scenarios tested, abemaciclib remained dominant or cost-effective compared with ribociclib and palbociclib. The absolute difference in QALYs between the CDK 4/6 inhibitors was very small, and the ranking of abemaciclib, ribociclib and palbociclib did change between scenarios. However, as the company note, the lower costs of abemaciclib are driven by a shorter time on treatment with ABE+NSAI. We note that this difference is based on weak evidence, as hazard ratios between treatments were estimated from reported median time to discontinuation.

1 Introduction to the ERG Report

This report is a critique of the company's submission (CS) to NICE from Eli Lilly and Company Limited on the clinical effectiveness and cost effectiveness of abemaciclib with an aromatase inhibitor for untreated advanced hormone receptor-positive (HR+), human epidermal growth factor receptor 2-negative (HER2-) breast cancer. It identifies the strengths and weaknesses of the CS. Clinical experts were consulted to advise the Evidence Review Group (ERG) and to help inform this report.

Clarification on some aspects of the CS was requested from the company by the ERG via NICE on (11th July 2018). A response from the company via NICE was received by the ERG on 26th July 2018 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 company presents an accurate overview of breast cancer and its pathogenesis in CS section B.1.3. Breast cancer is the most common cancer amongst women in the UK (age-standardised incidence rate of 95.0 per 100,000[1]) and is responsible for 7% of all cancer deaths in the UK (mortality rate of 17.1 per 100,000[1, 2]). The annual breast cancer incidence in England and Wales is 0.08% (~46,700 women),[3-5] of which approximately 90% of patients are diagnosed with invasive breast cancer.[3] The majority of these women (95%) are estimated to have early and locally advanced disease,[3] in which the cancer has not spread to other parts of the body. Approximately 35% of these women progress to advanced metastatic breast cancer,[3] where the disease has spread (metastasised) to other parts of the body (e.g. bones, liver, and lungs) or has grown into tissues and is unable to be removed completely by surgery.[6] An estimated 13% of women in the UK have advanced breast care at diagnosis.[3, 7] Advanced breast cancer is associated with poorer outcomes and is incurable, with a median overall survival (OS) of 2–3 years.[8]

The population of relevance to this appraisal is people with untreated advanced HR+ and HER2- breast cancer. Breast tumours are tested for oestrogen receptors (ER) and progesterone receptors (PgR), which stimulate tumour growth. ER+ or PgR+ tumours are commonly referred to as being HR+. The majority of HR+ tumours are both ER+ and PR+, while around 15% to 20% are ER+ only. Patients with HR+ breast cancer generally have an improved prognosis

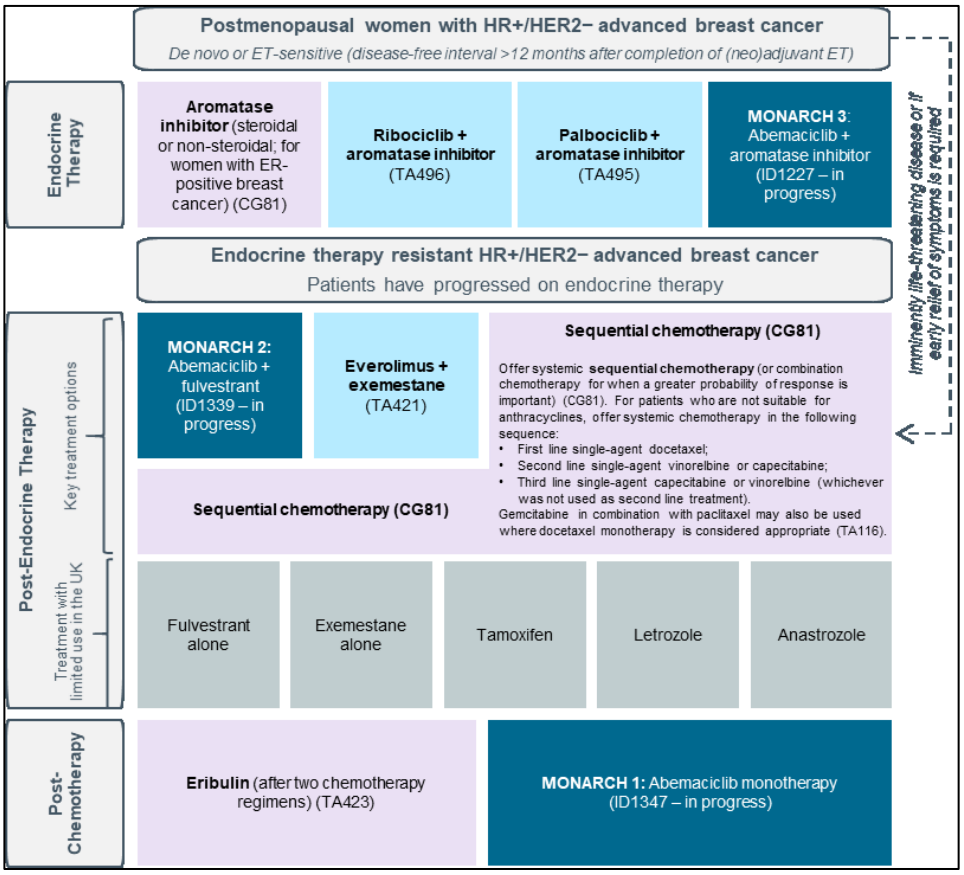
compared to those who are HR-negative (HR-), as tumours tend to grow more slowly and are more likely to respond to hormonal therapy (i.e. endocrine therapy). Endocrine therapy lowers the amount of available oestrogen or blocks existing oestrogen from binding to its receptor.

HER2 is a protein is found on the surface of breast cancer cells that can affect the growth of some cancer cells. Patients with HER2- breast cancer generally have an improved prognosis compared to those who are HER2+. The most common type of advanced breast cancer is ER+/HER2−, applying to approximately 64% of women with metastatic breast cancer in the UK.

The highest rates of breast cancer occur in older people, with ≥80% of cases in women over the age of 50 years (60 years or over for the majority of men) and 25% in women aged at least 75 years.[9] The CS describes the effects of breast cancer on patients and carers. Disease progression and side effects from treatment impact on the patient's ability to work, carry out of daily activities and on their emotional well-being. HER2- metastatic breast cancer is associated with worsening symptoms related to pain, fatigue, sleeplessness and acute distress.[10] This not only creates a burden for the patient, but also for their caregiver. Slowing disease progression and reducing treatment-related adverse events is therefore crucial for maintaining good health-related quality of life (HRQoL).[10]

2.2 Critique of company's overview of current service provision

The CS (Section B.1.3.3) describes the current treatment pathway for advanced breast cancer, based on current NICE guidance, and the intended position of abemaciclib in the pathway (Figure 1). Only abemaciclib in combination with an aromatase inhibitor (AI) as first-line treatment is relevant to this appraisal. Separate NICE appraisals will assess abemaciclib as a second-line and third-line treatment for advanced breast cancer (NICE ID1339 and ID1347, respectively). Expert clinical advisors to the ERG consider that the pathway is reflective of current clinical practice. However, they note that AI monotherapy would only now be used in a minority of patients given that ribociclib[11] and palbociclib[12] have been recommended by NICE for use in the NHS.



Source: CS Figure 2

Figure 1 Clinical pathway for patients with HR+/HER2− advanced or metastatic breast cancer being treated with abemaciclib + aromatase inhibitor

2.3 Critique of company's definition of decision problem

2.3.1 Population

The population described in the decision problem is post-menopausal women with advanced HR+/HER2− locoregionally recurrent or metastatic breast cancer, who have had no prior systemic therapy for advanced disease. (NB. in locoregional recurrent breast cancer the cells are identified in the same breast as the original tumour or in nearby lymph nodes, clarification question A1). Patients who have received treatment with endocrine therapy in the (neo)adjuvant setting with a disease-free interval >12 months from completion of endocrine therapy are included.

The ERG queried with clinical experts whether the inclusion of locoregionally recurrent breast cancer would potentially exclude patients with newly occurring (de novo) locally advanced breast cancer. The experts clarified that in routine practice the majority of these patients would be treated with chemotherapy in an attempt to downstage them and they would then receive surgery. The patients are unlikely to be entered into palliative treatment trials such as those relevant to this appraisal.

The company's decision problem reflects the patient population in the pivotal clinical trial of abemaciclib included in the CS (MONARCH 3[13] - see Table 2). While this approach appears reasonable, it does omit men with the disease potentially eligible under the NICE scope (the scope, which is aligned with the marketing authorisation, mentions "people with advanced hormone-receptor positive HER2-negative breast cancer"). The anticipated marketing authorisation does not exclude men (CS page10).

2.3.2 Intervention

The description of the intervention (abemaciclib + non-steroidal AI [ABE+NSAI]), is appropriate to the NHS and the NICE scope. Abemaciclib is a selective dual inhibitor of cyclin-dependent kinase 4 and 6 (CDK4 and 6). The starting dose of abemaciclib is 150 mg twice daily, reflecting the recommended dose of abemaciclib in the draft Summary of Product Characteristics (SmPC) when used in combination with endocrine therapy.[14] Abemaciclib should be taken continuously as long as the patient is deriving clinical benefit from therapy or until unacceptable toxicity occurs. Dose interruption and/or dose reduction due to adverse events are recommended (see Table 1), such as for hematologic toxicities, diarrhoea and increased alanine aminotransferase levels.

Table 1 Dose adjustment recommendations for adverse reactions

Draft SmPC[14]	**Abemaciclib dose combination therapy[a]**
Recommended dose	150 mg twice daily
1st dose adjustment	100 mg twice daily
2nd dose adjustment	50 mg twice daily
3rd dose adjustment	-

[a] dose reductions for monotherapy not presented here

The decision problem states that either anastrozole or letrozole can be chosen as the NSAI to be used in combination with abemaciclib.

2.3.3 Comparators

The comparators are palbociclib + NSAI (PAL-NSAI) (letrozole) and ribociclib + NSAI (RIB+NSAI) (letrozole). These are appropriate for the NHS and reflect the NICE scope. Clinical experts advising the ERG consider palbociclib and ribociclib equivalent in effectiveness and safety, and the choice between them would be down to patient and clinician preference.

2.3.4 Outcomes

The outcomes stated in the CS scope are overall survival (OS), progression-free survival (PFS), tumour response rate, adverse effects of treatment and health-related quality of life (HRQoL). These are standard outcomes measured in cancer treatment trials and reflect those in the NICE scope.

2.3.5 Economic analysis

The economic analysis described in the decision problem is appropriate for the NHS. Cost-effectiveness is expressed in terms of the incremental cost per quality-adjusted life years (QALY) and costs are considered from the perspective of the NHS and personal social services (PSS), with a 35-year time horizon, using a Markov state-transition model with a fixed 'pay-off' for post-progression survival (see section 4 of this report for further description of the economic analysis).

2.3.6 Other relevant factors

The NICE scope does not contain any patient subgroups. The CS presents a summary of subgroup analyses of PFS and OS from the MONARCH 3 trial of abemaciclib (CS Appendix E). These are discussed in further detail in section 3.1.6 and section 3.3.6 of this report.

The company does not identify inequality issues that could be associated with the introduction or provision of abemaciclib (CS Section B.1.4). However, incidence is relatively uncommon the ERG consider that there is a potential issue of excluding men with advanced breast cancer. Expert clinical advice to the ERG is that in practice men with advanced breast cancer would be treated with goserelin acetate and palbociclib or ribociclib.

3 CLINICAL EFFECTIVENESS

3.1 Critique of the company's approach to systematic review

3.1.1 Description of the company's search strategy

The CS reports four systematic literature searches:

- Clinical effectiveness evidence: searched from database inception to December 2015. Updated twice: March 2017 and January 2018 (CS Appendix D).
- Cost effectiveness: searched from 2010 to April 2016. Updated in November 2017 (CS Appendix G).
- Health related quality of life: searched from database inception to April 2016. Updated in November 2017 (CS Appendix H).
- Cost & healthcare resource identification measurement and valuation: searched from 2010 to 15th April 2016. Updated in November 2017 (CS Appendix I).

All four literature search strategies are of sound methodology, well documented and reproducible. An acceptable range of databases were searched with the application of appropriate syntax, good balance of descriptive terms and free text terms, with the use of suitable search filters. Key conferences were recorded as searched. The following ongoing trials databases were documented as searched: clinicaltrials.gov and the World Health Organisation International Clinical Trials Registry Platform (WHO ICTRP). No further published trials were identified by the ERG via an internet search and a Delphis database search (a University of Southampton cross-database search platform). The decision was therefore taken not to run full replicated update searches on the databases cited in the submission. An ongoing trials search, restricted to trials of abemaciclib that are currently recruiting patients, was undertaken by the ERG, to identify any other relevant trials not captured in the submission. Databases searched were clinicaltrials.gov, the UK Clinical Trials Gateway (UKCTG) and the WHO ICTRP (see section 3.3.9 for further details of ongoing studies). In summary, the ERG considers that the company's literature searches are all fit for purpose.

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

The CS clearly presents the eligibility criteria for the SLR (CS Appendix D.1.2, Table 9). The SLR was also used in the CS to identify studies of relevance to a network meta-analysis (NMA) which indirectly compares abemaciclib with relevant comparators. We describe this NMA in section 3.1.7 of this report.

3.1.2.1 Population

The company used wider population criteria for the SLR than the MONARCH 3 trial population criteria (see Table 2). The company justifies this because the specific characteristics of the patients in the MONARCH 3 trial meant that low returns of relevant literature were expected. The final included population was post-menopausal women with advanced HR+/HER2− locoregionally recurrent or metastatic breast cancer who have had no prior systemic therapy for advanced disease. Patients who had received treatment with endocrine therapy in the (neo)adjuvant setting with a disease-free interval of more than 12 months from completion of endocrine therapy were included). This reflects the patients in the MONARCH 3 trial, where the inclusion criteria were age ≥18 years, with patients required to be post-menopausal (either having had prior bilateral oophorectomy or aged ≥60 years).

3.1.2.2 Intervention

The inclusion criteria specify abemaciclib as single agent (not relevant to this appraisal) or combination therapy with NSAI. This is broader than the scope of this appraisal, but in line with the anticipated marketing authorisation which covers use of abemaciclib at first, second and third line treatment in locally advanced or metastatic breast cancer (see Figure 1).

3.1.2.3 Comparators

For inclusion studies had to compare to ≥1 listed treatments from below, or to placebo:

- Endocrine therapy (i.e. anastrozole; exemestane; fulvestrant; letrozole; megestrol acetate; tamoxifen; toremifene);
- Chemotherapy (i.e. capecitabine; docetaxel; doxorubicin; liposomal; gemcitabine; paclitaxel; nanoparticle bound; vinorelbine);
- Targeted therapy (i.e. buparlisib; ribociclib);
- Combination chemotherapy (i.e. AC (doxorubicin + cyclophosphamide); CAF (cyclophosphamide + doxorubicin + fluorouracil); docetaxel + capecitabine; gemcitabine + carboplatin; gemcitabine + paclitaxel);
- Combination endocrine and targeted therapy (i.e. buparlisib in combination with paclitaxel, or with ribociclib + letrozole, or with tamoxifen; exemestane + everolimus; palbociclib in combination with anastrozole, or with everolimus + exemestane, or with exemestane, or with fulvestrant, or with letrozole, or with tamoxifen; ribociclib in combination with anastrozole, or with capecitabine, or with exemestane, or with fulvestrant, or with letrozole, or with tamoxifen);
- Combination chemotherapy and targeted therapy (i.e. paclitaxel + bevacizumab).

Expert clinical advice to the ERG is that toremifene is no longer used, and that the chemotherapy drug eribulin is absent from the list. Also, buparlisib is not yet licensed; ribociclib is not licensed for use in combination with tamoxifen or capecitabine; and palbociclib is not licensed for use with exemestane + everolimus or tamoxifen.

Whilst the included comparators are broader than those listed in the NICE scope for this appraisal, the purpose was to identify relevant studies which could be included in the NMA (section 3.1.7.1). Additional comparators, even if not yet licensed or recommended by NICE, can link the NICE scoped treatments indirectly in networks.

3.1.2.4 Outcomes

The effectiveness and safety outcomes reflect those specified in the NICE scope and decision problem (OS, PFS, response rate; adverse effects of treatment; HRQoL).

3.1.2.5 Design

The eligibility criteria permits studies using randomised controlled trials (RCTs) and non-RCTs. Non-RCTs were identified in the first version of the SLR, but were not included in the updated SLR used to inform the CS as a sufficient number of RCTs were identified. The ERG considers this to be acceptable.

The CS presents a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram detailing the original literature search in 2015 and the two updated searches in 2017 and 2018. Details of excluded and included studies for all three searches are provided at each stage and a list of references for both are presented in the appendix of the CS (CS Appendix D.1.2, Table 10 and 11).

3.1.3 Identified studies

The company's SLR identified one phase III RCT, the MONARCH 3 trial, funded by Eli Lilly.[13] The CS presents sufficient summary details for the trial (CS section B.2.3: Table 4 trial inclusion and exclusion criteria; Table 5 trial design, intervention, population, outcomes, description of intention-to-treat (ITT) analysis, subgroups; Table 8 statistical analysis, power/sample size calculations, data management). A flow diagram details patient allocation and discontinuations (CS Appendix D.2, Figure 8). All relevant references were provided by the company electronically with the submission.

Patients in MONARCH 3 were randomised in a 2:1 ratio, with randomisation stratified by:

- site of metastases: visceral (lung, liver, pleural, peritoneal, or adrenal gland involvement); bone only, or other;
- prior (neo)adjuvant endocrine therapy: AI therapy (e.g. anastrozole, exemestane and letrozole), other, or no prior endocrine therapy.

A total of 328 patients were randomised to abemaciclib (150 mg taken orally twice daily) +NSAI and 165 to placebo+NSAI. The NSAIs were either letrozole (2.5 mg taken orally once daily) or anastrozole (1 mg taken orally once daily) in both treatment arms (investigator choice), with the majority of patients receiving letrozole (79.1%). The CS states that patients should have remained on the same NSAI throughout the study. Treatment was provided on a continuous 28-day treatment cycle.

Patients were not permitted to cross-over between trial arms; however, they were allowed to discontinue either abemaciclib/placebo or NSAI, and continue the other drug as a monotherapy. In response to a clarification question (A5) the company reported the percentage of patients receiving post-discontinuation therapies, ([redacted]% in the ABE+NSAI arm vs [redacted]% in the placebo+NSAI arm). The most common post-discontinuation therapies included endocrine therapy ([redacted]) (e.g. fulvestrant) and chemotherapy ([redacted]) (e.g. paclitaxel).

Patients' mean age was around 63 years, with the [redacted] of patients Caucasian ([redacted]) and [redacted] of included patients were enrolled at European sites ([redacted]), including four sites in the UK (Table 3). The company clarified that [redacted] patients from the UK were randomised in the MONARCH 3 trial; [redacted] were allocated to the ABE+NSAI arm and [redacted] to the placebo+NSAI arm (clarification question A2).

All patients with reported HR and HER2 receptors (missing data n=1, placebo arm) had breast cancer that was HER2- and around [redacted] had cancer that was positive for both ER and PgR hormone receptors. Baseline data for Eastern Cooperative Oncology Group (ECOG) performance status, disease setting, receptor status, initial diagnosis disease stage, metastatic site, number of organ sites, prior (neo)adjuvant chemotherapy/endocrine therapy and measurable disease were comparable between treatment arms. Median duration of disease was around [redacted] in the ABE+NSAI arm compared with placebo+NSAI arm ([redacted] months, respectively) and the proportion of patients with treatment-free interval

of ≥36 months higher (62.7 % vs 50.0% respectively). This suggests that the ABE+NSAI arm had some better prognostic factors at baseline, potentially favouring the treatment effects for this arm.

Around 40% of patients had de novo metastatic disease (slightly higher in the ABE+NSAI arm, Table 3) and approximately [redacted] had prior endocrine therapy in the neo(adjuvant) setting (slightly higher use of (neo)adjuvant NSAI in the placebo+NSAI arm).

The CS summarises selected categories of concomitant medication use (Table 3). Nearly all the patients received concomitant medication regardless of treatment allocation (ABE+NSAI [redacted], placebo+NSAI [redacted]), with details only reported for treatment received in [redacted] of patients. Differences between the treatment arms existed in the use of loperamide (an antidiarrhoeal) (ABE+NSAI [redacted] vs placebo+NSAI [redacted]) and therefore also in the antidiarrhoeal category (ABE+NSAI [redacted] vs placebo+NSAI [redacted], both [redacted] in patients receiving abemaciclib. Use of ≥1 antiemetics + anti-nauseants, erythropoietic agents, granulocyte-colony stimulating factor and granulocyte-macrophage colony stimulating factor [redacted] in patients receiving abemaciclib compared with placebo.

Table 2 Population as defined in the NICE scope, MONARCH 3, company decision problem and anticipated marketing authorisations

NICE final scope	Trial inclusion (MONARCH 3)	Company decision problem	Anticipated EMA marketing authorisation (CS p10)
People with advanced HR+/HER2− breast cancer that has not been previously treated with endocrine therapy	Postmenopausal women (≥18 years) with HR+/HER2− locoregionally recurrent or metastatic breast cancer who had no prior systemic therapy in the advanced setting Exclusion criteria: prior (neo) adjuvant ET with a disease free interval of ≤12 months from completion of treatment	Postmenopausal women with advanced HR+/HER2− locoregionally recurrent or metastatic breast cancer who have had no prior systemic therapy for advanced disease (patients who have received treatment with endocrine therapy in the (neo)adjuvant [a] setting with a disease-free interval >12 months from completion of ET are included).	Abemaciclib is expected to be indicated for the treatment of HR+/HER2− locally advanced or metastatic breast cancer: • **in combination with an aromatase inhibitor as initial endocrine-based therapy (current appraisal)** or in women who have received prior endocrine therapy • in combination with fulvestrant as initial endocrine-based therapy, or in women who have received prior endocrine therapy • as monotherapy following disease progression after endocrine therapy and one or two chemotherapy regimens in the metastatic setting

[a] As defined in the MONARCH 3 trial

Table 3 Baseline characteristics - MONARCH 3 trial

Baseline characteristic	Abemaciclib + NSAI (n=328)	Placebo + NSAI (n=165)
Age		
Mean (SD)	[redacted]	[redacted]
Race, n (%) [a,b]		
White	186 (56.7)	102 (61.8)
Asian	103 (31.4)	45 (27.3)
Other	11 (3.4)	7 (4.2)
Region, n (%)		
Europe	[redacted]	[redacted]
ECOG performance status		
0	192 (58.5)	104 (63.0)
1	136 (41.5)	61 (37.0)
Disease setting, n (%)[c]		
De novo metastatic	135 (41.2)	61 (37.0)
Metastatic recurrent	182 (55.5)	99 (60.0)
Locoregionally recurrent	11 (3.4)	5 (3.0)
Receptor status, n (%)		
ER+/PgR+	[redacted]	[redacted]
ER+/PgR-	[redacted]	[redacted]
ER+/PgR unknown	[redacted]	[redacted]
ER-/PgR+	[redacted]	[redacted]
HER2 receptor status		
Negative	[redacted]	[redacted]
Missing[d]	[redacted]	[redacted]
Duration of disease (months)		
Median (IQR)	[redacted]	[redacted]
Prior (neo)adjuvant chemotherapy, n (%)		
Yes	125 (38.1)	66 (40.0)
No	203 (61.9)	99 (60.0)
Prior (neo)adjuvant endocrine therapy, n (%)		
None	178 (54.3)	85 (51.5)
Aromatase inhibitor	85 (25.9)	50 (30.3)
Other endocrine therapy	65 (19.8)	30 (18.2)
Treatment-free interval, n (%)[e]		
<36 months	42/150 (28.0)	32/80 (40.0)
≥36 months	94/150 (62.7)	40/80 (50.0)
Unknown	14/150 (9.3)	8/80 (10.0)
Measurable disease, n (%)		
Yes	267 (81.4)	130 (78.8)
No	61 (18.6)	35 (21.2)
Selected categories of concomitant medications during trial (safety population), n (%)		
	(n=327)	**(n=161)**
Patients with ≥1 analgesic	[redacted]	[redacted]
Patients with ≥1 antidiarrheal	[redacted]	[redacted]
Patients with ≥1 antiemetics and anti-nauseants	[redacted]	[redacted]
Patients with ≥1 bone-modifying agents	[redacted]	[redacted]
Patients with ≥1 erythropoietic agents	[redacted]	[redacted]
Patients with ≥1 G-CSF/GM-CSF	[redacted]	[redacted]

Source: CS Table 6 and 7 based on Goetz et al. 2017[13] and CSR. ECOG, Eastern Cooperative Oncology Group; ER, oestrogen receptor; G-CSF, granulocyte-colony stimulating factor; GM-CSF, granulocyte-macrophage colony-stimulating factor; PgR, progesterone receptor; HER2, human epidermal growth factor receptor 2; IQR, Interquartile Range; NSAI, non-steroidal aromatase inhibitor; SD, standard deviation.

[a] Race was self-reported; [b] Data was missing for remaining patients; [c] Percentage does not equal 100% as the result of rounding; [d] For one patient in the placebo+NSAI arm, HR status and HER2 status were missing. The patient was not treated; [e] Treatment-free interval was calculated only for patients with prior endocrine therapy.

3.1.4 Description and critique of the approach to validity assessment

Quality assessment of MONARCH 3 was undertaken by the company using NICE recommended criteria. A comparison of the company and ERG judgements for MONARCH 3 can be seen in Table 4.

Table 4 Company and ERG assessment of trial quality for MONARCH 3

NICE QA Criteria for RCT[a]	CS response	ERG response
1. Was the method used to generate random allocations adequate?	Low	Low
2. Was the allocation adequately concealed?	Low	Low
3. Were the groups similar at the outset of the study in terms of prognostic factors, e.g. severity of disease?	Low	Low (for most characteristics but not duration of disease or treatment-free interval, see section 3.1.1)
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)?	Low	Unclear: adequate blinding described but high frequency of adverse events such as diarrhoea in the ABE+NSAI arm could lead to unblinding.
5. Were there any unexpected imbalances in drop-outs between groups? If so, were they explained or adjusted for?	Low	High: [redacted]
6. Is there any evidence to suggest that the authors measured more outcomes than they reported?	Low	High: The European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-Breast 23 (EORTC QLQ-BR23) was measured in MONARCH 3, but this is not mentioned in the CS or trial publication (mentioned in the CSR).
7. Did the analysis include an intention to treat analysis? If so, was this appropriate and were appropriate methods used to account for missing data?	Low	Low

[a] Low = low risk of bias, high = high risk of bias, unclear = uncertain risk of bias.

The ERG agrees with most of the company's judgements for MONARCH 3, but notes that the higher frequency of adverse events such as diarrhoea in the ABE+NSAI arm could have led to unblinding of patients and care providers. This may potentially increase the risk of performance bias and detection bias (particularly affecting self-reported outcomes such as HRQoL). The reasons for discontinuation were not presented by trial arm in the CS; these were requested by the ERG and provided in clarification response A3. [redacted]

[redacted]. The ERG judged the MONARCH 3 trial to have a high risk of selective reporting bias, as the EORTC QLQ-BR23 trial was measured but not reported. The ERG obtained these results from the CSR.

3.1.5 Description and critique of company's outcome selection

The outcomes selected by the company are appropriate to the NICE scope and are commonly measured in a cancer trial. The details in the CS generally concur with those reported in the MONARCH 3 trial publication[13] and CSR except where stated below. The ERG consider that the outcomes appear to have been predefined.

The primary outcome of the MONARCH 3 trial was investigator-assessed PFS as defined by RECIST (RECIST: Response Evaluation Criteria in Solid Tumours) version 1.1.[15] PFS was measured from the date of randomisation to the date of objective progression or death due to any cause. A randomly selected subset of scans (number of scans not stated) was independently and blindly reviewed by a panel of radiologists at the interim analysis, and at the final analysis a full independent review of PFS was performed. The CS provides results for both investigator and independently reviewed PFS at both interim and final analysis, which the ERG considers appropriate.

Baseline tumour measurements (RECIST 1.1) were performed within 28 days of randomisation and then on Day 21–28 of every second cycle (approximately every eight weeks) between cycle 2 and cycle 18 and on day 21–28 of every third cycle (approximately every 12 weeks) after cycle 18, and then within 14 days of clinical progression. The finding of a new lesion was required to be unequivocal and not attributable to something other than a tumour. In the non-measurable, bone only disease cases, appearance of one or more new lesions (in bone or outside of bone), or unequivocal progression of existing bone lesions was required.

For those patients with locoregionally recurrent disease (around 3%) the CS states that in those in whom surgery was performed while on study with evidence of residual disease postoperatively, new baseline measurements should have been assessed. The CSR also describes that in [redacted]

- [redacted]

- [redacted]

The ERG asked for clarification (question A8) and the company response suggests that no participants had surgery while on study.

Response outcomes definitions as per RECIST 1.1 criteria were as follows:

- Complete response (CR), disappearance of all target lesions;
- Partial Response (PR), ≥30% reduction in the sum of diameters of target lesions (taking baseline sum diameters as the reference);
- Clinical benefit rate (CBR), the proportion of patients with CR, PR, or stable disease (SD) ≥6 months;
- Duration of Response (DoR), date of first evidence of CR or PR ([redacted]) to the date of objective progression or death due to any cause, whichever was earlier.

Expert clinical advice to the ERG confirms that clinical benefit rate (CR + PR + SD ≥6 months) is a clinically relevant outcome and used in practice.

The CS also reports the best overall response (BOR) which was categorised as CR, PR, SD, PD, or not evaluable except for those with bone-only, non-measurable disease, where it was limited to CR, SD, PD and not evaluable (partial response is not a criterion in non-measurable disease). SD was further classified as ≥ 6 months (best response of SD and PFS ≥ 6 months) or < 6 months.

Safety and patient reported outcomes (PROs) were evaluated on a safety population (defined as all patients who received at least one dose of study drug, 327 abemaciclib + NSAI vs 161 placebo + NSAI). The ERG considers that the ITT population should have been used for the analysis of PROs although number of patients in the two analysis sets were similar, see section 3.3.5 of this report.

The CS says that PROs of HRQoL were measured with European Organisation for Research and Treatment of Cancer Quality of Life Core 30 questionnaire (EORTC QLQ-C30) and EuroQol 5-Dimensions 5-Levels (EQ-5D-5L) administered at baseline, day 1 of every second cycle between cycle 3-19 and day 1 of every third cycle thereafter. Both measures are validated tools, they are briefly described in the CS (p 33) although the details of scoring and transformation of the data are only reported in the CSR. The CSR says that [redacted] This is a validated module of the EORTC QLQ-C30 for breast cancer.

Adverse events were classified according to the Common Terminology Criteria for Adverse Events (CTCAE) version 4 and further classified as treatment emergent or serious. Treatment emergent adverse events (TEAE) were defined as any adverse event that began between the first dose and 30 days after treatment discontinuation, or any pre-existing condition that increased CTCAE grade between the first dose and 30 days after treatment discontinuation (except there was no time limit on treatment emergent serious events). Serious Adverse Events (SAE) were defined as any adverse event that resulted in death; a life-threatening experience; persistent or significant disability/incapacity; initial or prolonged inpatient hospitalisation; congenital anomaly/birth defect; or were considered significant by the investigator for any other reason.

PFS, OS and some adverse events inform the economic analysis, see Section 4.3.4.1.

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

3.1.6.1 Sample size and power calculation

The MONARCH 3 RCT was a superiority trial which was powered for an interim analysis of the primary outcome, PFS, to be undertaken after approximately [redacted] investigator-assessed PFS events had been observed. The final PFS analysis was to be performed after [redacted] investigator-assessed PFS events had been observed. The statistical power calculation assumed a hazard ratio (HR) of 0.67 for ABE+NSAI vs placebo+NSAI, median PFS for placebo-NSAI of 10 months to yield > 80% power of the 1-sided log-rank test at a type 1 error of 0.025 (the HR of 0.67 amounted to approximately five months [50%] improvement in median PFS for the ABE+NSAI under an additional assumption of exponential survival distribution). The ERG considers that the power calculation was defined apriori, though the source of the assumptions was not stated.

The interim analysis of PFS (31st January 2017) was undertaken on the ITT population (ABE+NSAI n=328; placebo+NSAI n=165). At this time 164 patients (50.0%) in the ABE+NSAI arm and 98 patients (59.4%) in the placebo+NSAI arm had discontinued treatment. The final PFS analysis (3rd November 2017) was undertaken on the ITT population by which time [redacted] in the ABE+NSAI arm and [redacted] in the placebo+NSAI arm had discontinued treatment.

3.1.6.2 Analysis populations

Interim and final efficacy analyses were performed on the ITT population (n=493), which included all randomised patients, including two patients in the abemaciclib arm and three patients in the placebo arm who did not receive treatment. There were no exclusions from

the ITT analysis and missing data were not imputed [redacted] As stated earlier, the safety population was defined as all patients who received at least one dose of study drug, ABE+NSAI n=327 vs placebo+NSAI n= 161).

3.1.6.3 Statistical test methods

PFS was analysed with a one-sided stratified log-rank test with a type I error rate of 0.025, stratified by nature of disease (visceral metastases vs. bone-only metastases vs. other) and prior (neo)adjuvant endocrine therapy. Kaplan-Meier curves were used to estimate median PFS for each treatment arm; rates were compared at 4-month intervals using a normal approximation for the difference between rates. A stratified Cox proportional hazard model was used to calculate the hazard ratio between groups. In CS Appendix D.1.5 the assessment of proportional hazards (for all trials in the first-line treatment NMA) suggested that the assumption was reasonable for MONARCH 3 PFS data, although data were immature for OS (see section 3.1.7 for discussion of proportional hazards in the NMA).

Censoring occurred where it was not known if there had been progression or death at the time of analysis, with participants being censored at the last known progression-free assessment. Data were also censored if there was death or progressive disease after two or more missed tumour assessments; no baseline tumour assessment; or no post-baseline tumour assessment. The ERG asked the company to clarify the choice of censoring criteria used (clarification question A7). The company reported that there was no specific request from regulatory agencies regarding the censoring criteria for PFS in MONARCH 3. Censoring rules from the US FDA regulatory guidance were followed, there were no specific censoring criteria in the available European Medicines Agency (EMA) guidance. The ERG considers the use of the FDA guidance is reasonable.

The methods of statistical analyses for the other outcomes were not reported in the CS. [redacted] (provided in response to clarification question 10).

[redacted]

The MONARCH 3 trial publication[13] states that stratified tests using the Cochran-Mantel-Haenszel test were performed to compare response outcomes between treatment arms and that tests were performed at the two-sided 0.05 level and used 95% confidence intervals (CI)

unless stated. [redacted]

3.1.6.4 Subgroup analyses

The CS presents clinical effectiveness results for pre-planned subgroup analyses of PFS (for baseline stratification factors and other factors such as disease setting, see section 3.3.6), and post hoc exploratory subgroups on other factors associated with prognosis or endocrine therapy sensitivity. Subgroup analyses for OS were performed, but not presented in the CS, as the data are immature.

[redacted] In response to clarification question A6 the company states that the p-value for the interaction term was derived from a Cox model with the treatment arm, the subgroup variable and treatment by subgroup interaction term as factors. No adjustment for multiplicity in the subgroup analyses was performed (i.e. no correction was made to avoid erroneous inferences being made from multiple simultaneous statistical tests). Many of the pre-planned and exploratory subgroups had small sample sizes, particularly in the placebo group, and confidence intervals around the HRs are wide which need to be considered when interpreting their results. The NICE scope did not include any subgroups.

> **ERG conclusions:** the statistical approach of the MONARCH 3 trial reasonable. The power calculation for the primary outcome is appropriate; an ITT population was used for efficacy analyses; standard survival analysis methods were used, and both investigator and central independent assessment of PFS was undertaken. Caution is required in the interpretation of subgroup analyses as these are not statistically powered to show effects.

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

As only one trial of abemaciclib in this indication was included in the submission, MONARCH 3, a direct meta-analysis of abemaciclib trials was not possible. The CS provides a narrative review of the trial, with data presented in tables and text.

The CS reports two NMAs indirectly comparing abemaciclib with other treatments:

- The MONARCH 3 trial-aligned NMA is the main focus in the CS (hereafter referred to in this report as the '**first-line treatment NMA'**), as it presents comparative evidence of abemaciclib as a first-line treatment for advanced breast cancer (i.e. within the scope of this NICE appraisal). The results of this NMA inform the economic model: PFS informs the time to first progression estimate and OS informs the deaths before first progression estimate (described in further detail in section 4.3.4.2 of this report).
- A separate NMA is reported, the MONARCH 2 trial-aligned NMA (hereafter referred to in this report as the '**second-line treatment NMA'**), to provide comparative evidence of abemaciclib as a second-line treatment in advanced breast cancer. This NMA is aligned with the patient population of the MONARCH 2 RCT,[16] which compares abemaciclib plus fulvestrant vs placebo plus fulvestrant in HR+ HER2- advanced breast cancer patients who had progressed following (neo)adjuvant or first lie advanced breast cancer endocrine treatment. The second-line treatment NMA was necessary as the OS data from the MONARCH 3 trial are immature and the economic model therefore includes a PFS2 health state to estimate OS from abemaciclib indirectly via the effects of second-line and subsequent treatment lines (we explain this in more detail in section 4.3.3 of this report). This NMA is briefly reported in CS Appendix N. The company provided the ERG with a separate confidential report[17] describing it (and an accompanying SLR report[18]) as part of their response to clarification questions.

In the following sub-sections we provide a description and critique of the first-line treatment NMA (see also Appendix 9.1 for a quality assessment checklist of this NMA). We provide a separate description and critique of the second-line treatment NMA in Appendix 9.2 of this report.

3.1.7.1 First-line treatment NMA evidence networks

The CS reports the results of five separate NMA networks, for the outcomes PFS, OS, ORR, CBR and CR, respectively. Following a feasibility assessment (details not reported in the CS) the CS concluded that networks for grade 3-4 adverse events, treatment discontinuation and HRQoL were not possible due to limited available data in primary studies (CS Appendix D.1.3). Only PFS and OS outcomes are used to inform the economic model, therefore the ERG's critique focuses mostly on these two networks.

The inclusion criteria for the NMA are reported in CS Appendix D.1.2. These criteria are broader than the NICE scope, and permit inclusion of a range of comparator treatments

including endocrine therapies, chemotherapies, targeted therapies, and combinations of these. The CS states that these additional comparators were included in the NMA "to generate a fully connected network and to make optimal use of available data" (CS page 56). The ERG considers this is a reasonable decision as it enables more data to be included for the anastrozole/letrozole reference comparator (see below for more information on this). The inclusion criteria also differed from the NICE scope in relation to patient characteristics (HER2 and HR status, and previous treatment status). We discuss these below in section 3.1.7.4 in relation to clinical heterogeneity.

The company's SLR search identified potentially relevant studies for inclusion in the NMA (CS Appendix D.1.1. and D.1.2). A total of 20 trials met the inclusion criteria for the SLR, of which 18 were included in the NMA (2 of the 20 were excluded as they could not be connected to the network). The number of trials contributing data for the respective outcome measures (individual networks) varied according to trial data availability:

- PFS n=8
- OS n=15
- (PFS or OS n=17)
- ORR n=17
- CBR n=10
- CR n=15

The network diagrams for PFS and OS are reproduced from the CS in Figure 2 and Figure 3 respectively below. As can be seen, the NMA networks include the three scoped treatments (abemaciclib, palbociclib and ribociclib) plus additional treatments outside of the scope of this appraisal: anastrozole/letrozole monotherapy; exemestane 2.5mg, fulvestrant 250mg/ 500 mg; megestrol acetate 160 mg; tamoxifen 20mg/40mg and toremifene 60 mg or 200 mg. Hereafter we refer to these as the non-scoped treatments.

Data for the scoped treatments are provided by their respective pivotal RCTs:

- Abemaciclib + anastrozole/letrozole vs placebo + anastrozole/letrozole - MONARCH 3[13]
- Ribociclib + letrozole vs placebo + letrozole - MONALEESA-2[19]
- Palbociclib + letrozole vs letrozole - PALOMA-1/TRIO-18[20] (NB. This is a single trial)
- Palbociclib + letrozole vs placebo + letrozole - PALOMA-2[21]

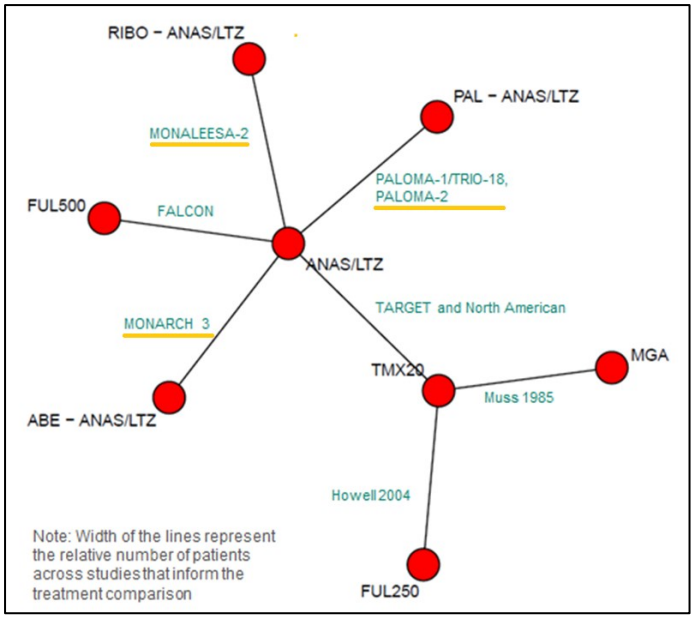
3.1.7.2 PFS network

The PFS network is a star-shaped network in which abemaciclib is connected to comparator treatments via a reference treatment, anastrozole/letrozole. There are no comparisons informed by both direct and indirect evidence in this network. The single abemaciclib trial included in the NMA, MONARCH 3, compared abemaciclib and anastrozole/letrozole with placebo and anastrozole/letrozole. This is the only trial in the network to have included both anastrozole and letrozole in a single trial arm. The other trials evaluated either anastrozole or letrozole as separate trial arms. To connect MONARCH 3 to the network anastrozole and letrozole were therefore pooled into one node and considered as one treatment arm. This is based on the assumption that the effectiveness of these two treatments is similar, and the CS notes that this assumption has been accepted in previous NICE appraisals in this indication (e.g. TA495[12] and TA496[11]). Clinical experts to the ERG in this current appraisal likewise agreed that they are equivalent in effectiveness.

The ERG notes that the reference treatment node in all the networks is anastrozole/letrozole monotherapy, however, in the MONARCH 3, MONALEESA-2 and PALOMA-2 trials the connecting comparator arm is placebo + NSAI (anastrozole/letrozole). This makes the assumption that the combination of placebo with NSAI is equivalent to NSAI alone. The CS does not discuss this assumption. The comparator arm in the PALOMA1/TRIO-18 trial was letrozole monotherapy (i.e. no placebo). The ERG considers the company's assumption to be acceptable for the purposes of connecting treatments in the networks.

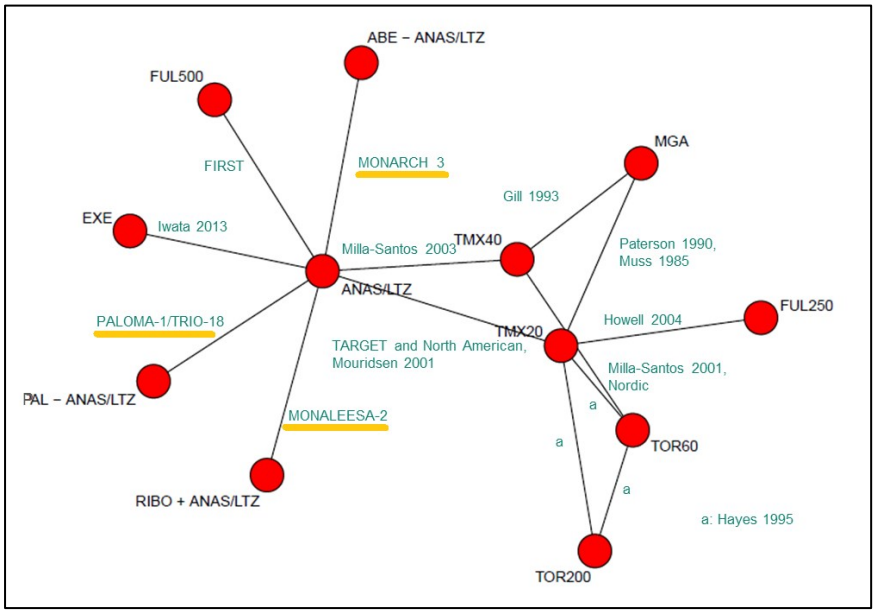
3.1.7.3 OS network

The OS network includes a larger number of treatments than the PFS network. Abemaciclib is connected to comparator treatments via an anastrozole/letrozole node (again, assuming equivalence in effectiveness of these two aromatase inhibitors). Comparisons between abemaciclib and palbociclib / ribociclib are only made indirectly, though other comparisons between non-scoped treatments are informed by both direct and indirect evidence, as illustrated by closed evidence loops. The ERG is not aware of any other studies of the scoped comparators that are eligible for inclusion.



Source: CS Appendix D.1.3 (Figure 3). The scoped treatment trials have been underlined by the ERG in yellow. ABE-ANAS/LTZ: abemaciclib plus anastrozole/letrozole; ANAS/LTZ: anastrozole/letrozole; FUL250: fulvestrant 250 mg; FUL500: fulvestrant 500 mg; TMX20: tamoxifen 20 mg; MGA: megestrol acetate 160 mg; PAL-ANAS/LTZ; palbociclib plus anastrozole/letrozole; RIBO-ANAS/LTZ: ribociclib plus anastrozole/letrozole

Figure 2 Network diagram for PFS, first-line treatment NMA network



Source: CS Appendix D.1.3 (Figure 4) The scoped treatment trials have been underlined by the ERG in yellow. ABE-ANAS/LTZ: abemaciclib plus anastrozole/letrozole; ANAS/LTZ: anastrozole/letrozole; EXE: exemestane; FUL250: fulvestrant 250 mg; FUL500: fulvestrant 500 mg; TMX20: tamoxifen 20 mg; TMX40: tamoxifen 40 mg; MGA: megestrol acetate 160 mg; PAL-ANAS/LTZ; palbociclib plus anastrozole/letrozole; RIBO-ANAS/LTZ: ribociclib plus anastrozole/letrozole; TOR60: toremifene 60 mg; TOR200: toremifene 200 mg.

Figure 3 Network diagram for OS, first-line treatment NMA network

3.1.7.4 Clinical heterogeneity assessment

The CS provides an assessment of clinical heterogeneity amongst the studies included in the NMA (CS Section B.2.9.3). This assessment is in relation to the MONARCH 3 trial and

the pivotal trials of the scoped comparators (MONALEESA-2; PALOMA-1/TRIO-18; PALOMA-2). In CS Appendix D.1.5 a heterogeneity assessment is provided for all the trials included in the NMA (i.e. the trials of the scoped and non-scoped treatments). Below we discuss clinical heterogeneity among the scoped treatment trials (n=4), then we discuss clinical heterogeneity among all 18 scoped and non-scoped treatment trials. We distinguish between the scoped and non-scoped treatments because the former are directly relevant to the decision problem and their results inform the economic evaluation. The latter are used to connect networks but are not directly relevant to the decision problem.

3.1.7.4.1 Heterogeneity assessment among scoped treatment trials

The scoped treatment trials are large, multi-centre, international, drug company-sponsored, double-blind, phase III trials, each containing several hundred patients. The exception is the PALOMA-1/TRIO-18 trial which was a smaller (n=165 patients) open-label phase II trial. [20]. The PALOMA and MONALEESA-2 trials were assessed in recent NICE appraisals of palbociclib (TA495)[12] and ribociclib (TA496),[11] respectively.

The CS considers that the trials of the scoped treatments are similar in terms of characteristics such as age, disease characteristics (e.g. cancer performance status; cancer stage; bone-only disease, menopausal status, HR+/HER2− status), and absence of previous endocrine therapy or chemotherapy in the advanced disease setting. The ERG has independently assessed these characteristics and agree that they are similar.

The CS highlights three areas where clinical heterogeneity was identified:

- **Required disease-free interval (DFI)** following adjuvant therapy. All trials enrolled patients with a DFI of over 12 months since adjuvant NSAI therapy. The MONARCH 3 trial also included patients with a DFI of over 12 months for (neo)adjuvant anti-oestrogen therapy. However, in the other trials the DFI for other hormonal therapies was not clear. The ERG assessed the proportion of patients at baseline in the trials within stated DFI categories. However, the trials report DFI according to different interval classes (e.g. < or ≥ 36 months;[13] ≤ or > 12 months [21]) making it difficult to compare trials.
- **Proportion of patients with visceral metastases**. This varied from 44% to 59% across treatment arms. The MONARCH 3 trial was towards the middle of this range (52%-54% of patients).
- **Site of disease**. Only the MONARCH 3 trial reported the proportion of patients with liver metastases (16%), although not by treatment arms and only reported as a post hoc subgroup analysis.

Expert clinical advice to the ERG is that these are key prognostic factors in this patient group. Visceral metastases confers a worse prognosis than bone metastases alone. The ERG considers that the difference between trials in visceral metastases in particular could potentially bias the results of the NMA, though it is not clear whether this would over or under-estimate the relative effectiveness of abemaciclib. The CS does not come to a firm conclusion about whether or not the above factors contribute to heterogeneity amongst the trials in the NMA, and what impact that might have on outcomes. The ERG asked the company to provide a discussion of the likely impact of any heterogeneity on the results of the NMA (clarification question A13). The company responded that the trials were homogenous for a large number of patient characteristics, with any differences 'anticipated to be minimal, thus lending reliability to the NMA results'. The company did not speculate what the impact of differences between the trials in visceral metastases might be.

Expert clinical advisors to the ERG are not aware of any additional key prognostic factors that should be noted in assessment of clinical heterogeneity.

The ERG also observes that the percentage of patients with newly diagnosed advanced/metastatic breast cancer varies between 34%[19] and 52%[20] across treatment arms in the trials. As noted earlier (section 2.1), this is a higher percentage than is commonly experienced in the UK (where rates of newly identified advanced/metastatic breast cancer are commonly in the range 10%-15%). This issue was discussed by the appraisal committee in NICE TA495[12] and clinical experts involved commented that they would not expect to see a difference in treatment effect for patients with newly diagnosed advanced/metastatic breast cancer. However, one clinical expert advising the ERG commented that patients who are newly diagnosed could be considered to have biologically different disease since this has never been exposed to hormonal therapy, whereas the remainder of the patients have remained disease free for at least 12 months after completing adjuvant hormonal therapy (most patients will have received two to five years of treatment).

3.1.7.4.2 Heterogeneity assessment among all trials including in the NMA (scoped and non-scoped treatments)

The reporting dates of the 18 trials vary from 1985 to 2017, reflecting the evolution of new treatments for advanced breast cancer over the decades (e.g. tamoxifen, anastrozole, letrozole, fulvestrant). Pivotal phase III and some phase II trials of these treatments have been included (CS Appendix Table 18). Given the long period of time over which the trials were conducted and published it is likely that delivery of clinical care and evaluation of

treatment effectiveness in the trials may have changed (e.g. introduction of testing for hormone receptors, broader use of a number of hormonal therapies, the use of bisphosphonates/denosumab and improvements in imaging/monitoring of patients to more clearly identify treatment benefit). This may be a source of heterogeneity in the network, though one of the expert clinical advisors to the ERG commented that changes to the supportive care environment would not significantly affect treatment effects.

CS Appendix Table 19 provides a very limited summary of patients' baseline characteristics across the 18 trials included in the NMA (age, performance status and menopausal status only). The CS states that the trials can be considered to be similar in terms of these characteristics. The ERG agrees with this assertion (however, see below).

The trials were also similar in other patient characteristics:

- The CS reports that none of the trials in the NMA included patients who had received prior endocrine therapy for advanced breast cancer.
- The CS also reports that all but two of the trials in the NMA omitted patients receiving prior chemotherapy for advanced breast cancer.

The ERG notes uncertainty about trial similarity for some patient characteristics:

- **Cancer performance status.** Whilst most of the patients in the trials had a favourable ECOG performance status (i.e. a PS <2) there was variability across the trials in the percentage of patients with PS 1, ranging between 15.4% to 57%. Seven trials did not report the performance status of patients. However, expert clinical advice to the ERG is that the difference between performance status of 0 and 1 is minimal and they can be grouped together for practical purposes.
- **HER2 status**. The inclusion criteria for the NMA specified that ≥80% of the trial study population should have HER2- breast cancer, but studies in which HER2 status was unknown were also eligible (HER2 testing was not routinely performed in older studies). The HER2 status of patients was unknown in 12 of the 18 trials in the NMA, and one trial permitted inclusion of HER2 +/- patients[22] (CS Appendix Table 24).
- **HR status.** All of the trials included in the NMA included women with HR+ breast cancer, though the CS does not report the percentage of women in each trial with HR+ breast cancer (the inclusion criteria for the NMA specified that trials in which ≥50% of women had HR+ breast cancer were eligible for inclusion). The company provided the percentage of women with HR+ breast cancer on request (clarification question A14) but only for the scoped treatment trials, not for all trials in the NMA as

was requested. For the scoped treatment trials the percentage of women with HR+ breast cancer (ER +) was [redacted]-100%, thus a high degree of similarity between trials.

The CS identifies areas of heterogeneity from consideration of the baseline characteristics of the trials (CS Appendix D.1.5). These include the same characteristics (prognostic factors) as identified for the scoped comparator studies discussed above: DFI; proportion of patients with visceral involvement and site of disease (e.g. liver metastases, bone only disease). The CS notes that there was incomplete reporting of these details in the studies, prohibiting a full assessment of clinical heterogeneity. Where details were reported there was variability between trials, such as DFI (reported in 6 of the 18 trials) where mean or median values ranged from 16 months (median) to 6.4 years (mean). The CS reports that meta-regression was not considered feasible due to the limited number of trials available. The ERG agrees with this as generally a minimum of 10 studies are required to perform meta-regression.[23]

The ERG asked the company to provide additional tabulated patient characteristics, including the proportion of patients with visceral involvement; liver metastases; DFI and prior therapy in the adjuvant setting (clarification question A14). The company provided these for the scoped treatment trials only, therefore the ERG is unable to make further comment on these characteristics in the non-scoped treatments in the NMA.

ERG conclusion

The ERG considers that the trials included in the NMA are similar in terms of patient characteristics such as age and previous treatment history for advanced breast cancer. However, due to reporting limitations in the trial publications a full assessment of clinical heterogeneity across the trials is not possible. The scoped treatment trials appear similar, though there was variation between them in the proportion of patients with visceral metastases. The impact of this on the results of the NMA are not clear. For this reason the results of the NMA should be interpreted with caution.

3.1.7.5 Critical appraisal of trials included in the first-line treatment NMA

CS Appendix Table 25 provides tabulated risk of bias assessments of all 18 included trials, using the NICE recommended criteria. The CS states that all trials were judged to be good quality with acceptable risk of bias (low or unclear risk of bias across the criteria). The CS reports that high risk of bias was mainly encountered with regard to blinding (of care providers, participants and outcome assessors) to treatment allocation “as several trials were open-label” (CS Appendix D.1.7). However, the ERG notes that only two of the trials were open-label (CS Appendix Table 18), 10 were double-blind, and in the remaining 6 trials

blinding was not reported. Thus, in half of the trials the risk of bias from lack of blinding was low and in the remaining half the risk of bias was mainly unclear; this is apparent in CS Appendix Table 25. It should be noted that outcomes such as OS are less prone to detection bias associated with lack of blinding than other outcomes such as PFS or tumour response, but performance bias (systematic differences in care) can occur.

The ERG did an independent critical appraisal of the scoped-comparator trials (Table 5). For the MONALEESA-2 trial,[19] the ERG largely agreed with the company's assessment, but, as with the MONARCH 3 trial, noted that adverse events may lead to unblinding. The ERG also considered there was evidence of selective outcome reporting.

For PALOMA-1/TRIO-18[20] and PALOMA-2,[21] the ERG gave more favourable assessments for randomisation and concealment of allocation, indicating a low risk of selection bias, but note a slightly higher proportion of patients with ECOG performance status 0 in the palbociclib + letrozole arm of PALOMA-2. In the PALOMA-1/TRIO18 trial,[20] an unplanned interim analysis was undertaken as almost twice as many patients in the control group of cohort 1 discontinued because of disease progression, therefore this trial is considered by the ERG to have a high risk of bias. The ERG considered that selective reporting bias was evident in PALOMA-1/TRIO-18[20] but not PALOMA-2,[21] and that ITT analysis was appropriate in PALOMA-2.

The ERG has not performed an independent critical appraisal of the non-scoped comparator trials included in the NMA. However, we note that the risk of bias as judged by the company was unclear in many trials for adequate randomisation, concealment of allocation, attrition, and use of ITT analysis. Thus, our conclusion is that the risk of bias in these trials is mostly uncertain.

3.1.7.6 Statistical NMA methods used

CS section B.2.9.1 and CS Appendix D.1.5 report details of the statistical methods used to conduct the NMA, citing methods described in NICE Decision Support Unit (DSU) Technical Support Documents 2, 3 and 4.[24-26] Binary outcomes (ORR, CBR, CR) were estimated using a logistic regression model using a binomial likelihood and a logit link function. For survival endpoints (i.e. PFS and OS) the CS cites a publication (itself cited in Technical Support Document 2) by Woods et al[27] which describes methods for NMA on the log-hazard scale combining count data (e.g. number of patients with an event at a point in time) and hazard ratio statistics (based on time to event data).

Table 5 Company and ERG assessment of trial quality for the NICE scoped comparator trials

NICE QA Criteria for RCT[a]		MONALEESA-2 [19] Ribociclib + letrozole vs letrozole	PALOMA-1/TRIO-18 Palbociclib + letrozole vs letrozole[20]	PALOMA-2 Palbociclib + letrozole vs letrozole[21]
1. Was randomisation carried out appropriately?	CS:	Low	Unclear	Unclear
	ERG:	Low	Low	Low
Comment: PALOMA-1/TRIO18: Interactive web-based randomisation system; PALOMA-2: Centralized internet/telephone registration system.				
2. Was concealment of treatment allocation adequate?	CS:	Low	Low	Unclear
	ERG:	Low	Low	Low
Comment: PALOMA-2: Centralized internet/telephone registration system.				
3. Were groups similar at outset in terms of prognostic factors?	CS:	Low	Unclear	Low
	ERG:	Low	Unclear	Unclear
Comment: PALOMA-1/TRIO18: slight imbalances in some characteristics; PALOMA-2: Slightly higher proportion with ECOG performance status 0 in palbociclib + letrozole group.				
4. Were care providers, participants and outcome assessors blind to treatment allocation?	CS:	Low	High	Low
	ERG:	Unclear	High	Low
Comment: MONALEESA-2: Adverse events may have led to unblinding.				
5. Were there any unexpected imbalances in drop-outs between groups?	CS:	Unclear	Unclear	Unclear
	ERG:	Unclear	High	Unclear
Comment: MONALEESA-2: Discontinuations due to progression were higher in the comparator group (therefore not unexpected); and discontinuations due to adverse events were higher in the ribociclib group. PALOMA-1/TRIO18: Publication states that an unplanned interim analysis was done as almost twice as many patients in the control group of cohort 1 discontinued because of disease progression. PALOMA-2: Discontinuations due to progression were higher in the comparator group (therefore not unexpected); and discontinuations due to adverse events were higher in the palbociclib group.				
6. Is there any evidence that authors measured more outcomes than reported?	CS:	Unclear	Low	Unclear
	ERG:	High	High	Low
Comment: MONALEESA-2: Most outcomes reported, but time to definitive deterioration of ECOG performance status in one category of the score not reported. PALOMA-1/TRIO18: Change from baseline in Modified Brief Pain Inventory in Pain Interference Scale (mBPI-sf) not reported.				
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	Low	Unclear
	ERG:	Low	Low	Low

[a] Low = low risk of bias, high = high risk of bias, unclear = uncertain risk of bias.

This method was designed to allow NMAs to include data from trials where the survival data is expressed in varying forms, thus potentially allowing a greater number of trials to be included in the same analysis. The CS doesn't explicitly state this as a rationale for using the Woods et al method.[27] The ERG considers that the use of Woods et al[27] method to be appropriate for the NMA of OS and PFS in this appraisal.

3.1.7.6.1 Proportional hazards assessment

An assessment of proportional hazards of survival data was conducted by the company. The company digitised Kaplan-Meier graphs for PFS and OS from the trials to estimate underlying individual patient data using a published algorithm.[28] The HR, median and proportion of patients event-free at a specific timepoint were checked against the published estimates to ensure internal validity. The CS did not report the results of this checking and the ERG requested this from the company (clarification question A20). The company provided a table describing discrepancies between the published and the digitised data. They report (unquantified) "small" discrepancies in HRs and CIs and median survival times in many of the trials. They also report quantified discrepancies (not described as "small") for some trials, including the scoped treatment trials. The company state that where there were discrepancies priority was given to the published data. The ERG considers this acceptable.

The CS reports that the proportional hazards assumption was tested by visual inspection of the Kaplan-Meier curves, log-cumulative hazard plots, Schoenfeld residual plots and the results of a weighted residual test based on standardised Schoenfeld residuals. The CS did not provide these curves, plots and the rest results, so the ERG requested these (clarification question A19). The company provided the requested information but did not provide any interpretation of them. The ERG's interpretation is as follows:

- The PFS log-log plots generally show initially overlapping lines which separate and become parallel over time (parallel lines indicates that the proportional hazard assumption is considered to hold). The OS log-log plots for mainly trials show overlapping lines over time.
- The Schoenfeld residual plots for both PFS and OS show variations over time in the residuals, illustrated by increasing and decreasing slopes in the curves between residual points. Horizontal shaped curves would indicate that hazards are proportional over time. The PFS curves are appear less variable over time than the OS curves suggesting that proportional hazards are more likely to hold.
- The results of the weighted residual tests for PFS showed no statistically significant trend between the residuals and time for any trials ($p>0.05$), indicating that the

proportional hazard assumption holds. For OS there were statistically significant trends ($p<0.05$) for four trials, including MONARCH 3.

The company's judgements on proportional hazards for PFS and OS in the 17 trials with OS or PFS data are presented in CS Appendix Table 23. Their judgements only appear to have been based on inspection of the Kaplan-Meier curves, rather than the other plots and statistical tests they provided (as discussed above).

- **PFS**: Kaplan-Meier data were unavailable for the whole trial population in nine trials and also for the HR+ subgroup of one trial. The proportional hazards assumption was accepted for the whole trial population in seven trials (including all of the pivotal scoped-comparator trials) and in the HR+ subgroup of one trial, and rejected in one trial (a non-scoped comparator). The ERG has visually inspected the Kaplan-Meier curves for the scoped-comparator trials and agree that proportional hazards appear to be supported.
- **OS**: Kaplan-Meier data were not available for three trials, and in a further eight trials the proportional hazards assumption was judged not to be supported. In the remaining six trials the assumption was supported.

The CS and the ERG have the following observations for the scoped-comparator trials:

- The ERG notes that in the MONARCH 3 trial the two [redacted] with final OS analysis to be done after 315 events). The CS states that due to immature survival data, conclusions regarding the proportional hazards assumption in this trial are uncertain. (NB. However, in CS Appendix Table 23 the proportional hazard assumption was accepted for the MONARCH 3 trial, despite noting that Kaplan-Meier curves cross after 20 months, immature survival and high level of censoring.)
- The CS notes that MONALEESA-2[19] trial OS data were immature at the time of analysis and that the Kaplan-Meier curves for the two treatments lie on top of each other until around month 24 when they begin to separate. The ERG also notes that updated results from the MONALEESA-2 trial were published in April 2018[29] (data cutoff 2nd January 2017) and that these show that the OS data are still immature (NB. these data are used in the NMA).

- For the PALOMA-2 trial[21] the CS states that no Kaplan-Meier OS data were available to inform assessment of proportional hazards. The trial publication states that data on OS were immature at the time of this analysis of the primary end point, and the final OS analysis will be performed when a total of 390 deaths occur. The ERG has not identified any updated OS results for this trial since this trial publication. This is a particular limitation for the indirect comparison between abemaciclib and palbociclib as the only OS data available for this comparison comes from a relatively small phase II open-label study (PALOMA-1/TRIO-18).[20]

The CS states that due to the immaturity of the data and the lack of a clinical rationale for explaining non-constant treatment effects over time between treatments they chose to conduct the NMA for OS based on an assumption of proportional hazards. They urge caution in the consideration of the results of the NMA due to the data immaturity. The CS states that there is no clinical rationale to justify a more complex NMA methodology assuming non-constant treatment effects over time between treatments. The ERG notes that company submissions for other NICE appraisals have used NMA methods such as fractional polynomials[30] for comparing treatments when proportional hazards are not supported or uncertain (e.g. TA463[31] and TA512[32]).

ERG conclusion:

It would have been appropriate for the company to have considered methods that incorporate time-varying hazards in the current appraisal as an alternative to the adopted methods. Nonetheless, the OS data from the MONARCH 3 and MONALEESA-2 trials would still be immature for this outcome and the NMA results – whichever approach was taken - would consequently be uncertain.

3.1.7.6.2 Outcome data used in NMA

CS Appendix tables 20 and 21 report the PFS and OS results (respectively) from the 18 trials (CS Table 22 reports the response rate results used in the NMA of response outcomes). Results for the ITT population and selected subgroups (e.g. patients with measurable disease) are tabulated. The ERG presumes that the results for the ITT population were used in the NMA, however, this is not explicitly stated in the CS. The ERG notes that the aforementioned assessment of proportional hazards (CS Appendix table 23) included both ITT populations and HR+ subgroup populations for two (non-scoped treatment) trials.

The ERG has checked the PFS and OS data in CS Appendix tables 20 and 21 for the scoped-comparator trials and note the company has used the most up to date data available where available.

The company also provided a summary of whether investigator or independent committee PFS assessments were available in the included trials (clarification response A13 Table 4). This was not reported by 10 trials and was reported as investigator-assessed by eight trials. The company state 'the heterogeneity of PFS assessments is not considered to have had a significant impact on the conclusions made'. The ERG notes that in the NICE appraisal of palbociclib (TA495[12]) the appraisal committee expressed a preference for blinded independent assessments of PFS, given that the higher rate of specific adverse events associated with palbociclib which may have unblinded some patients and investigators in the PALOMA trials. As discussed earlier in this report (section 3.1.4), there was a higher rate of diarrhoea in the abemaciclib arm of the MONARCH 3 trial potentially leading to unblinding. Although blinded independent committee PFS assessments are available in this trial these were not used in the NMA.

3.1.7.6.3 Bayesian modelling methods

Observed data were included in the model using a normal likelihood. The treatment effect model had a linear regression structure with the predicted log HR equal to the sum of the difference between the two treatment coefficients (CS Appendix D.1.5). The CS reports that a vague prior $\beta \sim N(0, 10^4)$ was to be used for the treatment effect coefficients. The CS does not provide a justification for the prior chosen and it is not stated whether choice of prior was explored by in sensitivity analysis. However, the ERG notes that vague priors are recommended by NICE DSU guidelines for treatment effect measures in NMAs.[24]

The ERG requested the company to provide more information about the Bayesian methods used to conduct the NMA (clarification question A15). The company provided the information requested. The ERG notes that the procedures reported for choosing initial parameter values and assessing convergence within and between chains as described are recommended by NICE DSU guidelines.[24]

The company reported that an assessment of the consistency of the direct and indirect evidence was performed in accordance with NICE DSU guidelines,[26] but did not provide further information on it in response to a clarification question (clarification question A16). The company's justification was that closed evidence loops containing both direct and

indirect evidence were only present for comparisons between non-scoped treatments in the networks. The ERG considers this justification reasonable.

As stated above, all treatments included in the NMA were compared to a reference treatment, anastrozole/letrozole monotherapy. The results are presented as pairwise comparisons between each treatment and the reference treatment (CS Figures 10-14). The ERG requested the company to provide NMA results for the indirect comparison of ABE+NSAI vs the comparators in the scope of the appraisal (i.e. palbociclib and ribociclib). The company provided these and they have been summarised later in this report (section 3.3.7).

OpenBUGS software (software package version 3.2.3) was used to conduct the analysis and the company provided the programming code on request from the ERG (clarification question A18).

3.1.7.6.4 Model fitting

The choice between a random effects and a fixed effect model was informed by the Deviance Information Criterion (DIC). The DIC is commonly used to compare the fit of Bayesian statistical models, whereby the model with the smallest DIC is estimated to be the model that would best predict a replicate dataset which has the same structure as that currently observed.[33] The company provided the DIC values upon request (clarification question A17).

The CS presents random effects NMA for all but one outcome measure. For the PFS outcome a fixed effect model was presented. Random effects models are appropriate when it is suspected that included trials may be heterogeneous. The ERG therefore regards use of random effects models to be more appropriate for this set of trials. The ERG requested the results for both random and fixed effect models for all outcomes, to permit comparison of their results (clarification question A17). The company supplied the random effects PFS results only. The ERG notes that these results provide similar point estimates to the fixed effect results, though wider credible intervals are generated by the random effects model (as would be expected) and they now cross the null line showing no statistically significant effects for ABE+NSAI and each of the scoped comparator treatments (see section 3.3.7 of this report for the results). The ERG also notes that the random effects PFS credible intervals are very wide, but in comparison, the width of the random effects OS credible

intervals are of a much smaller magnitude (they are more in-keeping with the PFS fixed effects credible intervals). There is no explanation given for this inconsistency.

3.1.7.7 Summary of the ERG's appraisal of the first-line treatment NMA

The ERG considers that, overall, the NMA has been adequately conducted. Standard Bayesian methods have been used, as recommended by the NICE Decision Support Unit. The pivotal trials of the scoped treatments have been included, and the ERG regards these to be generally at low risk of selection bias but may be at risk of other biases. The ERG is not aware of any relevant trials that have been omitted from the NMA.

However, there are some limitations and uncertainties:

- For many trials it was not possible to ascertain similarity, or otherwise, of patient characteristics. Notably, there is variation between trials in the proportion of patients with visceral metastases, and the effect of this on the results is uncertain.
- The NMA method used assumes the proportional hazards assumption holds for survival outcomes. However, this assumption could not be supported by available data for some trials. Amongst the scoped-treatment trials proportional hazards appeared to hold for PFS, but not for OS, where OS data are currently immature. The CS concludes that there is no clinical rationale to justify using an NMA approach that assumes non-constant treatment effects. However, the ERG considers that an alternative approach assuming time-varying hazards would have been informative (albeit with immature OS data).
- Due to the immaturity of the OS data in the scoped treatment trials the ERG considers the results of the OS NMA to be highly uncertain.

Although there were limitations to the NMA, the results were considered by clinical experts advising the ERG to be clinically plausible (we summarise these results later in section 3.3.7 of this report).

Finally, the ERG notes that recent NICE appraisals of treatments in this indication (palbociclib TA495 and ribociclib TA496) did not include an NMA. Therefore, no comparison of the methods and results of the NMAs in the current appraisal with previous NMAs has been possible.

3.2 Summary statement of company's approach to systematic review

Table 6 provides a critical appraisal of the company's SLR of clinical effectiveness.

Table 6 Quality assessment (CRD criteria) of CS review

CRD Quality Item; score Yes/No/Uncertain with comments	ERG comments
1. Are any inclusion/exclusion criteria reported relating to the primary studies which address the review question?	**Yes** The eligibility criteria were set apriori. The eligibility criteria were used to identify trials of relevance to the decision problem, including trials for the NMA. Eligibility of potential trials for the NMA was wider than the NICE scope, including a number of other potential treatment options. The eligibility criteria suggest that all studies were required to have abemaciclib as the intervention and the other potential treatment options were listed only as comparators. However, the SLR included studies of the other potential treatments as interventions.
2. Is there evidence of a substantial effort to search for all relevant research?	**Yes** The searches were of sound methodology, well documented and reproducible with an acceptable range of databases searched. As such the ERG did not consider it necessary to replicate the main searches. An update for the searches of ongoing studies was completed.
3. Is the validity of included studies adequately assessed?	**Yes** The studies were adequately assessed although the ERG differs in assessment on more than one risk of bias criterion (selective reporting bias, blinding and drop outs for the pivotal RCT). Risk of bias was assessed for all of the studies included in the NMA.
4. Is sufficient detail of the individual studies presented?	**Yes**
5. Are the primary studies summarised appropriately?	**Partly** The CS omits some of the pre-specified outcomes but these were available in the CSR.

CRD = Centre for Reviews and Dissemination

The company's evidence synthesis is clearly reported and presents the key information that the ERG would expect to see. It is unlikely there is any error in the inclusion of studies from the SLR and the ERG does not consider that any key trials are likely to have been missed. The NMA included all studies of possible relevance in this population group, which were broader than those specified in the NICE scope. The ERG considers this to be appropriate.

The review processes reported in CS Appendix D.1.2 appear appropriate. Two reviewers independently assessed studies for inclusion through a two-stage process. One reviewer extracted data into a piloted data extraction worksheet and a second reviewer checked extractions. Excluded studies with reasons were reported and a PRISMA style flow chart. It is unclear whether one or two reviewers assessed each study for risk of bias, however, the ERG considers that it is unlikely that the CS have introduced biases from the processes used for the SLR.

3.3 Summary of submitted evidence

In the following sub-sections we summarise the results of the MONARCH 3 trial.

3.3.1 Progression-free survival

The CS provides interim and final efficacy analyses for both investigator assessed (primary outcome) and Independent Central Review assessed PFS; the final analyses only are summarised in Table 7. At a median follow-up of [redacted] months, investigator assessed median PFS was [redacted] months in the abemaciclib + NSAI group compared with [redacted] in the placebo + NSAI group; HR [redacted] (95% CI [redacted], 2-sided [redacted]), giving a reduction in the risk of progression of disease or death of 46%. PFS survival rate at 24 months was [redacted] vs [redacted] respectively. Outcomes by Independent Central Review were slightly more favourable than investigator assessment in both treatment arms (Table 7).

3.3.2 Overall survival

Overall survival data were immature at the final data cut, with median survival [redacted], HR [redacted] (95% CI [redacted] 2-sided stratified log-rank [redacted]), Table 7. [redacted] overall survival rate at 24 months (abemaciclib + NSAI [redacted] vs placebo + NSAI [redacted]

Table 7 Survival at final analysis

	Abemaciclib + NSAI (n=328)	Placebo + NSAI (n=165)	Treatment Effect / Difference /p-value
Progression-free survival			
Median PFS, months Investigator assessed	[redacted]	[redacted]	[redacted]
Median PFS, months Independent Central Review	[redacted]	[redacted]	[redacted]
24 month PFS rate, % Investigator assessed	[redacted]	[redacted]	[redacted]
24 month PFS rate, % Independent Central Review[a]	[redacted]	[redacted]	[redacted]
Overall survival			
Median OS, months	[redacted]	[redacted]	[redacted]
24 month OS rate, % (95% CI)	[redacted]	[redacted]	[redacted]

[a] Source: CSR addendum
CI, confidence interval; NSAI, non-steroidal aromatase inhibitor; OS, overall survival; PFS, progression free survival.

3.3.3 Response rates

The objective response rate (RECIST 1.1 complete response or partial response) by investigator assessment was [redacted] with abemaciclib +

NSAI compared with placebo + NSAI ([redacted]) (Table 8). [redacted]

[redacted]

8[redacted]

[redacted]

3.3.4 Duration of response

Among patients with an investigator assessed response (abemaciclib + NSAI n=163, placebo + NSAI n=61), the median duration of response was [redacted] months (95% CI [redacted]) in the abemaciclib + NSAI arm compared with [redacted] months (95% CI [redacted]) in the placebo + NSAI arm (Table 8).

Table 8 Best overall response and duration of response (investigator assessment)

	Abemaciclib + NSAI N=328		Placebo plus NSAI N=165		OR	p-value
	n (%)	95% CI	n (%)	95% CI		
Objective response rate, CR + PR	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Disease control rate, CR + PR + SD	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Clinical benefit rate, CR + PR + SD ≥6 months	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
CR	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
PR	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
SD	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
SD ≥6 months	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
PD	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Not evaluable	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	Months	**95% CI**	**Months**	**95% CI**		
Duration of response	[redacted]	[redacted]	[redacted]	[redacted]	-	-

CI, confidence interval; CR, complete response; NA, CS states 'the computations were not done because there were fewer than 2 non-missing levels in the data'; NSAI: non-steroidal aromatase inhibitor; PD, progressive disease; PR, partial response; SD, stable disease.

3.3.5 Health related quality of life

The CS states that patient completion rates for HRQoL instruments were high, apart from in cycle 22. In response to clarification question A4, the company provided further details on

the completion rates for each arm during this cycle, including reasons for non-completion, and noted that the low rate reported in the CS was for one arm (placebo + NSAI) for one of the three instruments (EQ-5D scale). The ERG notes that the completion rates for each instrument were lower in the placebo group, but the reasons for this are not clear. HRQoL measures were analysed on the safety population set (without imputation of missing data), rather than the ITT analysis set.

3.3.5.1 EORTC QLQ-C30

Global health status

[redacted]

[redacted] in the CS. The ERG notes that there is no category for a 'large' difference (unequivocal clinical relevance) for this symptom.[34]

[redacted]

3.3.5.2 EQ-5D-5L

There was [redacted] in change from baseline in the EQ-5D-5L index score or visual analogue scale (Table 9).

3.3.5.3 EORTC QLQ-BR23

Compared with placebo + NSAI,

[redacted] with

abemaciclib + NSAI. There were no significant differences in between group changes on the other function and symptom scales (Table 9).

Table 9 HRQoL outcomes change from baseline[a] (safety population)

LS Mean (SE)	Abemaciclib + NSAI (n=327)	Placebo + NSAI (n=161)	Between group change difference[a]	p-value
EORTC QLQ-C30				
Global health status	[redacted]	[redacted]	[redacted]	[redacted]
Functional scales (higher score = better)				
Physical functioning	[redacted]	[redacted]	[redacted]	[redacted]
Role functioning	[redacted]	[redacted]	[redacted]	[redacted]
Emotional functioning	[redacted]	[redacted]	[redacted]	[redacted]
Cognitive functioning	[redacted]	[redacted]	[redacted]	[redacted]
Social functioning	[redacted]	[redacted]	[redacted]	[redacted]
Symptom scale items (higher score = worse)				
Fatigue	[redacted]	[redacted]	[redacted]	[redacted]
Nausea and vomiting	[redacted]	[redacted]	[redacted]	[redacted]
Pain	[redacted]	[redacted]	[redacted]	[redacted]
Dyspnoea	[redacted]	[redacted]	[redacted]	[redacted]
Insomnia	[redacted]	[redacted]	[redacted]	[redacted]
Appetite loss	[redacted]	[redacted]	[redacted]	[redacted]
Constipation	[redacted]	[redacted]	[redacted]	[redacted]
Diarrhoea	[redacted]	[redacted]	[redacted]	[redacted]
Financial difficulties	[redacted]	[redacted]	[redacted]	[redacted]
EQ-5D-5L				
Index value	[redacted]	[redacted]	[redacted]	[redacted]
Visual analogue scale	[redacted]	[redacted]	[redacted]	[redacted]
EORTC QLQ-BR23[b]				
Functional scales (higher score = better)				
Body image	[redacted]	[redacted]	[redacted]	[redacted]
Sexual functioning	[redacted]	[redacted]	[redacted]	[redacted]
Future perspective	[redacted]	[redacted]	[redacted]	[redacted]
Symptom scale items (higher score = worse)				
Systemic therapy side effects	[redacted]	[redacted]	[redacted]	[redacted]
Breast symptoms	[redacted]	[redacted]	[redacted]	[redacted]
Arm symptoms	[redacted]	[redacted]	[redacted]	[redacted]

See CS Table 12 p.53 and CS Table 13 p.54 for baseline values.

EORTC QLQ-BR23: European Organisation for Research and Treatment of Cancer Quality of Life Breast cancer; EORTC QLQ-C30: European Organisation for Research and Treatment of Cancer Quality of Life Core 30 questionnaire; EQ5D-5L: EuroQol 5-Dimensions 5-Levels; LS: least squares. [a] Across all post-baseline visits. [b] From CSR addendum.

3.3.6 Sub-group analyses results

3.3.6.1 Progression-free survival

Pre-planned subgroup analyses for PFS were undertaken for the following subgroups (see CS Appendix E Figure 9):

- Baseline stratification factors
    - Site of metastases (visceral metastases, bone-only metastases, other)
    - Prior (neo)adjuvant endocrine therapy (aromatase inhibitor, other, no prior endocrine therapy)
- Other subgroups:
    - NSAI received at Cycle 1 (letrozole, anastrozole) *(note this is missing from CS Appendix E Figure 9)*
    - Disease setting (de novo metastatic, recurrent metastatic, locoregionally recurrent) *(note that locoregionally recurrent was not a category in CS Appendix E Figure 9)*
    - Measurable disease at baseline (yes, no)
    - Number of organs involved (1, 2, 3+) *(note this is missing from CS Appendix E Figure 9)*
    - Age (<65 years, ≥65 years)
    - Region (North America, Europe, Asia)
    - Race (Caucasian, Asian, and other) *(note this is missing from CS Appendix E Figure 9)*
    - Progesterone receptor status (positive, negative)
    - Baseline ECOG PS (0, 1)

In addition, the CS describes additional exploratory subgroup analyses on factors associated with prognosis and/or sensitivity to endocrine therapy; these are not described as pre-planned (see CS Appendix E Figure 10):

- Disease diagnosis (<10 years, ≥10 years, de novo metastatic)
- Tumour grade (high-grade tumour, low/intermediate grade, unknown)
- Disease free interval (de novo metastatic, <3 years, ≥3 years, recurrent with no adjuvant chemotherapy)

- Presence of liver metastasis at baseline (yes, no)

The direction of the treatment effect was consistent across all of the above subgroups (favouring ABE+NSAI). However, there were two statistically significant interactions (the following data are from the CSR, and reflect the final PFS assessment in the ITT population): race ([redacted]) and geographic region ([redacted]). A greater treatment effect was found in Asian patients ([redacted]) compared with Caucasian patients ([redacted]), and in Asia ([redacted]) compared with Europe ([redacted]) and North America ([redacted]).

The CS explains this difference as being driven by the reduced performance of the control arm in Asian patients (particularly in Taiwan and Korea), in which there were a larger proportion of patients with a poorer prognosis at baseline.

No further details are provided in the CS, however the CSR states that [redacted] The ERG notes that the interaction p-values were not adjusted for multiplicity testing (clarification question A6).

3.3.6.2 Overall survival

Subgroup analyses were not presented for OS due to immature data; the ERG consider this to be reasonable.

3.3.6.3 Response rate

The CS provided response rates for the subgroup with measurable disease at baseline (n=397, 81%), these are summarised in Table 10.

Table 10 Response rates in subgroup with measurable disease at baseline

	Abemaciclib + NSAI N=267		Placebo plus NSAI N=130		OR	p-value
	n (%)	95% CI	n (%)	95% CI		
Objective response rate, CR + PR	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Disease control rate, CR + PR + SD	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Clinical benefit rate, CR + PR + SD ≥6 months	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

[a] Source CSR addendum.
CR, complete response; PR, partial response; SD, stable disease; OR= Odds ratio

The ORR and CBR, but not the DCR, were significantly higher in the abemaciclib + NSAI arm compared with placebo + NSAI in this subgroup. As non-measurable disease cannot have a best response of partial response, these outcomes cannot be assessed for the subgroup with non-measurable bone-only disease.

3.3.7 Network meta-analysis results

The treatment effects of ABE+NSAI and each of the scoped comparators relative to placebo+NSAI are summarised in Table 11. Effects for the other (non-scoped) treatments included in the NMA can be seen in CS Figures 10 to 14. The CS did not present indirect comparisons between the scoped treatments; these were requested by the ERG and were provided in clarification question response (A12) for PFS and OS only (Table 12).

For PFS using the fixed effects model, all three treatments showed similar and statistically significant HRs improving PFS relative to placebo+NSAI (Table 11). Using the random effects model (provided in response to clarification question A17) resulted in similar point estimates but much wider credible intervals, and statistically nonsignificant differences relative to placebo+NSAI for each of the three treatments. There were no significant differences for the indirect comparisons between ABE+NSAI and PAL+NSAI or ABE+NSAI and RIBO+NSAI using either fixed or random effects models (Table 11).

There were no statistically significant differences in OS for any of the three treatments relative to NSAI (Table 11). However, OS data are currently immature in the trials therefore these results are uncertain. Similarly, there were no significant differences in OS for the indirect comparisons between ABE+NSAI and PAL+NSAI or ABE+NSAI and RIBO+NSAI using either fixed or random effects models (Table 12).

There were also no statistically significant differences in ORR, CBR or complete response for any of the three treatments relative to NSAI (Table 11). The estimate for abemaciclib + NSAI complete response was highly uncertain due to low event counts.

Table 11 Summary of treatment effects relative to placebo+NSAI for the scoped treatments

Outcome, number of studies in NMA	Abemaciclib + NSAI	Palbociclib + NSAI	Ribociclib + NSAI
PFS, FE 8 studies, HR (95% CrI)	[redacted]	[redacted]	[redacted]
PFS, RE 8 studies, HR (95% CrI)[a]	[redacted]	[redacted]	[redacted]
OS, RE 15 studies, HR (95% CrI)	[redacted]	[redacted]	[redacted]
ORR, RE 17 studies, OR (95% CrI)	[redacted]	[redacted]	[redacted]
CBR, RE 10 studies, OR (95% CrI)	[redacted]	[redacted]	[redacted]
CR, RE 15 studies, OR (95% CrI)	[redacted]	[redacted]	[redacted]

FE = Fixed effects model; RE = Random effects model; HR = Hazard ratio; OR = Odds ratio
[a] clarification response A17.

Table 12 Treatment effects for ABE+NSAI vs PAL+NSAI and RIBO+NSAI for PFS and OS

Comparator	HR (95% CrI) (*fixed* effects model)	HR (95% CrI) (*random* effects model)
PFS		
PAL+NSAI	[redacted]	[redacted]
RIBO+NSAI	[redacted]	[redacted]
OS		
PAL+NSAI	[redacted]	[redacted]
RIBO+NSAI	[redacted]	[redacted]

Source: clarification question response A12

3.3.8 Summary of adverse events

Adverse events were reported for the safety population, which was all patients who received at least one dose of study drug (327 abemaciclib + NSAI; 161 placebo + NSAI), at the final analysis. Summary treatment emergent adverse events (TEAEs) are described in Table 13.

Table 13 Summary rates of key treatment emergent adverse events (safety population)

Percent of participants[a]	Abemaciclib + NSAI (n=327)	Placebo + NSAI (n=161)
Patients with ≥1 TEAE	[redacted]	[redacted]
TEAEs related to study treatment [b]	[redacted]	[redacted]
Patients with ≥1 Grade 3 or higher TEAE	[redacted]	[redacted]
Grade 3 or higher TEAE related to study treatment [b]	[redacted]	[redacted]
Patients with ≥1 serious adverse event	[redacted]	[redacted]
Serious adverse events related to study treatment [b]	[redacted]	[redacted]
Discontinuations of all study treatment due to an AE	[redacted]	3.1

Discontinuations of study treatment due to a SAE	[redacted]	[redacted]
Deaths due to adverse event	[redacted]	[redacted]

[a] Patients may be counted in >1 category. [b] Includes events that were considered related to study treatment as judged by the investigator.

Rates appeared to differ between groups for:

- Proportions with at least one TEAE related to treatment as judged by the investigator ([redacted] abemaciclib + NSAI vs [redacted] placebo + NSAI);
- Proportions with grade ≥3 TEAEs (abemaciclib + NSAI arm [redacted] vs placebo + NSAI arm [redacted], with [redacted] and [redacted] considered related to study treatment as judged by the investigator, respectively);
- Proportions with at least one serious adverse event (SAE) (abemaciclib + NSAI arm [redacted] vs placebo + NSAI arm [redacted]);
- Serious adverse events considered related to study treatment as judged by the investigator (abemaciclib + NSAI group [redacted] placebo + NSAI group [redacted];
- Discontinuations of all study treatments (abemaciclib plus NSAI arm [redacted] vs placebo plus NSAI arm 3.1%).

The CS provides details of TEAEs (grades 1-4 and all grades) occurring in at least 15% of participants in CS Table 16 (CS p69), not reproduced here. All TEAS, with the exception of arthralgia and back pain, occurred more frequently in the abemaciclib + NSAI arm. At any grade, diarrhoea ([redacted]), infections/infestations ([redacted]), neutropenia ([redacted]), fatigue ([redacted]) and nausea ([redacted]) were the most frequently experienced TEAEs in the abemaciclib plus NSAI arm. Infections/infestations ([redacted]), fatigue ([redacted]), diarrhoea ([redacted]), nausea ([redacted]) and arthralgia ([redacted]) were the most frequently experienced TEAEs of any grade in the placebo plus NSAI arm. At grade 3 or higher, the most commonly experienced TEAEs in the abemaciclib + NSAI arm were neutropenia ([redacted] grade 3 / [redacted] grade 4); diarrhoea ([redacted] grade 3 / [redacted] grade 4, see below for more details of diarrhoea); leukopenia ([redacted] grade 3 / [redacted] grade 4); infections and infestations ([redacted] grade 3 / [redacted] grade 4) and anaemia ([redacted] grade 3 / [redacted] grade 4), Table Y. Rates of grade 3 or 4 TEAEs in the placebo + NSAI arm were low; there were no events that were reported more commonly than others, see Table 14 for those most commonly reported in the abemaciclib + NSAI arm.

Specific TEAEs related to study treatment were not reported in the CS but were identified in the CSR addendum from the final analysis. Any grade diarrhoea made up the majority of these events in both the abemaciclib + NSAI arm ([redacted]) and the placebo + NSAI arm

([redacted]); the majority of which were grade 1 or 2. Rates of other TEAEs related to study treatment that were commonly experienced included any grade neutropenia ([redacted] and [redacted] in the abemaciclib + NSAI arm and placebo + NSAI groups respectively, with [redacted] of ≥grade 3 in the former group) fatigue ([redacted] and [redacted] in the abemaciclib + NSAI arm and placebo + NSAI groups respectively, mostly < grade 3) and nausea ([redacted] and [redacted] in the abemaciclib + NSAI arm and placebo + NSAI groups respectively, mostly < grade 3).

Grade 3 or higher rates of anaemia, ALT and AST increase, diarrhoea, hypertension, leukopenia, neutropenia and lymphopenia are used in the economic model.

3.3.8.1 Diarrhoea

Diarrhoea was more common in the abemaciclib plus NSAI group than the placebo group (CS Table 16). The majority of diarrhoea was grade 1 and 2 ([redacted] and [redacted] respectively), see above for rates at grade 3 and 4. The CS says that the median onset of diarrhoea was 8.0 days and the median duration was 10.5 days for grade 2 and 8.0 days for grade 3. In the abemaciclib plus NSAI arm 76.3% did not undergo any treatment modifications due to diarrhoea; [redacted] had a dose reduction and [redacted] had a dose omission. [redacted] discontinued treatment due to diarrhoea. The ERG clinical experts confirmed that abemaciclib is associated with diarrhoea and that this is worse in the first few weeks and it then settles down. Antidiarrhoeal medications were used in [redacted].

3.3.8.2 Serious adverse events

Rates of participants experiencing at least one SAE were higher in the abemaciclib + NSAI group ([redacted]) than the placebo + NSAI group ([redacted]) (Table 13). Specific SAEs by treatment group are presented for those occurring in at least 1% of participants in CS Table 17 (p 70); rates of all events were higher in the abemaciclib + NSAI arm. Lung infection [redacted], embolism ([redacted]), anaemia ([redacted]), diarrhoea ([redacted]) and acute kidney injury ([redacted]) were the most commonly reported SAEs in the abemaciclib + NSAI group, and dehydration ([redacted]), abdominal pain ([redacted]) and vomiting [redacted]) were most common in the placebo + NSAI group.

The CS concludes that abemaciclib + NSAI was well tolerated with an acceptable TEAE profile. Expert clinical advice to the ERG agrees with this conclusion, though it was noted that the relatively high proportion of patients receiving abemaciclib reporting diarrhoea ([redacted]) is clinically important.

3.3.8.3 Comparator treatment summary adverse events

The CS did not present adverse events for the scoped comparators. The ERG has summarised the key events here from their pivotal phase III RCTs, for information (Table 14).

Table 14 Grade 3 or higher adverse events reported in the four included trials (most commonly experienced in the MONARCH 3 RCT abemaciclib arm)

AE, %	MONARCH 3		MONALEESA-2		PALOMA 1 and 2	
	Abemaciclib + NSAI	Placebo + NSAI	Ribociclib + letrozole	Placebo + letrozole	Palbociclib + letrozole	Placebo + letrozole
Neutropenia	[redacted]	[redacted]	59.3[a]	0.9[a]	1: 54.2[a] 2: 66.4[a]	1: 1 2: 1.4[a]
Diarrhoea	[redacted]	[redacted]	1.2[a]	0.9[a]	1: 4.0[a] 2: 1.4[a]	1: 0 2: 1.4[a]
Leukopenia	[redacted]	[redacted]	21.0[a]	0.6[a]	1: 19[a] 2: 24.8[a]	1: 0 2: 0
Infections + infestations	[redacted]	[redacted]	4.2[ab]	2.4[ab]	1: NR 2: NR	1: NR 2: NR
Anaemia	[redacted]	[redacted]	1.2[a]	1.2[a]	1: 6.0[a] 2: 5.4[a]	1: 1.0[a] 2: 1.8[a]

[a] Calculated by ERG; [b] Reported as 'infections'.

3.3.8.3.1 Ribociclib

In the MONALEESA-2 trial 98.5% of patients in the ribociclib + letrozole arm and 97.0% of patients in the placebo + letrozole group experienced an adverse event.[19] The proportions experiencing any grade 3 or higher event was higher in the ribociclib + letrozole group than the placebo group (81.1% vs 32.7%). The most commonly reported adverse event was neutropenia, with ≥ grade 3 neutropenia experienced in 59.3% and 0.9% in the two groups respectively. Other commonly reported adverse events at grade 3 or higher included leukopenia (21.0% and 0.6%, respectively) and hypertension (9.9% and 10.9%). As an adverse event of interest in the current appraisal diarrhoea at any grade was experienced in 35% in the ribociclib + letrozole group and 22.1% in the placebo group. SAEs were experienced in 21.3% in the ribociclib group and 11.8% in the placebo group. Rates of discontinuation of treatment due to adverse events was 7.5% in the ribociclib + letrozole group and 2.1% in the placebo + letrozole group.

3.3.8.3.2 Palbociclib

In the two RCTs of palbociclib + letrozole the proportions experiencing any adverse events were similar; in PALOMA-1/TRIO-18[20] 100% in the palbociclib + letrozole arm and 84.4% in the placebo + letrozole arm; in PALOMA-2[21] 98.9% in the palbociclib + letrozole arm and 95.5% in the placebo + letrozole arm. The most common adverse events in the palbociclib + letrozole groups of each trial were neutropenia, leukopenia, and fatigue. Diarrhoea was

experienced in 20.5% of participants in the palbociclib + letrozole group and 10% in the placebo + letrozole group in the PALOMA-1/TRIO-18.[20] In the PALOMA 2 trial[21] rates were 26.1% and 19.4% in the two groups respectively. SAEs were experienced in 19.6% of participants in the palbociclib + letrozole group and in 12.6% of participants in the placebo + letrozole group of PALOMA-2. Rates of discontinuation owing to TEAEs were 13% in the palbociclib + letrozole group and 2% in the placebo + letrozole group in PALOMA-1/TRIO-18. In PALOMA-2, discontinuation of any study treatment due to adverse events occurred in 9.7% in and 5.9%, respectively.

3.3.9 Ongoing studies

The company states that there are currently five ongoing studies in the UK investigating the efficacy and safety of abemaciclib in breast cancer patients (CS Section B.2.11). One of these is the MONARCH 3 trial, as follow-up for overall survival is still ongoing. The other four studies are not relevant to this appraisal. An update search for ongoing trials was undertaken by the ERG (restricted to trials of abemaciclib currently), which did not identify any additional ongoing studies with relevant comparators.

4 COST EFFECTIVENESS

4.1 Overview of company's economic evaluation

The company's submission to NICE includes:

i) a review of published economic evaluations of treatment options for the management of HR+/HER2- advanced breast cancer (CS section B.3.1).

ii) a report of an economic evaluation undertaken for the NICE STA process. The cost effectiveness of ABE+NSAI is compared with RIBO+NSAI and PAL+NSAI for untreated HR+/HER2- advanced breast cancer (CS section B.3.2).

4.2 Company's review of published economic evaluations

The company report a systematic literature review conducted to identify cost-effectiveness evidence relevant to treatment options for the management of HR+/ HER2- locally advanced or metastatic breast cancer. The methods of systematic review and results are reported in CS Appendix G and summary information about included cost-effectiveness studies relevant to the UK setting is presented in CS Table 18 (B.3.1). This included seven NICE technology appraisals (TA214; TA239; TA250; TA263; TA295; TA421; TA423), one paper (Das et al. 2013)[35] and an abstract (Polyani et al. 2014)[36], none of which related to comparators in the current appraisal. Three of the non-UK publications related to scoped-comparators: Bhattacharya (2016); Mamiya (2017) and CADTH (2016), all of which on palbociclib. However, none of these papers reported useful information about model input parameters that would add to the existing information in NICE TA495.

4.3 Critical appraisal of the company's submitted economic evaluation

4.3.1 NICE reference case

The ERG considers that the company's economic evaluation meets NICE's reference case requirements (Table 15).

Table 15 NICE reference case requirements

NICE reference case requirements:	Included in submission	ERG comment
Decision problem: As per the scope developed by NICE	Yes	However, population is restricted to postmenopausal women
Comparator: As listed in the scope developed by NICE	Yes	Palbociclib or ribociclib with an aromatase inhibitor (letrozole)
Perspective on costs: NHS and PSS	Yes	
Evidence on resource use and costs: Costs should relate to NHS and PSS resources and should be valued using the prices relevant to the NHS and PSS	Yes	
Perspective on outcomes: All direct health effects, whether for patients or, when relevant, carers	Yes	
Type of economic evaluation: Cost utility analysis with fully incremental analysis	Yes	
Synthesis of evidence on outcomes: Based on a systematic review	Yes	
Time horizon: Long enough to reflect all important differences in costs or outcomes between the technologies being compared	Yes	
Measuring and valuing health effects: Health effect should be expressed in QALYs. The EQ-5D is the preferred measure of HRQoL.	Yes	
Source of data for measurement of HRQoL: Reported directly by patients and/or carers.	Yes	Yes for PFS1, but PFS2 and PPD use general public valuations [37]
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	

4.3.2 Decision problem

4.3.2.1 Population

While the NICE scope considers a broad population of people with advanced HR+/HER2- breast cancer, the decision problem addressed by the company is narrowed to address postmenopausal women with advanced HR+/HER2− locoregionally recurrent or metastatic breast cancer who have had no prior systemic therapy for advanced disease. No patient subgroups are included in the NICE scope of the CS.

The modelled cohort is women of 65 years and above. To estimate drug doses for intravenous treatments, a body surface area (BSA) of 1.70 m^2 were assumed. Given that BSA data were not collected directly from the MONARCH 3 trial, height and body weight were used to estimate BSA. An average weight of 67.99kg and a height of 158.41cm were used for this estimation.

4.3.2.2 Interventions and comparators

The comparators in the model are palbociclib or ribociclib with an aromatase inhibitor, which are currently licensed for use in the UK NHS and correspond to the NICE scope.

The first-line NMA and economic model treat the NSAIs letrozole and anastrozole as a single class (i.e. similar in efficacy and safety). This reflects conclusions in previous NICE appraisals that in clinical practice AIs are considered to be equivalent, with similar effectiveness and acquisition costs (NICE TA495 and TA496).

In the previous NICE appraisals TA495 and TA496, the committees also considered NSAI monotherapy as a comparator for ribociclib + NSAI and palbociclib + NSAI. However, NSAI monotherapy is not specified as a comparator in the scope for this current appraisal. The company includes NSAI as a reference treatment in the first-line NMA and in the economic model. We therefore report input parameters and results for NSAI to provide context for the included comparators.

4.3.3 Model structure and assumptions

The company describes the model structure and key characteristics in CS section B.3.2.2. Monthly cycles are used to reflect state-transitions and the accrual of costs and outcomes. With the exception of treatment costs, which are incurred at the beginning of each month, a half-cycle correction is incorporated. Costs and QALYs are discounted at an annual rate of 3.5%. The model uses a 35-year time horizon, so could be said to reflect a lifetime since survival approaches 0.1% for all arms in base case by the end of this time period.

The company's illustration of the model is reproduced in Figure 4 below.

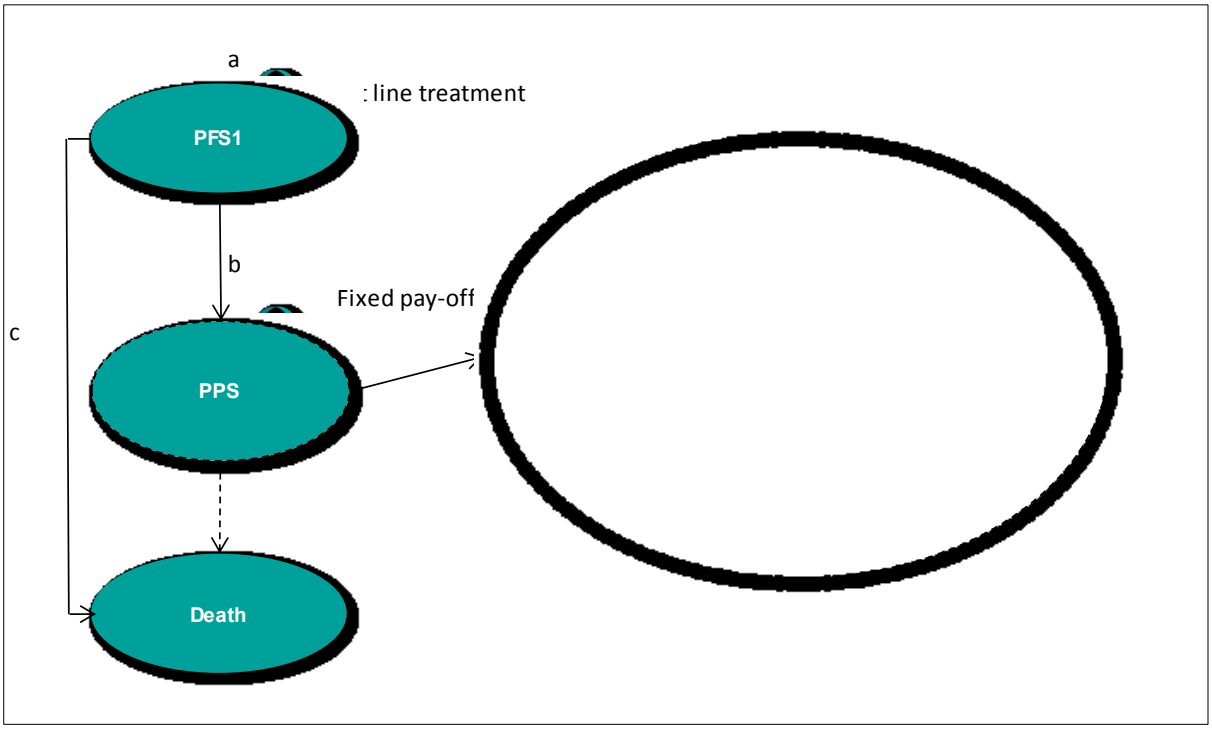


Figure 4 Illustration of model structure (CS Figure 15)

Abbreviations: PFS1: first-line progression-free survival; PFS2: second-line progression-free survival; PPS: post-progression survival

The model can be thought of as encapsulating a main model starting from first-line treatment and a 'fixed pay-off' sub-model starting from second-line treatment.

First-line model: This comprises three states; progression free survival (PFS1), post-progression survival (PPS) and death. A cohort of patients enters the model in the PFS1 health state at the start of first-line treatment with one of the included comparators (ABE+NSAI, PAL+NSAI or RIBO+NSAI) or NSAI. Patients may then:

a. Remain progression free.
b. Experience disease progression. Time to progression from first-line treatment (TTP1) is estimated as a survival curve, but unlike conventional progression-free survival, death is treated as a censoring event in the calculation of TTP1.
c. Die before disease progression. The progression-free death rate (PFD1) is conditional on the patient not having progressed. Unlike OS, progression is treated as a censoring event in the calculation of PFD1.

Methods used to estimate TTP1 and PFD1 are discussed in section 4.3.4.2 below.

When patients experience a first disease progression a 'fixed pay-off' is applied, representing health outcomes and costs that are incurred while patients receive second-line treatment and subsequent treatment and care. This pay-off is calculated in a separate sub-model (the dashed circle in Figure 4).

Fixed pay-off sub-model: This accounts for treatment and outcomes after the first disease progression. It is a conventional three-state partitioned-survival model, with transition probabilities calculated from:

- Overall survival from the start of second-line treatment (OS2). This includes deaths that occur before and after progression.
- Progression-free survival from the start of second-line treatment (PFS2). This includes deaths that occur before progression as events. For logical consistency, PFS2 is constrained in the model to be no more than OS2.
- The proportion of PFS2 events that are deaths is used to separate probabilities of progression, pre-progression deaths and post-progression deaths. This proportion is estimated from two other survival curves: time to progression and progression-free death from the start of second-line treatment (TTP2 and PFD2), defined and estimated in the same way as TTP1 and PFD2.

Methods used to estimate the post-progression transition probabilities using OS2, PFS2, TTP1 and PFD2 are discussed in section 4.3.4.3.

Transition probabilities and costs in the fixed-pay-off model are weighted according to the proportions of patients assumed to start each of the included second-line treatments. The

model includes costs for a third line of treatment (within the PPS state), but outcomes related to third-line treatment are not modelled explicitly.

The treatment pathway illustrated below shows the classes of treatment offered at first, second and third-line.

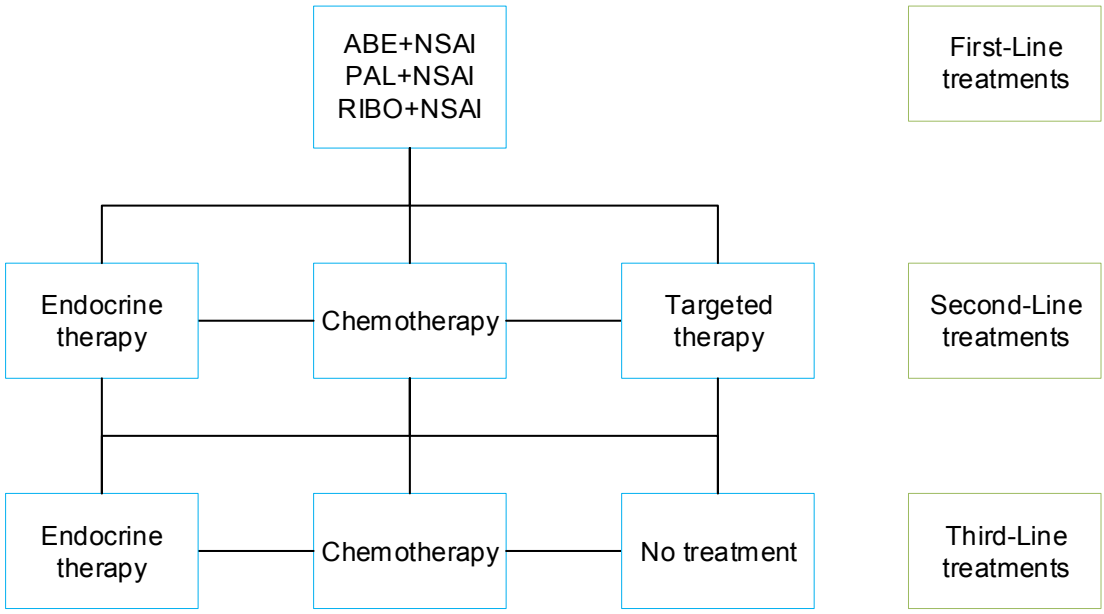


Figure 5 Treatment pathway (adapted from CS Figure 17)

Patients are assumed to stop first-line treatment when their disease progresses, or earlier if, for example, they experience intolerable adverse effects. The time to discontinuation of first-line treatment (TTD1) is estimated from trial data but constrained so that it cannot exceed TTP1. Similarly, time to discontinuation of second-line treatment (TTD2) cannot exceed PFS2. Time on third line treatment is estimated as a fixed proportion of time spent in the PPS state in the fixed pay-off sub-model.

The company's assumptions about initiation, utilisation and discontinuation of drugs are discussed in section 4.3.6.2 below.

4.3.3.1 Appropriateness of model structure and assumptions

Earlier models for NICE technology appraisals in breast cancer, including the palbociclib appraisal (NICE TA495), have taken the conventional three-state (PFS, OS and death) partitioned survival approach.[38] The ribociclib appraisal (NICE TA496) and this current company submission explicitly model a second-line of treatment and time to second progression (PFS2) using multi-state modelling. This approach is motivated as a way of

reducing uncertainty over immature first-line overall survival data. The CS cites immaturity of the MONARCH 3 OS data as the main reason for adopting this approach. The ribociclib model was an individual-level simulation. In this current appraisal, the company applies similar principles but implemented in a cohort model. They note that building a strictly Markov state-transition model would require a 'memoryless' assumption, where the probability of death would be the same for every individual after a first progression, regardless of how long they had spent in the PFS1 state. To overcome this problem, a 'compartmental' approach is used to keep track of successive cohorts of patients entering the fixed pay-off sub-model.

Calibration is used to adjust the time spent in the pay-off sub-model to reflect an assumed relationship between PFS and OS. In the base case, the company assumes a 'partial surrogacy' relationship, with the gain in OS being a fixed proportion of the gain in PFS (27.5%). This follows the approach in the ribociclib appraisal (TA496) and the DSU report for that appraisal.[39, 40] Without calibration, the model would automatically assume a direct gain in OS equal to the gain in PFS (100% surrogacy), which is not justified due to the uncertain and immature OS data from MONARCH 3. See section 4.3.4.3.4 below for a description of the source of the 27.5% surrogacy assumption and of how partial surrogacy is implemented.

ERG conclusions: The model structure and assumptions are appropriate. Given the immaturity of overall survival data for abemaciclib and for the comparators ribociclib and palbociclib, a conventional partitioned-survival approach would be subject to high uncertainty. The 'PFS2' fixed pay-off sub-model incorporates additional information about the effectiveness of second-line treatments. It also has the advantage that different patterns of second-line treatment use can be explored, modifying both the costs and outcomes. This is important as current UK practice differs from that in the RCTs on which the model is based. In addition, the calibration enables manipulation and exploration of uncertainty over the relationship between PFS and OS.

4.3.4 Clinical effectiveness

4.3.4.1 Overview of clinical parameters

The company summarise sources for transition probability estimates in CS Table 20 (CS section B.3.3.2) and base case values in CS Table 57 (B.3.6.1). Further detail is provided in CS sections 3.3.2 through to 3.3.7.

4.3.4.2 First-line transition probabilities

Table 16 Base case transition probabilities: first-line treatment

		Treatment	Base case	Source
TTP1	Time to first progression	NSAI	[redacted]	Exponential survival estimated from MONARCH 3 data adjusted for interval-censoring but not for patient baseline characteristics (CS Figure 21 B.3.3.5)
		ABE+NSAI	[redacted]	
		ABE+NSAI[c]	[redacted]	PFS hazard ratios compared with NSAI from first-line NMA, fixed effects (CS Table 23 B.3.3.5)[a]
		PAL+NSAI	[redacted]	
		RIBO+NSAI	[redacted]	
PFD1	Death rate before first progression	NSAI	0.002 per month	Negative-binomial regression of MONARCH 3 data not adjusted for patient baseline characteristics (CS Table 24 B.3.3.5)
		ABE+NSAI	0.005 per month	
		ABE+NSAI[c]	[redacted][b]	OS hazard ratios compared with NSAI from first-line NMA (CS Figure 11 B.2.9.2)
		PAL+NSAI	[redacted]	
		RIBO+NSAI	[redacted]	

[a] HRs for TTP1 in the model (as cited in CS Table 23) differ from those reported in CS Figure 10 B.2.9.2. These differences are small and we test the impact in ERG scenario analysis.

[b] The HR for ABE+NSAI ([redacted]) implies a pre-progression death rate only slightly higher than that for NSAI. The reason for this discrepancy is unclear.

c Not used in company base case (included here for reference).

4.3.4.2.1 Time to first progression (TTP1)

The company used individual patient data from the MONARCH 3 trial to estimate time to first progression for abemaciclib + NSAI and for NSAI. This analysis was conducted on the final PFS dataset (3rd November 2017) for the ITT population, with investigator assessment of progression.

MONARCH 3 analysis: investigator vs independent assessment

The company state that they use investigator-assessed progression as this is the primary endpoint in MONARCH 3 and is consistent with most trials included in the first-line NMA,

whereas independently-assessed progression was reported for fewer trials. However, there are arguments in favour of independent assessment of progression. Concerns were raised about the robustness of investigator assessment in NICE TA495 and TA496 due to the potential for un-blinding caused by the higher incidence of haematological adverse events with palbociclib and ribociclib than with NSAI (TA495 paragraph 4.3 and TA496 3.4). There are similar issues for MONARCH 3 trial because of the higher incidence of diarrhoea in the abemaciclib arm.

> **ERG conclusion**: We prefer independent assessment of progression outcomes due to the potential for loss of blinding caused by imbalances in adverse events. However, we acknowledge the importance of aligning outcomes in the NMA and the absence of independently-assessed outcomes for some trials. We therefore use investigator-assessed outcomes in the ERG preferred analysis. This is conservative because PFS is less favourable with investigator assessment than with independent assessment (CS B.2.6.1 Figures 6 and 7).

MONARCH 3 analysis: Adjustment for interval censoring

Tumour assessment in MONARCH 3 was conducted periodically (every other cycle up to cycle 18, then every third cycle and within 14 days of clinical progression) (CS Table 5). This explains the 'stepped' appearance of the Kaplan-Meier curves for PFS (CS Figures 6 and 7). In reality, progression will have occurred between assessments, thus recorded time to progression will tend to overestimate true time to progression. The company adjust for this using interval-censoring (CS B.3.3.4): the timing of progression events is coded as an interval between the preceding tumour assessment and the assessment at which the progression was recorded. The company use interval censored (IC) TTP1 estimates in their base case analysis and present a scenario without interval-censoring.

> **ERG conclusion**: We agree with the company's use of IC adjustment to estimate time to progression and use this in our preferred analysis. Like the company, we run a scenario without IC adjustment to test its impact on cost-effectiveness results.

MONARCH 3 analysis: Adjustment for baseline characteristics

The company do not adjust for baseline patient characteristics in their base case estimate of TTP1 from MONARCH 3, but they include baseline covariates in a scenario analysis. The covariates that were considered for inclusion are listed in CS section B.3.3.3. They include

variables for pre-planned subgroup analyses and additional prognostic factors identified by a literature review and discussion with experts. The company state that they performed backward and forward stepwise procedures to covariate selection but favoured the backward approach as this tends to include fewer variables. The variables included in the final covariate-adjusted equation were: age, liver metastases, measurable disease at baseline, PgR receptor status and tumour grade. In addition, a treatment group indicator was included as an explanatory variable.

> **ERG conclusion**: We agree with the company's approach to adjustment for baseline characteristics. The methods used to select covariates are reasonable and important prognostic variables are included. However, we support the use of unadjusted estimates in the base case, as this is more conservative.

MONARCH 3 analysis: Parametric survival functions

The company fitted parametric models for TTP1 in MONARCH 3, including a treatment indicator to provide joint estimates for abemaciclib + NSAI and NSAI. They tested three proportional hazards (PH) models (exponential, Weibull and Gompertz) and three accelerated failure time models (log-normal, log-logistic and gamma). Interval-censored adjusted curves are shown in CS Figures 19 and 20 and unadjusted curves are in CS Figures 25 and 26, M.2.2. CS Appendix M also includes supporting evidence for their choice of curves, sections M.1.1, M.2.1 and M.2.2.

The company concludes:

- Exponential and Weibull provide the best fit based on AIC (Akaike information criterion) and BIC (Bayesian inference criteria) statistics;
- In addition, Gompertz and gamma appear to fit the observed data well;
- Log-normal and log-logistic appear to overestimate survival after about 30 months;
- Proportional hazards models are compatible with the use of hazard ratios to estimate treatment effects for the comparators, whereas the accelerated failure time models are not.

For their base case, the company use an exponential function for time to first progression, with Weibull and Gompertz scenarios. Parameter values for these distributions are given in CS Tables 61, 62 and 62 for interval-censored, interval-censored and covariate adjusted and unadjusted models respectively (CS M.2.1). The CS and model do not include parameters

for the fitted log-normal, log-logistic or gamma distributions, and the company did not provide these parameters in response to a clarification question.

ERG considerations on the choice of parametric functions for TTP1:

- Cumulative hazards and log-log plots support the assumption of proportional hazards (CS Appendix M.1.1 Figures 15 and 16). We accept the company's argument that PH functions should be preferred because they are compatible with the use of hazard ratios to estimate results for the comparator treatments.
- For the interval-censored model, AIC and BIC statistics suggest that the exponential curve has the best fit to the trial data, followed by Weibull and Gompertz models (Table 60 M.2.1).
- Visual inspection of the curves (see Figure 8 in Appendix 9.3 of this report) indicates that all functions except log-normal have a good fit to the abemaciclib + NSAI arm. The fit is less good for the NSAI arm, particularly for the log-normal and log-logistic.
- Table 17 below shows estimated proportions of patients whose disease has not progressed within 1, 2, 3, 5 and 10 years of initiation of first-line treatments. Results are broadly consistent across the parametric functions, with the exception of the log-normal and log-logistic, which predict fewer late progressions. Clinical advisors to the ERG have suggested that 1% to 4% survival without progression at 10 years is more realistic than 9%: indicating that the exponential extrapolation may be appropriate.

We note an error in the coding of Gompertz TTP1 interval-censored adjusted survival in the company's submitted model. The formulae incorrectly reference the shape parameter for the baseline covariate model. There is also an error in reporting of the shape parameter for the Weibull curve in Table 61 in Appendix M (CS M.2.1). However, the value in the model (0.951) seems correct as the resulting curve fits the Kaplan-Meier curve and matches that in CS Figures 19 and 20.

ERG conclusion: The exponential, Weibull and Gompertz estimates of time to first progression provide a good fit to MONARCH 3 trial results. On the basis of statistical fit and clinical judgement on long-term extrapolations, we agree with the use of exponential as a base case, with Weibull and Gompertz as scenarios.

Due to a coding error, the company scenario with Gompertz TTP1 is not reliable. We present corrected results Section 4.4.1 below.

Table 17 Proportion of cohort without progression: (interval-censored adjusted)

	Year	Kaplan-Meier	Exp.	Weibull	Gompertz	Gamma	Log-normal	Log-logistic
NSAI (MONARCH 3)	0	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	3		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	5		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	10		[redacted]	[redacted]	[redacted]			
ABE+NSAI (MONARCH 3)	0	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	5		[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	10		[redacted]	[redacted]	[redacted]			
ABE+NSAI (HR vs. NSAI)	0		[redacted]	[redacted]	[redacted]			
	1		[redacted]	[redacted]	[redacted]			
	2		[redacted]	[redacted]	[redacted]			
	3		[redacted]	[redacted]	[redacted]			
	5		[redacted]	[redacted]	[redacted]			
	10		[redacted]	[redacted]	[redacted]			
PAL+NSAI HR vs. NSAI)	0		[redacted]	[redacted]	[redacted]			
	1		[redacted]	[redacted]	[redacted]			
	2		[redacted]	[redacted]	[redacted]			
	3		[redacted]	[redacted]	[redacted]			
	5		[redacted]	[redacted]	[redacted]			
	10		[redacted]	[redacted]	[redacted]			
RIBO+NSAI HR vs. NSAI)	0		[redacted]	[redacted]	[redacted]			
	1		[redacted]	[redacted]	[redacted]			
	2		[redacted]	[redacted]	[redacted]			
	3		[redacted]	[redacted]	[redacted]			
	5		[redacted]	[redacted]	[redacted]			
	10		[redacted]	[redacted]	[redacted]			

Source: Company model with log-normal, log-logistic and gamma distributions digitised from CS Figures 19 and 20

PFS hazard ratios from first-line NMA

TTP1 for ribociclib + NSAI and palbociclib + NSAI are estimated using PFS hazard ratios relative to NSAI estimated from the first-line NMA. This entails the assumption of equal relative treatment effects for PFS and TTP. We consider this a reasonable approximation given the rarity of pre-progression death (21 out of 493 patients in MONARCH 3): the difference between PFS and TTP lies in how deaths before progression are analysed (an event in PFS but censored in TTP).

There are small differences between the PFS hazard ratios used in the model (CS Table 23 section B.3.3.5) and the fixed effect results reported in CS section B.2.9.2 Figure 10 – see Table 18 below. We also show results from the random effects model, as reported in response to a clarification question (A17). Although the random effects model converged, the credible intervals were implausibly wide. We therefore focus on the fixed effects model.

Table 18 PFS hazard ratios reported in CS: first-line NMA

Comparator	Median hazard ratio (95% credible interval)		
	CS Table 23 B.3.3.5 (as in model)	Fixed effects CS Figure 10 (B.2.9.2)	Random effects (clarification question A17 response Figure 2)
Abemaciclib + NSAI	[redacted]	[redacted]	[redacted]
Palbociclib + NSAI	[redacted]	[redacted]	[redacted]
Ribociclib + NSAI	[redacted]	[redacted]	[redacted]
NSAI	Reference	Reference	Reference

Source: CS Table 23 and CS Figure 3

Base case estimates of first-line TTP for the comparators are shown in CS Figure 21. We reproduce this graph adding a curve for abemaciclib estimated relative to NSAI using the PFS HR (see Figure 9 in Appendix 9.3 of this report). This shows that the base case estimate of TTP1 for abemaciclib from MONARCH 3 data is more favourable than the NMA estimate relative to NSAI, calculated in the same way as the other comparators. The reasons for this difference are unclear given that the only data for abemaciclib in the first-line NMA comes from MONARCH 3. Possible explanations include: the use of median HRs from the NMA but means for regression coefficients from MONARCH 3; and differences in relative treatment effects for TTP and PFS. The company conducted a scenario analysis with the NMA-based estimate of PFS1 for abemaciclib as well as for the other comparators. This made abemaciclib relatively less cost-effective.

ERG conclusion: The first-line NMA indicated that the three treatments have very similar effects on extending PFS compared with NSAI. Therefore, the large difference in time to first progression for comparators in the company's base case is questionable. This occurs because different estimation methods are used for ABE+NSAI (regression analysis of MONARCH 3 data) and for the PAL+NSAI and RIBO+NSAI (hazard ratios relative to NSAI from NMA). For a more reliable comparison, we use NMA-based estimates of TTP1 for all comparators relative to NSAI in ERG preferred analysis.

Uncertainty over the relative effects of the three comparators on PFS is not properly reflected in the company's probabilistic sensitivity analysis, because the HRs are sampled independently, not accounting for correlations between NMA results. We conduct additional deterministic sensitivity analysis to investigate the impact of uncertainty over the PFS HRs.

4.3.4.2.2 Deaths before first progression (PFD1)

Estimation of PFD1 from MONARCH 3

Pre-progression death rates for abemaciclib+NSAI and for NSAI were estimated from MONARCH 3 data. As few deaths before progression were observed - 17 deaths out of 328 patients in the intervention arm and 4 out of 165 patients in the control arm - the company used a negative binomial regression rather than a parametric survival model. They included follow-up time as an exposure variable and a treatment indicator. Forward stepwise regression identified ECOG status, prior adjuvant endocrine therapy and the type of NSAI received as co-variates. However, the company chose to use the simpler model without covariates for their base case. Parameters for the models with and without covariates are shown in CS Tables 65 and 66, M.2.3.

OS hazard ratios from first-line NMA

Hazard ratios for OS were used to estimate PFD1 rates for ribociclib and palbociclib: OS random effects model, see Table 16 above and CS Figure 11 B.2.9.2. This entails the assumption of equal treatment effects for OS and the rate of deaths before progression. However, the OS HRs were incorrectly applied relative to the progression-free death rate for abemaciclib, rather than for NSAI. This can be seen in CS Figure 22, as the survival rate is

shown to be highest for NSAI despite hazard ratios below 1 for ribociclib and palbociclib. We show a corrected version of this graph in Figure 9 (Appendix 9.3 of this report).

We also note that the rate of pre-progression deaths for abemaciclib estimated from the MONARCH 3 negative binomial regression is very different to that estimated using a HR relative to NSAI: [redacted] and [redacted] respectively. The reason for this difference is unclear, though we note that the regression approach uses mean coefficient values whereas the NMA approach uses median HRs. The assumption that relative treatment effects are the same for pre-progression deaths as for overall survival may also be wrong. Whatever the correct value for abemaciclib, we are concerned that the use of a different estimation method for the other comparators may bias relative estimates of cost-effectiveness.

ERG conclusion: We agree with the company's approach to estimating pre-progression death rates from MONARCH 3 data: the constant hazard estimated by negative binomial regression and omission of covariates in the base case is appropriate given the rarity of this event. However, the estimated death rates for palbociclib and ribociclib are higher than they should be because the model applies hazard ratios to the wrong comparator. We correct this in ERG analysis.

As was the case for TTP1, different methods are used to estimate PFD1 for abemaciclib for the comparators. The pre-progression death rate for abemaciclib and is considerably higher in the base case (estimated directly from MONARCH 3 data) than when it is estimated in the same way as the other comparators (with HRs relative to NSAI). We highlighted uncertainty and limitations in the first-line OS NMA (section 3.1.7 above). Nevertheless, we believe that the first-line NMA still provides the best available foundation for comparisons between abemaciclib, ribociclib and palbociclib. We therefore use first-line NMA-based estimates relative to NSAI for all three comparators in ERG preferred analysis.

4.3.4.3 Post-progression transition probabilities

Table 19 reports the transition probabilities used to estimate the effects of second-line treatment.

Table 19 Post-progression transition probabilities (before calibration)

		Treatment	Base case [a]		Source
Distribution of second-line treatments (used to weight costs and transitions)		FUL	10.9%		ERG scenario in NICE fulvestrant appraisal (TA503). With additional assumption that NSAI is not repeated at second-line. (CS Table 35 B.3.5.1)
		ANAS	-		
		LTZ	-		
		EXE	37.0%		
		TMX	18.5%		
		EVE+EXE	8%		
		Chemo	25.7%		
PFS2	Progression free survival from start of second-line treatment	FUL	[redacted] per month		MONARCH 2 subgroup parametric survival regression, exponential (CS Figure 29 B.3.3.6)
		ANAS	[redacted]	[redacted]	PFS HRs relative to fulvestrant (from model)[b] estimated from second-line NMA (see Appendix 9.2 below)
		LTZ	[redacted]	[redacted]	
		EXE	[redacted]	[redacted]	
		EVE+EXE	[redacted]	[redacted]	
		TMX	[redacted]	[redacted]	Estimated from Milla-Santos 2001[41]
		Chemo	1.64	(0.85, 3.15)	HR vs. EVE+EXE, Li et al. 2015 [42]
OS2	Overall survival from start of second-line treatment	FUL	[redacted] per month		MONARCH 2 parametric survival regression, exponential
			[redacted] per month		CONFIRM hazard after maximum MONARCH 2 follow-up (27.95 months)
		ANAS	[redacted]	[redacted]	OS hazard ratios relative to fulvestrant (from model) [b], estimated in second-line NMA (see Appendix 9.2 below)
		LTZ	[redacted]	[redacted]	
		EXE	[redacted]	[redacted]	
		EVE+EXE	[redacted]	[redacted]	
		TMX	[redacted]	[redacted]	Milla-Santos 2001[41] and second-line NMA
		Chemo	1.89	(0.72, 5.00)	HR vs. EVE+EXE, Li et al. 2015[42]
PFD2	Progression free death rate at second-line	EVE+EXE	0.005 per month		Rate of on-treatment progression in BOLERO-2[43]
		EXE	0.003 per month		
		Chemo	1.64	(0.85, 3.15)	PFS2 HR for chemo (Li et al. 2015).[42] Other treatments assumed equal to EXE.

a Base case probabilities before calibration adjustment for partial surrogacy

b Values and credible intervals from model. Note these differ from the values in the forest plots in CS Appendix N Figures 33 and 35 (N.1.2)

4.3.4.3.1 Progression-free survival on second-line (PFS2)

Methods used to estimate PFS2 are described in CS section B.3.3.6.

Estimation of PFS2 from MONARCH 2

MONARCH 2 was a randomised, placebo-controlled trial comparing abemaciclib + fulvestrant with placebo + fulvestrant (see Appendix 9.2.4 of this report for a comparison of the MONARCH 3 and MONARCH 2 populations).[16]

For the economic model, the company fitted parametric survival curves to MONARCH 2 data for a subgroup of patients (38% of the randomised population) in this trial who had progressed on prior endocrine therapy for advanced disease to reflect the population at second-line in the current decision problem. Models were fitted with and without IC adjustment, although the unadjusted results are not used. The regression included a treatment indicator, but only estimates for the control arm (fulvestrant 500mg) are used in the model. The resulting PFS2 curves for fulvestrant are shown in CS Figure 29 (CS section B.2.2.6).

The company used similar methods to select the parametric function for PFS2 as for the first-line survival functions, concluding:

- Cumulative and log-log hazard plots show no evidence of violation of proportional hazards, so a proportional hazards model may be appropriate (CS Figures 21 and 22, M.1.4);
- Exponential, Weibull and Gompertz provide the best fit based on AIC and BIC statistics (CS Table 72, CS Appendix M.2.6);
- Exponential was chosen for the company base case, as it has the most favourable BIC. Weibull and Gompertz were used in scenario analyses. CS Table 73 (CS Appendix M.2.6) for parameters for these three survival functions.

PFS2 hazard ratios from second-line NMA

The CS states that the HRs for second-line treatments were obtained from the company's second-line NMA (CS Appendix N). See Appendix 9.2 for the ERG critical appraisal of the second-line NMA. We note concerns over clinical heterogeneity between the included trials, also highlighted by the company. The network included MONARCH 2, which had narrower inclusion criteria than other included trials. It also appears that data for the ITT population from this trial were used in the NMA, rather than the subgroup of patients who progressed

on endocrine therapy for advanced disease which was used to fit the fulvestrant curves for the economic model.

We note that the PFS HRs in the model for treatments in the second-line NMA (anastrozole, letrozole, everolimus and exemestane + everolimus) are similar but not exactly the same as values in the forest plot for the second-line NMA (CS Figure 33, CS Appendix N.1.2).

Tamoxifen and chemotherapy are included in the economic model but not in the company's second-line NMA. Chemotherapy was eligible for inclusion in the network and some chemotherapy trials (n=9) were identified, but they could not be connected to the network as no study compared an endocrine therapy to chemotherapy monotherapy or combination treatment. The model uses hazard ratios from a paper by Li et al. (2015) [42] for chemotherapy: 0.61 (95% CI: 0.32-1.17) for PFS and 0.53 (95% CI: 0.20-1.39) for OS for "everolimus based therapy' vs chemotherapy (the inverse of these hazard ratios were used in the model and applied relative to everolimus + exemestane (EVE+EXE)). The Li at al. paper was also used in the submission for the previous NICE appraisal of ribociclib (TA496). The ERG for that appraisal criticised the lack of rationale for the selection of this study as the source of evidence for second-line treatment effects of chemotherapy. They also commented on the lack of clarity in the Li et al. paper about whether 'everolimus based therapy' refers only to everolimus monotherapy or if it also includes everolimus combination therapy. We share these concerns.

We note that the confidence intervals for the chemotherapy HRs are incorrectly entered in the economic model, with the lower and upper limits the wrong way round. This has the effect of excluding uncertainty over this parameter from the probabilistic sensitivity analysis. We correct this error in ERG analysis.

The CS does not state why NMA results are not reported for tamoxifen, as this is listed as one of the treatments for inclusion and some of the included trials had tamoxifen arms. The source of relative treatment effects for tamoxifen is not discussed in the CS. The model specifies that HRs were obtained from a paper by Milla-Santos (2001), which is a report of an RCT comparing tamoxifen with toremifene in a first-line setting for advanced breast cancer.[41] The company does not justify choice of this source. The model indicates that PFS and OS hazard ratios for tamoxifen relative to fulvestrant were calculated by multiplying HRs relative to toremifene from the Milla-Santos paper by HRs for toremifene relative to

fulvestrant from the second-line NMA.[17] We could not replicate the values in the model (as in Table 19 above), although our results were similar.

BOLERO-2 scenario analysis

The company present a scenario for second-line PFS and OS based on the BOLERO-2 trial: a phase III RCT comparing everolimus + exemestane with exemestane in postmenopausal women with HR+/HER2- advanced breast cancer with recurrence or progression during or after treatment with an NSAI.[43, 44] See Appendix 9.4 of this report for a comparison of the MONARCH 2 and BOLERO-2 patient populations.

It is difficult to judge whether MONARCH 2 or BOLERO-2 provide a better source for extrapolation of post-progression outcomes. Having considered the available evidence, we consider that patients in the MONARCH 2 subgroup are broadly representative of patients progressing in the MONARCH 3 trial, with the caveat that this only applies to a relatively small sub-group of patients of the MONARCH 2 trial who had progressed on endocrine therapy for advanced disease (38%) (see Appendix 9.4 of this report). In NICE TA496, the committee accepted the assumption that patients in BOLERO-2 are representative of patients in MONALEESA-2 (the pivotal first-line trial that compared RIBO+NSAI with placebo+NSAI). We cannot verify whether the analysis of BOLERO-2 data for TA496 is consistent with that in the current appraisal, due to lack of detail in the current CS and redactions in the TA496 committee papers.

The company fitted parametric PFS curves to reconstructed BOLERO-2 data (from digitised Kaplan-Meier curves) for everolimus and exemestane + everolimus. They use a log-normal survival function for the scenario and parameters for log-logistic and gamma survival functions are also provided in the model. The company does not justify the choice of the log-normal distribution or provide any statistics or graphs to assess model fit. Values for fulvestrant, anastrozole and letrozole are estimated using hazard ratios relative to exemestane from the company's second-line NMA: cited in the model as [redacted], [redacted] and [redacted] for PFS and [redacted], [redacted] and [redacted] for OS. The company assumes equal treatment effects for exemestane, tamoxifen and letrozole.

We compare long-term PFS estimates from the company base case (MONARCH 2 exponential) and scenario (BOLERO-2 log-normal) in Table 20 below. The results are broadly similar, although the BOLERO-2 scenario gives slightly less favourable projections

than the MONARCH 2 base case. A clinical advisor to the ERG has indicated that the EVE+EXE estimates seem unrealistically high.

Table 20 PFS from second-line treatment: (interval-censored adjusted)

	Year	FUL	ANAS	LTZ	EXE	TMX	EVE+EXE	Chemo
MONARCH 2	0	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	5	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
BOLER- 2	0	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	5	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Source: produced by the ERG from survival curve estimates in the company model

ERG conclusions:

The company extrapolation of PFS for second-line fulvestrant is reasonable. We consider that the MONARCH 2 subgroup is broadly representative of patients progressing in the MONARCH 3 trial. And we agree with the use of IC adjustment and selection of the exponential survival curve for the base case, with Weibull and Gompertz scenarios.

There is uncertainty over the relative effects of other second-line treatments. We have concerns over the robustness of the second-line NMA, due to clinical heterogeneity (see Appendix 9.2). There are also small discrepancies between the PFS hazard ratios used in the model and values reported in the CS (CS Appendix N). The company has not provided justification for the choice of sources for chemotherapy and tamoxifen.

The BOLERO-2 trial analysis provides a useful cross-check for the MONARCH 2 results, particularly as BOLERO-2 was used for the assessment of post-progression outcomes in the NICE appraisal of ribociclib (TA492). However, the company has not provided any supporting evidence for the use of a log-normal curve for extrapolation of PFS.

4.3.4.3.2 Overall survival on second-line treatment (OS2)

Estimation of OS2 from MONARCH 2

The company estimates second-line OS curves for second-line treatments using a similar approach as for PFS. Fitted parametric curves for OS in the fulvestrant arm of MONARCH 2 are shown in CS Figure 33. Evidence for the fit of these curves is provided in CS Appendix M. The company concludes that there is no evidence of a violation of the proportional hazards assumption in MONARCH 2 OS data (though note that in their separate second-line treatment NMA report[17] they state that, based on Schoenfeld residual plots, the proportional hazards assumption for OS could not be supported for this trial). They note that the Gompertz curve has the best fit based on AIC and BIC statistics and Cox-Snell residual plots (CS Appendix M.2.5), but that the exponential, log-normal and log-logistic extrapolations are plausible, based on key opinion leader input. The company chose to use the exponential in their base case, with log-logistic and Gompertz in scenario analyses.

It is difficult to draw any meaningful conclusions about the validity of the proportional hazards assumption for OS from the MONARCH 2 trial, as the treatments are intertwined in the cumulative hazard and the log-log hazard plots (CS Figure 19 and 20 CS Appendix M), but this is not important, as the model only uses estimates for the fulvestrant arm. We question the decision to use an exponential curve for the company's base case, as this had a poor fit to MONARCH 2 survival data.

Long-term OS2 extrapolation from CONFIRM

Due to immaturity of the MONARCH 2 survival data (CS Figure 31 B.3.3.6), the company make use of data from the CONFIRM trial for extrapolation of OS2.[45, 46] CONFIRM was a randomised trial comparing fulvestrant 250 mg with fulvestrant 500 mg in postmenopausal women with HR+ advanced breast cancer. The company state that they chose this source as it is the only study from the second-line NMA that provided long-term OS data for fulvestrant (500 mg): reporting data up to 80 months, by which time around 20% of patients remained in the trial. We present information about the CONFIRM population in Appendix 9.4.

The company state that they fitted parametric distributions to reconstructed Kaplan-Meier data from the CONFIRM fulvestrant 500 mg arm. The CS states that they chose the Weibull

distribution for the CONFIRM extrapolation, but no information is provided to justify this choice. The company base case uses the MONARCH 2 exponential survival curve for fulvestrant up to the maximum follow-up (27.95 months). Extrapolation after this time is based on applying the hazard rate from the CONFIRM extrapolation. The resulting survival curve for fulvestrant is shown in CS Figure 34 (B.3.3.6).

OS2 hazard ratios from second-line NMA

The company estimates OS curves for other second-line treatments by applying hazard ratios relative to the survival curve for fulvestrant. The hazard ratios are estimated from the same sources as for PFS and we have the same concerns about differences between hazard ratios in the model and those cited in the CS and the sources of estimates for chemotherapy and tamoxifen (see Table 19).

Table 21 below shows second-line survival estimates from the company's base case model (MONARCH 2 exponential with Weibull extrapolation from CONFIRM) and also log-logistic and Gompertz extrapolations (MONARCH 2 without CONFIRM extrapolation). Clinical advice to the ERG suggests that the exponential and log-logistic estimates seem to overestimate long-term survival. One clinical advisor suggested to us that the Gompertz extrapolations are more reflective of current clinical experience, although another clinical expert has noted that they appear overly pessimistic.

Table 21 OS from second-line treatment

	Year	FUL	ANAS	LTZ	EXE	TMX	EVE+EXE	Chemo
Exponential with CONFIRM	0	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	5	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	10	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Log-logistic	0	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	5	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	10	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Gompertz	0	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	1	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	2	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	3	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	5	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
	10	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Source: produced by the ERG from survival curve estimates in the company model

ERG conclusions:

We disagree with the company's choice of an exponential survival function to model second-line OS for fulvestrant, as this has a poor fit to the MONARCH 2 data.

We are also concerned about the lack of evidence regarding the choice of Weibull distribution for the CONFIRM trial extrapolation. No evidence is provided regarding the goodness-of-fit of this or alternative parametric functions.

The Gompertz distribution has the best fit to MONARCH 2 data and clinical advice to the ERG is that the long-term survival predictions from the Gompertz are maybe more realistic than the alternatives presented by the company, although they may be rather too pessimistic. We therefore use Gompertz OS extrapolations in the ERG preferred analysis and include the log-logistic and exponential with CONFIRM extrapolations in scenario analysis.

4.3.4.3.3 Progression-free death rate on second-line (PFD2)

Additional information is required for the fixed pay-off model to estimate the three sets of transition probabilities (PFS2 to death, PFS2 to PPS and PPS to death) from PFS and OS curves – an issue that always arises with partitioned survival models. The approach taken is not discussed in the CS, but inspection of the model shows that PFS2 events are split into progressions and deaths using estimates of second-line time to progression (TTP2) and progression-free death rates (PFD2). The company use similar methods to estimate TTP2 as for PFS2: understandably as these outcomes only differ in that pre-progression deaths are included in the latter but not the former. In the base case, an exponential survival model fitted to data from the fulvestrant control arm in the MONARCH 2 trial is used for TTP2, but this yields the same results as for PFS2. Weibull and Gompertz parameters do differ between TTP2 and PFS2, but these are not used in the model.

The second-line pre-progression death rate is therefore estimated from external data. A simple monthly mortality rate estimated from on-treatment death rates in the BOLERO-2 trial: 0.005 per month (22 per 378 patient years) for everolimus + exemestane and 0.003 per month (4 per 103 patient years) for exemestane (Piccart et al. 2014).[43] The company assumes a higher mortality rate with chemotherapy (0.008 per month), based on the PFS hazard ratio from Li et al. (2001).[47] Rates for other second-line treatments are assumed to be the same as for exemestane. The overall probability of pre-progression deaths on second-line treatment is 0.005 per month, weighting by the company's assumed distribution of second-line treatments. A clinical advisor to the ERG has noted that this is a bit higher than expected.

> **ERG conclusions**: We agree with the use of BOLERO-2 trial data to estimate pre-progression death rates on second-line treatment, as this trial is larger with more mature survival data than MONARCH 2. We have some concerns over the source of relative effects between second-line treatments. We also note that uncertainty over the second-line pre-progression death rate is not factored into the company's deterministic or probabilistic sensitivity analysis. However, given the rarity of pre-progression deaths and the fact that rates do not differ between the first-line comparators, this parameter is very unlikely to affect cost-effectiveness results.

4.3.4.3.4 Overall survival calibration

A 'partial surrogacy' assumption is applied by calibrating the time spent in the fixed-pay-off sub-model until a desired ratio between median PFS gain and median OS gain for the first-line comparators relative to NSAI is achieved. The target for the calibration is 27.5% in the company base case. To achieve this target, the calibration weights are: 1.22 for ABE+NSAI; 1.41 for PAL+NSAI; 1.45 for RIBO+NSAI; and 1 for the reference treatment NSAI (CS Table 25, CS section B.3.3.7). For each comparator, the same weight is applied to all second-line event rates (progressions, deaths before progression and deaths after progression), thus holding the proportion of time spent in the three second-line health states (PFS2, PPS and death) constant. The calibration is implemented using the Excel 'goal seek' function. This is also applied within each PSA iteration; so, a different set of calibration factors is estimated for each iteration. Uncertainty over the calibration target itself is not reflected in the PSA. The company conducts a scenario analysis with 'full surrogacy' (i.e. calibration weights of 1 for all comparators).

The base case target of 27.5% surrogacy reflects the 'lower bound' specified by the committee for the NICE appraisal of palbociclib (TA495), based on fitting an exponential curve to final OS and PFS data from the PALOMA-1 trial. The TA495 committee concluded that the extension of PFS1 is likely to result in some improvement in OS, although the choice between the lower bound (27.5%) and upper bound (100%) is a source of uncertainty. The NICE DSU reviewed evidence on the relationship between PFS and OS, concluding that evidence on full surrogacy is 'inconclusive'.[39] Similarly, the NICE committee for appraisal TA496 concluded that ribociclib + NSAI improves PFS, that this is likely to result in some improvement in OS, that a degree of partial surrogacy is 'probably more likely' than full surrogacy, but that the magnitude of the relationship is highly uncertain.

> **ERG conclusion**: We consider that the company have correctly implemented the calibration and that they test an appropriate the range of assumptions about the magnitude of the surrogacy relationship between OS and PFS, as requested by previous NICE appraisal committees TA495 and TA496 (from 27.5% to 100% surrogacy). We also test the conservative assumption of no surrogacy and other intermediate values in our analyses.

4.3.4.4 Adverse event rates

The model applies adverse event (AE) related QALY decrements and costs as one-off penalties at the start of first-line treatment. Grade 3-4 treatment-related AEs that occurred for at least 5% of patients for at least one comparator are included, based on the main publication for each comparator in the NMA: see Table 22 below (CS Table 29, B.3.4.4). Adverse events were not modelled explicitly for second or third line treatments.

Table 22 Adverse event probabilities in the model (adapted from CS Table 29)

Event	ABE+NSAI	NSAI	PAL+NSAI	RIBO+NSAI
Alanine aminotransferase increased	[redacted]	[redacted]	0.2%	9.0%
Anaemia	[redacted]	[redacted]	5.9%	2.4%
Aspartate aminotransferase increased	[redacted]	[redacted]	0.0%	6.0%
Diarrhoea	[redacted]	[redacted]	1.4%	2.4%
Hypertension	[redacted]	[redacted]	0.0%	10.0%
Leukopenia	[redacted]	[redacted]	24.8%	21.0%
Lymphopenia	[redacted]	[redacted]	0.0%	7.0%
Neutropenia	[redacted]	[redacted]	67.1%	59.0%

Sources: ABE+NSAI and NSAI, ITT population from MONARCH 3 CSR; PAL+NSAI from PALOMA 2;[48, 49] RIBO+NSAI from MONALEESA-2.[19, 50]

Incidence of neutropenia and leukopenia were high for all three of the CDK4/6 inhibitors, but particularly so for palbociclib and ribociclib. Abemaciclib is associated with a high incidence of diarrhoea The committee for the palbociclib appraisal (TA495) concluded that although incidence of neutropenia is high, adverse events are manageable and treatment discontinuation in practice will tend to be lower than in the trials. Clinical experts advising in TA496 stated that AEs are more common at treatment initiation and are usually resolved with dose reductions and interruptions (TA496). This view was supported by the clinical advisers to the ERG.

Other adverse effects that are important to patients are omitted from the model: in particular, fatigue, nausea, vomiting and infection. Almost all of the events included are measurements

that often do not impact on how the patient feels, whereas nausea/vomiting and fatigue are symptoms that patients have to live with/adapt to and infection often causes symptoms that make patients feel less well. Raised serum creatinine was another toxicity reported in a significant proportion of patients treated with abemaciclib that was not seen with palbociclib or ribociclib but is important to note as this treatment will potentially be used in older patients with HR+ metastatic breast cancer who may have existing renal impairment. This suggests that the effects of adverse treatment effects may have been underestimated in the model.

4.3.5 Health related quality of life

4.3.5.1 Health state utilities

The company report a systematic literature review of utility studies (CS B.3.4.1 and Appendix H) but conclude that studies found were not representative of the population of interest. Instead, utilities for the model are estimated from analysis of EQ-5D-5L data from MONARCH 3 and MONARCH 2 and from previous NICE appraisals – reported in CS Tables 26, 27 and 28 (B.3.4.2). We summarise sources in Table 23 and discuss further below.

Table 23 Health state utility estimates

Source	PFS1	PFS2[a]	PPS	Comments
Company analysis				
Base case	[redacted] Overall	0.745[a]	0.505	MONARCH 3 Model 1 for PFS1. Others from TA496
Scenario 1	[redacted] NSAI [redacted] Other	0.745[a]	0.505	Treatment specific PFS1 from MONARCH 3 (Model 2)
Scenario 2	0.774	0.745[a]	0.505	PFS1 assumed equal to PFS2 (without chemotherapy)
Scenario 3	[redacted]	[redacted][a]	0.505	PFS2 from MONARCH 2 pre-progression utility
Scenario 4	[redacted]	0.745[a]	[redacted]	PPS estimated from MONARCH 3 progression disutility applied to PFS1[c]
Company estimates form trial data [b]				
MONARCH 3	[redacted] Overall [redacted] NSAI [redacted] ABE+NSAI	[redacted]		EQ-5D-5L adjusted for repeated measures, baseline utility and progression, with / without treatment arm
MONARCH 2		[redacted]	[redacted]	As above, without treatment
Previous NICE appraisals				
TA495 (palbociclib)	0.72 Overall 0.71 NSAI 0.74 PAL+NSAI	0.505	0.505	PALOMA 2 EQ-5D-3L, mean baseline values for PFS1. Estimated from Lloyd et al.[37] by ERG.[51]
TA496 (ribociclib)	Redacted in committee papers	0.774 initial 0.690 final, suggested by DSU	0.505	PFS1 from MONALEESA-2 EQ-5D-5L mixed model for repeated measures. PFS2 based on Lloyd et al. model [37] adjusted for BOLERO-2 age and response. DSU proposed reduction.[39]

[a] Weighted mean with disutility of 0.113 (Peasgood et al. 2010 [39]) applied for patients on chemotherapy at secondline (25.66%). Utility assumed equal for other second-line treatments.

[b] Values from CS Tables 26 to 28 and model.

[c] CS states scenario is based on MONARCH 2, but model applies MONARCH 3 disutility

4.3.5.1.1 Analysis of EQ-5D-5L data from MONARCH 3

EQ-5D-5L was administered in MONARCH 3 at baseline, at the start of alternate 28-day cycles up to cycle 19 and then at every third cycle. There were no significant differences between arms in change from baseline to final PFS EQ-5D-5L index or visual analogue scores - see section 3.3.5 above.

To inform the economic model, the company further analysed these data using a mixed model for repeated measures, with adjustment for baseline utility and progression (Model 1) and with an additional treatment variable to provide separate estimates for NSAI and ABE+NSAI (Model 2): see CS B.3.4.2 for methods and CS Appendix M.4.2 for results. We note that the CS omits important information about the methods of analysis: the population (safety or ITT); and the extent of missing data or whether attempts were made to impute missing values. The company states that utilities were calculated for the base case using the 'cross-walk' procedure, as recommended by NICE for consistency with UK EQ-5D-3L index values (CS B.3.4.2).[52] However, the CS reports the same results for the 'crosswalk' (CS M.4.2) as 'EQ-5D-5L' (CS M.4.1), and similarly in the model.

The company use the pre-progression utility from Model 1 ([redacted]) for PFS1 in their base case for all first-line interventions. They state that this is conservative, as there was no significant difference between treatments in Model 2. [redacted] The company use Model 2 results in a scenario, applying the ABE+NSAI PFS1 utility to all first-line treatments, which increased the ICER for ABE+NSAI vs NSAI.

The company do not use MONARCH 3 post-progression estimates for the economic analysis. Estimates were consistent between the utility models: mean [redacted] and [redacted] from Model 1 and Model 2 respectively. This results in an overall post-progression utility of [redacted], without treatment adjustment. We note that it is not obvious whether this estimate applies to the PFS2 or PPS health state, since some patients may have experienced a second progression during trial follow up.

ERG conclusions: The general approach to utility estimation from MONARCH 3 EQ-5D-5L data is appropriate, with use of a mixed model for repeated measures and

adjustment for baseline utility, progression and treatment group. We do have some reservations however about the reliability of the results because the CS omits information about the analysis population and handling of missing data.

Analysis of EQ-5D-5L data from MONARCH 2

The company reports a similar analysis of EQ-5D-5L data from the MONARCH 2 trial to inform second-line utility estimates (CS B.3.4.2 and M.4.2). This was conducted on the subgroup of patients who had progressed on prior endocrine therapy in the locally advanced or metastatic setting. A mixed regression model for repeated measurements was used, with adjustment for baseline utility and an indicator variable for progression, but no treatment indicator.

The company does not use MONARCH 2 utility estimates for their base case, although the pre-progression utility ([redacted]) is used for in a scenario for PFS2. This is assumed to apply for patients on endocrine or targeted therapies. As in previous NICE appraisals, including TA495 and TA496, an additional decrement of 0.113 (Peasgood et al. 2010)[53] is applied for the 25.66% of patients assumed to have chemotherapy at second-line. This results in an overall mean PFS2 utility estimate of [redacted]. The company also state that they use MONARCH 2 to estimate PPS utility in another scenario (see CS Table 28, B.3.4.2). However, examination of the model shows that this scenario actually uses the progression disutility from the MONARCH 3 analysis applied to the PFS1 utility: [redacted]. The post-progression utility estimated from the MONARCH 2 analysis is higher: [redacted].

ERG conclusions: The MONARCH 2 utility analysis shares the same strengths and weaknesses as the MONARCH 3 analysis described above. However, there is additional uncertainty about the compatibility of the MONARCH 2 subgroup with the MONARCH 3 population. The fact that pre-progression utilities from MONARCH 3 ([redacted]) are lower than pre-progression utilities from MONARCH 2 ([redacted]) is problematic. This might be a chance finding for two independent trial samples, or it might reflect a more structural incompatibility of patient selection or recruitment. Either way it is not realistic to assume a lower utility for PFS1 than for PFS2, as this implies that patients have a worse quality of life when progression-free at first-line than after disease progression at second-line.

4.3.5.1.2 Utility estimates from previous NICE appraisals

Utilities for the PFS1 health state in the appraisals for palbociclib and ribociclib (TA495 and TA496) were estimated using EQ-5D data from the PALOMA-2 and MONALEESA-2 trials respectively. The results are not available for ribociclib, because they are redacted in the NICE committee papers. For TA495, the company submission reports PFS utilities for palbociclib plus letrozole (0.74) and letrozole (0.71). The ERG for TA495 argued that this difference was not statistically significant and used a mean averaged across both arms (0.72). We also note that utility estimates from PALOMA-2 were just the treatment baseline values, assumed to apply for the duration of the pre-progression state.

For the post-progression health states, the company in the present appraisal relies on precedent for their base case: 0.774 for PFS2 (endocrine or targeted therapy) and 0.505 for PPS. These values are the same as in the Novartis submission for the NICE appraisal of ribociclib (TA496), derived in previous appraisals from a standard gamble study by Lloyd et al. (2006).[37] In this study, members of the UK general public were asked to value hypothetical health states for patients with metastatic breast cancer, described in vignettes. Results were analysed in a mixed model with a logistic transformation to estimate changes in utility related to the age of the respondent, stage of disease and treatment toxicities. The utility of 0.505 for progressed disease was calculated from the Lloyd et al. formula by the ERG in TA495 (by adjusting for the mean age of participants in the EQ-5D-3L UK value set survey).[51]

The estimate of 0.774 for PFS2 originated in TA421, calculated from the Lloyd et al. formula for stable disease, allowing for the treatment response rate in the BOLERO-2 trial. However, we note that the PFS2 value of 0.774 was not used in the final analysis for TA496. This was because the PFS1 utility estimated from the MONALEESA-2 trial exceeded 0.774, which was considered unrealistic. The DSU[39] suggested a revised value of 0.69 for PFS2, which was accepted by Novartis. The TA496 committee concluded that this assumption was appropriate for decision making but that the resulting utilities may undervalue the quality of life for patients in the progression-free state.

In the current appraisal, the company acknowledge the inconsistency in their base case of using a PFS2 utility that is higher than the PFS1 utility. They address this in a scenario in which they increase the PFS1 utility to 0.774. However, an alternative approach, as in TA496, would be to reduce the PFS2 utility.

ERG conclusions:

We consider that MONARCH 3 is the best source for the PFS1 utility ([redacted]): it complies with the NICE reference case (assuming crosswalk values are used); uses EQ-5D-5L data collected directly from participants in the pivotal trial; and the methods of analysis are appropriate, although we do have some reservations about lack of detail in reporting. We have a general preference for the treatment-specific utility estimates from MONARCH 3, because they reflect benefits and harms of treatments directly assessed by patients. However, equivalent treatment-specific utilities are not available for all comparators. We therefore agree with the company's decision to use the overall PFS1 utility for all comparators in their base case.

For post-progression utilities, the company's decision to use estimates from previous NICE appraisals derived from the Lloyd et al. formula has the merit of consistency between appraisals, although it does not comply with the NICE reference case, as health state measures are not obtained from patients. We consider that the company fails to address the inconsistency between the pre and post-progression utilities in their base case, as they use a PFS2 value that is higher than the PFS1 value. This same problem arose in TA496 and resulted in revision of the PFS2 value from 0.774 to 0.690. We suggest that this value should also be used in the current appraisal. We conduct one-way sensitivity analysis for PFS1 and PFS2, changing them from upper to lower bounds while respecting the assumption that the utility for PFS1>PFS2>PPS.

4.3.5.2 Adverse events

Assumptions underlying the estimation of QALY loss associated with treatment-related adverse events for first-line treatment options are described in CS B.3.4.4. In addition to the probability of modelled adverse events (see section 4.3.4.4 above), this includes a disutility and duration for each AE (CS Table 30 and 31 respectively). The company report that a systematic literature review was consulted to identify sources for these parameters, but that no relevant studies were identified. No further details of this search are provided. Cited sources for AE parameters are Hudgens et al. (2016)[54], Swinburn et al. (2010)[55] and NICE appraisals TA306 (pixantrone for non-Hodgkins lymphoma[56]) and TA503 (fulvestrant for untreated HR+ advanced breast cancer[57]).

We summarise the modelled AE QALY loss per person starting first-line treatment in Table 24 below. These values are very low due to the short duration (from 0 to 34 days) and small

disutility (0 to 0.153) attached to the events. Clinical advice to the ERG suggests that differences in the adverse event profiles of comparators can affect HRQoL. For abemaciclib, diarrhoea is more frequent, but this is easily controllable and usually short-lived. Patients on palbociclib and ribociclib may have low white cell count but not episodes of sepsis that could affect HRQoL.

Table 24 Adverse event QALY loss

First-line treatment	QALY loss per person starting treatment	Source
NSAI	-0.00062	Weighted means based on AE probabilities, utility decrements and durations (CS Tables 29-31)
ABE-NSAI	-0.00008	
PAL-NSAI	-0.00054	
RIBO-NSAI	-0.00100	

4.3.6 Resource use and costs

4.3.6.1 Use of second and third line treatment options

The company's base case assumptions about the proportions of patients receiving second and third-line treatment options are summarised in CS Tables 35 and 39 respectively – summarised in Table 25 below.

Table 25 Use of second and third-line therapies (adapted from CS Table 35 and 39)

	Company base case		ERG scenario	
	Second-line	Third-line	Second-line	Third-line
Chemotherapies	**25.7%**	**30.4%**	**25%**	**50%**
Capecitabine	12.3%	24.8%	12%	41%
Paclitaxel	6.2%	0.0%	6%	0%
Docetaxel	7.2%	0.0%	7%	0%
Eribulin	0.0%	5.6%	0%	9%
Endocrine therapies	**66.3%**	**24.0%**	**35%**	**25%**
Fulvestrant	10.9%	10.1%	0%	0%
Anastrozole	0.0%	0.0%	0%	5%
Letrozole	0.0%	0.0%	0%	5%
Exemestane	37.0%	6.2%	15%	5%
Tamoxifen	18.5%	7.7%	20%	10%
Everolimus + exemestane	**8.0%**	**0.0%**	**40%**	**10%**
No treatment	**0.0%**	**45.6%**	**0%**	**15%**

These are based on assumptions in the NICE appraisal of fulvestrant for untreated HR+ advanced breast cancer (TA503)[57] and the company's assumption that NSAIs would not be used following use at first line.

> **ERG conclusion**: Clinical advice to ERG suggests that these distributions do not reflect current NHS practice and policy. Fulvestrant is not used at second or third line, because it is not recommended by NICE (TA239) and fewer patients have exemestane monotherapy now that everolimus + exemestane are recommended by NICE (TA421). At third-line, a greater proportion of patients have chemotherapy (around 50%), with few patients receiving no treatment (10-15%). NSAIs may also be used sometimes at third-line. We test the impact of a scenario based on this clinical advice in ERG analyses.

4.3.6.2 Duration of treatment

We summarise methods used to model treatment duration in Table 26. For first- and second -line treatments, similar methods are used as for TTP and PFS: with parametric survival curves fitted to MONARCH 3 (NSAI and ABE+NSAI) and MONARCH 2 (FUL), adjusted for other comparators with hazard ratios. However, as time to discontinuation is not reported in trial publications, hazard ratios were estimated based on reported median treatment durations. Third line treatment is only included in the model as a cost, applied for a fixed proportion of time spent in the PPS health state.

Table 26 Time to treatment discontinuation

		Treatment	Base case		Source
TTD1	Time to discontinuation of first-line treatment	NSAI ABE+NSAI	Gamma survival curves (joint fit)		MONARCH 3, IC-adjusted (CS Figures 24 & 25)
		[redacted]	[redacted]		Hazard ratios relative to NSAI estimated from median times on treatment (CS Appendix M Table 68 M.2.4)
		[redacted]	[redacted]	[redacted]	
		PAL+NSAI	19.8 months	HR 0.81	
		RIBO+NSAI	20.3 months	HR 0.79	
TTD2	Time to discontinuation of second-line treatment	FUL			
		[redacted]	[redacted]		Hazard ratios relative to fulvestrant, estimated from median times on treatment (CS Appendix M Table 78 M.2.4)
		ANAS	5.6 months	[redacted]	
		LTZ	5.9 months	[redacted]	
		EXE	4.4 months	[redacted]	
		TMX	4.4 months	[redacted]	
		EVE+EXE	7.8 months	[redacted]	
		CAP	4.8 months	[redacted]	
		PAC	4.8 months	[redacted]	
		DOC	4.8 months	[redacted]	

Third line: proportion of time in PPS spent on treatment	37%	

[a] Not used in company base case (included here for reference).

Time to discontinuation of first-line treatment (TTD1) with ABE+NSAI and NSAI is estimated using parametric survival models fitted to MONARCH 3. Estimation methods are similar to those for TTP1 (see CS section B.3.3.5 and CS Appendix M.1.2 and M.2.4). The company concludes that treatment effects are multiplicative over time, rather than proportional, and that the log-normal, gamma and Gompertz models provide a good fit to the observed data. However, as treatment continuation is constrained by progression (modelled as an exponential), the company ruled out the log-normal and Gompertz curves for the base case (they 'overshoot' progression). They therefore chose the gamma distribution for TTD1, with log-normal, Gompertz and exponential curves used as scenarios. Note the model does also constrain time to discontinuation to not exceed time to progression. Time to discontinuation of the other first-line comparators (PAL+NSAI and RIBO+NSAI) was estimated relative to NSAI using hazard ratios estimated from median times to discontinuation. The resulting TTD1 extrapolation curves are shown in CS Figure 26.

The process for fitting time to discontinuation of second-line treatment (TTD2) was similar to that for PFS2 (CS section B.3.3.6 and CS Appendix M.1.5 and M.2.8). Joint parametric survival curves were fitted to MONARCH 2 data, although only the curve for the fulvestrant curve was used in the model. The company concluded that there was no evidence of violation of the proportional hazards assumption and that the Gompertz curve has the best fit to trial data. However, this overshoots progression, modelled with an exponential curve. The company decided to use an exponential curve for TTD2 in the base case and Gompertz and log-logistic curves for scenario analysis. Consideration of CS Figure 37, which shows the fitted parametric curves in relation to the Kaplan-Meier curve for the fulvestrant arm of MONARCH 2, indicates that exponential does provide a reasonable fit for TTD2.

4.3.6.2.1 Duration of third-line treatment

The company estimates time on third-line therapy, calculated based on an assumption that patients spend approximately 37% of their time on treatment after progression from second-line treatment. This assumption was based on clinical expert opinion. Estimated time on treatment based on this assumption is presented in Table 27.

ERG conclusion:

We agree with the company's choice of survival curves for time to discontinuation of first and second-line treatments and apply the same base case and scenarios in ERG analysis. However, clinical advice to the ERG is that it would be unusual for patients to spend as much as 63% of time after a second disease progression without treatment. Thus, the cost of treatment in the PPS health state is probably underestimated. We vary the proportion of PPS spent on treatment (from 10 to 50%) to assess the impact of uncertainty around this parameter.

Table 27 Time on third-line treatment (CS Table 43)

First-line treatment	Time in PPS (months)		
	On treatment	Off treatment	Total
ABE+NSAI	12.17	20.72	32.89
PAL+NSAI	12.26	20.88	33.15
RIBO+NSAI	12.26	20.88	33.15
NSAI	12.17	20.72	32.89

4.3.6.3 Drug costs

Table 28 Drug acquisition and administration costs

Drug	Dose Per cycle	Cycle	Dose intensity[b]	Drug cost		Admin. Per month
				Per cycle	Per month	
ABE	8,400 mg	28 days	[redacted]	[redacted]	[redacted]	
PAL	2,625 mg	28 days	93%	£2,950	£3,205	
RIBO	12,600 mg	28 days	88%	£2,950	£3,205	
LTZ	70 mg	28 days	[redacted]	£2.71	£2.94	
ANAS	28 mg	28 days	[redacted]	£1.34	£1.46	
CAP	59,437 mg	21 days	100%	£21.56	£31.22	£237
PAC	297 mg	21 days	78%	£0.39	£0.56	£376
DOC	127 mg	21 days	78%	£4.65	£6.74	£376
ERI	4 mg	21 days	87%	£1,714	£2,482	£752
FUL	500 mg	28 days	[redacted]	£522	£568	£238[a]
EXE	700 mg	28 days	100%	£3.44	£3.74	
TMX	608 mg	30 days	100%	£1.61	£1.61	
EVE	280 mg	28 days	100%	£2,495	£2,710	

[a] Loading dose only
[b] Not applied in base case (includes wastage)

Table 28 above, summarises drug acquisition and administration costs, including:

- Treatment regimens and acquisition costs: CS section B.3.5.1; Tables 33 and 34 for first-line; and Tables 36 and 37 for second-line.

- The base case assumes wastage (100% of dose for oral therapies and disposal of unused vial contents for IV therapies). The company conducts a scenario analysis with reduced costs according to the relative dose intensities shown in Table 28, derived from the primary trial publications for first-line therapies, MONARCH 2 for fulvestrant, Beuselinck et al. (2009)[58] for paclitaxel, Kaufmann et al. (2015)[59] for eribulin and assumptions for other second and third line treatments.
- No administration costs were applied for oral treatments, except for capecitabine, for which the NHS Reference cost for oral chemotherapy was incurred. Paclitaxel, docetaxel and eribulin incurred a cost per cycle for delivery of simple chemotherapy. A cost for face-to-face, first medical oncology visit was assumed for the first, loading dose of fulvestrant.

4.3.6.4 Health care costs

The model includes additional costs for follow up and care. These include:

- **Follow up care and monitoring**. See CS Tables 52 and 53. This includes diagnostic tests, outpatient oncology consultation, GP surgery visits, community and clinical nurse specialist care at home and therapy. These costs are related to health state, with resource used informed by MONARCH 3 and MONARCH 2 data and packages of care defined in the NICE Advanced Breast Cancer guideline (CG81).
- **Treatment for adverse reactions**. See CS Table 56, B.3.5.3. The cost of treatment adverse events was modelled as a one-off fixed cost at the start of treatment. The company assumed one outpatient visit for grade 3/4 hypertension, leukopenia, lymphopenia and neutropenia and a blood transfusion for anaemia. In their base case, the company assumed that grade 3/4 diarrhoea would be treated with loperamide, but they conducted a sensitivity analysis assuming a non-elective short-stay hospital admission.
- **Hospitalisation**. CS B.3.5.2 and Tables 47 to 51. Admission rates and lengths of stay were estimated by health state (PFS1, PFS2 and PPS) based on observations in the MONARCH 3 and 2 studies.
- **Best supportive care**. CS B.3.5.2 and Tables 45 and 46. This included palliative medications, with rates of use taken from the MONARCH 3 and 2 data.
- **End of life care.** Place of death and and packages of care were based on the NICE Advanced breast cancer clinical guidelines, CG81.

With the exception of first-line drug costs and treatment for adverse events, costs were the same for the first-line comparators. We summarise the monthly non-drug costs by health state in Table 29.

Table 29 Average monthly health care costs

	Average cost per month		
	PFS1	**PFS2**	**PPS**
Follow up care	£443	£635	£691
Adverse events	£106	-	-
Hospitalisation	£33	£46	£40
Best supportive care	£146	£146	£69
Total	**£728**	**£828**	**£800**
End of life	£4,379	£4,379	£4,379

For comparison, in NICE TA495 (palbociclib) the ERG estimated a mean cost per cycle of £1200 per cycle for active treatment states and £975 for best supportive care. The committee noted that these estimates were similar to confidential estimates by the Cancer Drugs Fund clinical lead in consultation with experts from the Chemotherapy Clinical Reference Group of NHS England. The committee agreed that the ERG estimates for post-progression costs are plausible. In NICE TA496 (ribociclib), the committee tested monthly costs in the PPS state in the region of £1140 to £1200 (ERG TA495 estimate) in decision making.

4.3.7 Model validation

The company report an external validation of their model was conducted by an analyst who was not initially involved in the model design or programming. The CS describes a series of iterations between analysts to identify and address areas of disagreement. The company also sought the opinion of their clinical experts to review the outputs from survival extrapolations.

The ERG checked the company's economic model for transparency and validity. The model was developed in Microsoft Excel and the visual basic codes were accessible.

We conducted a range of 'white box' tests to verify model inputs, calculations and outputs which consisted of:

- Cross-checking of all parameter inputs against values in the CS and cited sources;

- Checking that model outputs such as base case deterministic results and results of scenario analysis reported in the CS were reproducible by manually running the model;
- Checking individual equations and formulas within the model;
- Testing the logic of formulas in the model by substituting model inputs with a range of extreme values;
- Checking that visual basic codes did what they were designed to do.

Generally, we found the economic model to be of a good quality, with very few errors in input parameters, logic or coding. We identified a few small errors that we report and correct in section 4.4.1 below. However, these errors did not make any substantive difference to the results of cost-effectiveness analysis.

4.3.8 Company cost effectiveness results

Results from the economic model are presented in Section B.3.7, page 140 of the CS.

The base case results, presented in terms of incremental cost per QALY gained (Table 30) show that PAL+NSAI and RIBO+NSAI are both dominated by ABE+NSAI (that is, it has lower costs and higher QALYs). Model outputs from ERG corrections are reported in section 4.4.1 of this report and show minor variations from the company's results, however these differences do not alter the company's conclusions.

Table 30 Company base case results – deterministic (CS Table 59)

Treatment	Total		Incremental analysis ICER (£/QALY)	Pairwise ICERs ABE+NSAI vs. comparator (£/QALY)
	Costs (£)	QALYs		
NSAI	£56,449	2.997	Referent	£250,065
PAL+NSAI	£145,266	3.225	Dominated	ABE+NSAI Dominant
RIBO+NSAI	£148,170	3.222	Dominated	ABE+NSAI Dominant
ABE+NSAI	£129,803	3.291	£250,065	-

The CS summarises the results of the PSA stating that there is a 82% probability of ABE+NSAI being cost-effective, relative to PAL+NSAI and RIBO+NSAI, at a threshold willingness to pay of £30,000 per QALY gained.

Table 31 Company base case results – probabilistic (CS Table 61)

Treatment	Total		Incremental analysis ICER (£/QALY)	Pairwise ICERs ABE+NSAI vs. comparator (£/QALY)
	Costs (£)	QALYs		
NSAI				
PAL+NSAI	£139,631	3.15	-	-
RIBO+NSAI	£142,571	3.16	£397,144	£397,144
ABE+NSAI	£125,581	3.21	Dominant	Dominant

One-way sensitivity analyses

The company does not present one-way sensitivity analyses in the CS. In response to the ERG's clarification question B3, the company states that it does not believe one-way sensitivity analysis are crucial to decision making.

Scenario analysis

The CS reports a deterministic scenario analysis to explore the impact of base case assumptions in 29 scenarios. Results of these analyses are presented below (Table 32).

Table 32 Company scenario results (Adapted from CS Table 63)

Scenario	Base case value	Scenario	ICER (£ per QALY gained)		
			ABE+NSAI	PAL+NSAI	RIBO+NSAI
Base-case	N/A	N/A	£250,065	Dominated	Dominated
Discount rates	3.50%	0.00%	£212,582	Dominated	Dominated
Discount rates	3.50%	6.00%	£279,248	Dominated	Dominated
ABE+NSAI PFS1 treatment effect	Joint model MONARCH 3	NMA	£341,342	£1,378,635	Dominated
IC adjustment	IC-adjusted analysis	Unadjusted analysis	£250,065	Dominated	Dominated
Covariate adjustment	IC-adjusted analysis	Covariate and IC-adjusted analysis	£222,795	Dominated	Dominated
TTP1 (scenario 1)	Exponential	Weibull	£240,007	Dominated	Dominated
TTP1 (scenario 2)	Exponential	Gompertz	£571,795	Dominated	Dominated
PFS2 (scenario 1)	Exponential	Weibull	£256,368	Dominated	Dominated
PFS2 (scenario 2)	Exponential	Gompertz	£278,660	Dominated	Dominated
OS2 (scenario 1)	Exponential + CONFIRM	Exponential	£282,398	Dominated	Dominated
OS2 (scenario 2)	Exponential + CONFIRM	Log-logistic	£245,869	Dominated	Dominated
Second-line OS (scenario 3)	Exponential + CONFIRM	Gompertz	£197,053	Dominated	Dominated
TTD1	Gamma	Gompertz	£263,628	Dominated	Dominated
TTD1	Gamma	Log-normal	£254,708	Dominated	Dominated
TTD1	Gamma	Exponential	£223,727	Dominated	Dominated
TTD2	Exponential	Log-logistic	£250,065	Dominated	Dominated
TTD2	Exponential	Gompertz	£250,065	Dominated	Dominated
HRs for TTD2	Vs FUL based on median ToT	Vs second-line PFS	£248,546	Dominated	Dominated
Utility model	Overall	Treatment-specific	£269,922	Dominated	Dominated
PPS utility source	Lloyd, 2006	MONARCH 2	£411,806	Dominated	Dominated

Second-line PFS utility	TA496	MONARCH 2	£248,716	Dominated	Dominated
PPS hospital length of stay	MONARCH 2	MONARCH 3	£248,499	Dominated	Dominated
Relative dose intensity	OFF	ON	£196,532	Dominated	Dominated
PFS1 utility value	MONARCH 3	Equal to PFS in second-line treatment	£209,593	Dominated	Dominated
Source of clinical outcomes in PPS	MONARCH 2	BOLERO-2	£182,754	Dominated	Dominated
Apply PFS–OS surrogacy	Yes (27.5%)	No (100%)	£159,286	Dominated	Dominated
PFS 1 utility source	EQ-5D-3L (crosswalk)	EQ-5D-5L	£250,065	Dominated	Dominated
Management of diarrhoea	Loperamide	Hospitalisation and loperamide	£251,084	Dominated	Dominated

The scenarios are clearly stated and justified. PAL+NSAI and RIBO+NSAI are dominated by ABE+NSAI in all scenarios. We reran the company's scenarios after effecting our corrections and they are reported in section 4.4.2 of this report.

Probabilistic sensitivity analysis

The company's model computes PSA results based on 10,000 iterations. The ERG finds that running the PSA is computationally challenging (running the 28 scenarios takes over 2 hours) due to the calibration calculations required to adjust OS.

The ERG is of the opinion that 1000 iterations are sufficient to produce reasonably stable results. Our rerun of the PSA at 1000 iterations takes about 30 minutes.

4.4 Additional work undertaken by the ERG

4.4.1 ERG corrections to company model

We identified some minor errors in the company's model, as shown in Table 33 below.

Table 33 ERG corrections to company model

Aspect of model	Problem	ERG Correction
1.Hazard ratios and relative risks	Upper and lower confidence interval values for second-line chemotherapy are entered the wrong way round for all clinical outcomes in the model. This wrong entry only affects the results of the probabilistic analysis.	Reordered chemotherapy hazard ratio and relative risk confidence interval values for second-line time to progression, second-line progression-free deaths, second-line progression-free survival and second-line overall survival.
2.Pre-progression deaths	PAL and RIBO estimated relative to ABE.	Corrected so that the extrapolated hazards of pre-progression deaths for patients on PAL and RIBO are estimated relative to NSAI.
3. TTP1	Extrapolations from Gompertz distributions for ABE+NSAI (unadjusted) and NSAI (unadjusted) use shapes from IC and covariate adjusted calculations.	We corrected the formulas to so that the appropriate shapes are used.
4. The percentage of PFS events that are deaths	The company model estimates this from an incorrect denominator – PFS2 events instead of the sum of patients experiencing progression and pre-progression deaths in the payoff sub model).	We corrected the appropriate formulas in the model. This gives a fixed proportion of 4.4% of the people leaving PFS each month which matches the input assumptions.

4.4.2 Results from ERG corrected company base case

The results of the company's base case with ERG corrections are presented in Table 34.

Table 34 Company base case results (ERG corrected) - deterministic

Treatment	Total		Incremental analysis ICER (£/QALY)	Pairwise ICERs ABE+NSAI vs. comparator (£/QALY)
	Costs (£)	QALYs		
NSAI	£56,152	2.997	Referent	£250,352
PAL+NSAI	£152,268	3.273	Dominated	ABE+NSAI Dominant
RIBO+NSAI	£154,559	3.285	Dominated	ABE+NSAI Dominant
ABE+NSAI	£129,590	3.291	£250,352	-

Table 35 shows the ERG corrected version of the company's scenario analyses.

Table 35 Company scenario results (ERG corrected)

Scenarios	Treatments	Total costs (£)	Total QALYs	Incremental ICER (£/QALY)	Pairwise ICER ABE+NSAI vs. comparator
Discount rates: 0.00%	NSAI	63,783	3.381	Referent	£212,804
	PAL+NSAI	170,307	3.721	Dominated	ABE+NSAI Dominant
	RIBO+NSAI	172,946	3.735	Dominated	ABE+NSAI Dominant
	ABE+NSAI	144,531	3.760	£212,804	-
Discount rates: 6.00%	NSAI	51,717	2.774	Referent	£367,282
	PAL+NSAI	141,688	3.014	Dominated	£187,961
	RIBO+NSAI	143,775	3.025	£6,988,613	-
	ABE+NSAI	120,879	3.021	£279,586	£6,988,613
ABE+NSAI treatment effects for PFS: NMA	NSAI	56,152	2.997	Referent	£342,211
	PAL+NSAI	152,268	3.273	Ex Dominated	£188,241
	RIBO+NSAI	154,559	3.285	£343,915	-
	ABE+NSAI	130,514	3.215	£341,663	£343,915
Interval censoring unadjusted	NSAI	56,152	2.997	Referent	£250,352
	PAL+NSAI	152,268	3.273	Dominated	ABE+NSAI Dominant
	RIBO+NSAI	154,559	3.285	Dominated	ABE+NSAI Dominant
	ABE+NSAI	129,590	3.291	£250,352	-
Covariate and interval censoring adjusted	NSAI	58,122	3.127	Referent	£223,086
	PAL+NSAI	159,934	3.400	Dominated	ABE+NSAI Dominant
	RIBO+NSAI	161,058	3.400	Dominated	ABE+NSAI Dominant
	ABE+NSAI	142,262	3.504	£223,086	-
TTP1 Weibull	NSAI	56,305	3.018	Referent	£330,052
	PAL+NSAI	155,494	3.311	Dominated	£170,309
	RIBO+NSAI	158,148	3.327	£5,606,781	-
	ABE+NSAI	129,213	3.322	£240,299	£5,606,781
TTP1 Gompertz	NSAI	56,506	3.051	Referent	£311,553
	PAL+NSAI	162,059	3.396	£2,469,570	£935,832
	RIBO+NSAI	165,016	3.399	£935,832	-
	ABE+NSAI	127,893	3.382	£215,479	£2,184,412
PFS2 Weibull	NSAI	55,987	3.007	Referent	£256,648
	PAL+NSAI	152,229	3.273	Dominated	ABE+NSAI Dominant
	RIBO+NSAI	154,529	3.285	Dominated	ABE+NSAI Dominant
	ABE+NSAI	129,528	3.294	£256,648	-
PFS2 Gompertz	NSAI	55,226	3.045	Referent	£278,905
	PAL+NSAI	152,010	3.284	Dominated	ABE+NSAI Dominant
	RIBO+NSAI	154,329	3.295	Dominated	ABE+NSAI Dominant
	ABE+NSAI	129,214	3.310	£278,905	-

OS2 Exp.	NSAI	71,084	3.584	Referent	£282,820
	PAL+NSAI	165,287	3.804	Dominated	ABE+NSAI Dominant
	RIBO+NSAI	167,238	3.801	Dominated	ABE+NSAI Dominant
	ABE+NSAI	142,943	3.838	£282,820	-
OS2 Log-logistic	NSAI	57,047	3.031	Referent	£246,160
	PAL+NSAI	153,251	3.322	Dominated	ABE+NSAI Dominant
	RIBO+NSAI	155,397	3.327	Dominated	ABE+NSAI Dominant
	ABE+NSAI	130,419	3.329	£246,160	-
OS2 Gompertz	NSAI	40,049	2.350	Referent	£244,796
	PAL+NSAI	140,748	2.761	£1,250,081	-
	RIBO+NSAI	142,614	2.750	Dominated	ABE+NSAI Dominant
	ABE+NSAI	117,466	2.743	£197,123	£1,250,081
TTD1 Gompertz	NSAI	56,150	2.997	Referent	£263,915
	PAL+NSAI	151,324	3.273	Dominated	ABE+NSAI Dominant
	RIBO+NSAI	153,716	3.285	Dominated	ABE+NSAI Dominant
	ABE+NSAI	133,567	3.291	£263,915	-
TTD1 Log-normal	NSAI	56,152	2.997	Referent	£254,995
	PAL+NSAI	152,038	3.273	Dominated	ABE+NSAI Dominant
	RIBO+NSAI	154,263	3.285	Dominated	ABE+NSAI Dominant
	ABE+NSAI	130,952	3.291	£254,995	-
TTD1 Exp	NSAI	56,148	2.997	Referent	£224,015
	PAL+NSAI	136,447	3.273	Dominated	ABE+NSAI Dominant
	RIBO+NSAI	139,204	3.285	Dominated	ABE+NSAI Dominant
	ABE+NSAI	121,861	3.291	£224,015	-

TTD2: Log-logistic	NSAI	56,152	2.997	Referent	£250,352
	PAL+NSAI	152,268	3.273	Dominated	ABE+NSAI Dominant
	RIBO+NSAI	154,559	3.285	Dominated	ABE+NSAI Dominant
	ABE+NSAI	129,590	3.291	£250,352	-
TTD2 Gompertz	NSAI	56,152	2.997	Referent	£250,352
	PAL+NSAI	152,268	3.273	Dominated	ABE+NSAI Dominant
	RIBO+NSAI	154,559	3.285	Dominated	ABE+NSAI Dominant
	ABE+NSAI	129,590	3.291	£250,352	-
TTD2 vs 2nd line PFS	NSAI	56,728	2.997	Referent	£248,834
	PAL+NSAI	152,179	3.273	Dominated	ABE+NSAI Dominant
	RIBO+NSAI	154,444	3.285	Dominated	ABE+NSAI Dominant
	ABE+NSAI	129,720	3.291	£248,834	-
Treatment specific utility	NSAI	56,152	3.009	Referent	£270,232
	PAL+NSAI	152,268	3.263	Dominated	ABE+NSAI Dominant
	RIBO+NSAI	154,559	3.275	Dominated	ABE+NSAI Dominant
	ABE+NSAI	129,590	3.281	£270,232	-
PPS MONARCH 2	NSAI	56,152	3.425	Referent	£539,015
	PAL+NSAI	152,268	3.597	Dominated	£218,068
	RIBO+NSAI	154,559	3.608	£5,621,400	-
	ABE+NSAI	129,590	3.603	£412,280	£5,621,400
PFS utility MONARCH 2	NSAI	56,152	2.992	Referent	£249,002
	PAL+NSAI	152,268	3.269	Dominated	ABE+NSAI Dominant
	RIBO+NSAI	154,559	3.281	Dominated	ABE+NSAI Dominant
	ABE+NSAI	129,590	3.287	£249,002	-
PPS LOS MONARCH 3	NSAI	57,858	2.997	Referent	£248,787
	PAL+NSAI	153,562	3.273	Dominated	ABE+NSAI Dominant
	RIBO+NSAI	155,846	3.285	Dominated	ABE+NSAI Dominant
	ABE+NSAI	130,836	3.291	£248,787	-
Relative dose intensity	NSAI	55,697	2.997	Referent	£196,802
	PAL+NSAI	145,059	3.273	Dominated	ABE+NSAI Dominant
	RIBO+NSAI	141,672	3.285	Dominated	ABE+NSAI Dominant
	ABE+NSAI	113,427	3.291	£196,802	-
PFS1 utility = PFS2 utility	NSAI	56,152	3.077	Referent	£209,834
	PAL+NSAI	152,268	3.406	Dominated	ABE+NSAI Dominant
	RIBO+NSAI	154,559	3.419	Dominated	ABE+NSAI Dominant
	ABE+NSAI	129,590	3.427	£209,834	-
PPS BOLERO-2	NSAI	49,909	2.660	Referent	£199,854
	PAL+NSAI	144,078	3.113	Ex Dominated	£55,929
	RIBO+NSAI	145,475	3.138	£278,607	-
	ABE+NSAI	122,096	3.055	£183,093	£278,607

Full surrogacy	NSAI	56,152	2.997	Referent	£156,794
	PAL+NSAI	159,387	3.633	Ex Dominated	£70,232
	RIBO+NSAI	162,269	3.674	£156,794	-
	ABE+NSAI	133,339	3.481	Ex Dominated	£150,253
Utility source EQ5D-5L	NSAI	56,152	2.997	Referent	£250,352
	PAL+NSAI	152,268	3.273	Dominated	ABE+NSAI Dominant
	RIBO+NSAI	154,559	3.285	Dominated	ABE+NSAI Dominant
	ABE+NSAI	129,590	3.291	£250,352	-
Diarrhoea Hosp. and loperamide	NSAI	56,196	2.997	Referent	£251,371
	PAL+NSAI	152,320	3.273	Dominated	ABE+NSAI Dominant
	RIBO+NSAI	154,648	3.285	Dominated	ABE+NSAI Dominant
	ABE+NSAI	129,933	3.291	£251,371	-

4.4.3 ERG preferred assumptions and scenario analyses

Table 36 below summarises ERG preferred assumptions and scenario analyses, as discussed earlier in this report.

Table 36 ERG preferred assumptions and scenarios (NB. changes to base case in bold)

	Company base case	ERG preferred and scenarios	ERG comments
Decision problem			
Population	HR+/HER2- untreated advanced breast cancer (median age 63 at baseline)	No change	As per scope
Comparators	PAL+NSAI and RIBO+NSAI	No change	As per scope. We also report NSAI, as this is used in the model for reference
Model structure			
Health states & transitions	PFS1, PFS2, PPS, Death	No change	The model structure is appropriate for the decision problem and NICE reference case. It is also consistent with the ribociclib model (TA496)
Time horizon	35 years (lifetime)	No change	
Cycle length	Monthly, half cycle correction	No change	
Discount rates	3.5% per year costs & effects	No change	
Time to first progression			
Interval-censoring	IC-adjustment applied	Scenario: no IC adjustment	IC-adjustment for potential bias due to delayed identification of progression
Baseline adjustment	No baseline covariates	Scenario: baseline covariates included	Adjusts for imbalance in important prognostic factors (see 4.3.4.2.1)
TTP1 extrapolation	NSAI and ABE+NSAI exponential survival curves, joint fit to MONARCH 3 data	**ABE+NSAI estimated relative to NSAI with NMA1 PFS HR**	Exponential has best fit with a plausible extrapolation. But more reliable to use same method for ABE+NSAI curve as for comparators
		Scenario: ABE+NSAI direct fit	Company base case for comparison
		Scenario: Weibull	Alternative curves with good fit
		Scenario: Gompertz	

	Company base case	ERG preferred and scenarios	ERG comments
PFS1 HRs	RIBO+NSAI & PAL+NSAI vs. NSAI from NMA1 PFS HRs (as reported in CS Table 23)	Scenario: Use NMA1 results for all treatments in CS Fig 10	To test impact of inconsistency
		Range: Vary PFS HR for ABE+NSAI between 0.5 and 0.6	To test sensitivity to relative effects for key driver of clinical effectiveness
Death rate before first progression			
PFD1	0.2% per month NSAI 0.5% per month ABE+NSAI	**ABE+NSAI estimated relative to NSAI with NMA1 OS HR**	Fixed rate from negative binomial regression on MONARCH 3 data is appropriate. But more reliable to use same method for ABE+NSAI and comparators
		Scenario: ABE+NSAI direct fit	
Second-line survival			
PFS2 extrapolation	FUL fitted to control arm of MONARCH 2 - exponential	Scenario: Weibull	Agree with exponential as base case. Test other well-fitting distributions
		Scenario: Gompertz	
OS2 extrapolation	FUL exponential fitted to MONARCH 2 to 27.95 months CONFIRM after	**Gompertz**	Alternative assumption with better fit to observed data and clinical judgment on plausibility of extrapolation
		Scenario: log-logistic	
		Scenario: exponential +CONFIRM	
Source for PFS2 and OS2	MONARCH 2	BOLERO-2	Alternative source for second-line outcomes, as in company scenario
PFS-OS surrogacy			
Median gain in OS as % of median gain in PFS	27.5%	Range: 10%, 50%, 100%	High uncertainty over surrogacy assumption (TA496)
Treatment duration			
TTD1 survival	Gamma	Scenario: Lognormal	Same as company
		Scenario: Gompertz	
		Scenario: exponential	
TTD2 survival	Exponential	Scenario: log-logistic,	Same as company
		Scenario: Gompertz	

	Company base case	ERG preferred and scenarios	ERG comments
TTD3 survival	Assumes 37% of PPS on third-line treatment	Range: 10%, 50%	Clinical advice to the ERG indicates that patients would spend less time without treatment.
Adverse events			
AE rates	CS Table 29	Range: upper and lower 95% confidence interval limits for ABE+NSAI AE rates	Given uncertainty over relative AE rates test sensitivity of results to upper and lower limits for ABE+NSAI
Utilities			
Health state utilities	PFS1: [redacted] PFS2: 0.774 (ET/targeted) PFS2: 0.661 (chemotherapy) PPS: 0.505	PFS1: [redacted] **PFS2: 0.690** (ET/targeted) PFS2: 0.577 (chemotherapy) PPS: 0.505	Apply DSU assumption about PFS2 utility from TA496 to ensure that PFS1>PFS2>PPS
		Range: PFS1 0.690, 0.774	One-way extreme value sensitivity analysis to explore uncertainty
		Range: PFS2 0.505, [redacted]	
Resource use & costs			
Drug use	Second and third line as per TA503, with additional assumption of no NSAI	Scenario: ERG clinical scenario (see Table 25 above)	Test sensitivity of results to increased use of targeted therapy at second line and chemotherapy at third-line
AE costs	Assumes cost of loperamide only for grade 3/4 diarrhoea	Scenario: Add cost of admission for grade 3/4 diarrhoea	Test sensitivity to higher AE costs for ABE+NSAI

4.4.4 Results from ERG analysis

4.4.4.1 ERG preferred assumptions

Table 37 reports the company's original base case results, the ERG's corrected company base case results and, cumulatively, a series of ERG preferred assumptions. The final part of the table (labelled 'ERG 2L drug use') represents the ERG's base case results. As can be seen, abemaciclib + NSAI remains dominant.

Table 37 Cumulative ERG assumptions – deterministic at list prices

Analysis	Treatments	Total costs	Total QALYs	Incremental ICERs (£/QALY)	Pairwise ICERs ABE vs. comparator
Company original base case	NSAI	£56,449	2.997	Referent	£250,065
	PAL+NSAI	£145,266	3.225	Dominated	ABE+NSAI dom.
	RIBO+NSAI	£148,170	3.222	Dominated	ABE+NSAI dom.
	ABE+NSAI	£129,803	3.291	£250,065	-
ERG corrected company base case	NSAI	£56,152	2.997	Referent	£250,352
	PAL+NSAI	£152,268	3.273	Dominated	ABE+NSAI dom.
	RIBO+NSAI	£154,559	3.285	Dominated	ABE+NSAI dom.
	ABE+NSAI	£129,590	3.291	£250,352	-
+ ABE+NSAI TTP1 from NMA	NSAI	£56,152	2.997	Referent	£341,663
	ABE+NSAI	£130,514	3.215	£341,663	-
	PAL+NSAI	£152,268	3.273	Ext. dom.	£376,720 (SW)
	RIBO+NSAI	£154,559	3.285	£343,915	£343,915 (SW)
+ ABE+NSAI PFD1 from NMA	NSAI	£56,152	2.997	Referent	£289,982
	PAL+NSAI	£152,268	3.273	Dominated	ABE+NSAI dom.
	ABE+NSAI	£138,597	3.282	£289,982	-
	RIBO+NSAI	£154,559	3.285	£4,909,402	£4,909,402 (SW)
+ OS2 Gompertz	NSAI	£40,049	2.350	Referent	£208,333
	RIBO+NSAI	£142,614	2.750	Dominated	ABE+NSAI dom.
	PAL+NSAI	£140,748	2.761	Dominated	ABE+NSAI dom.
	ABE+NSAI	£127,062	2.768	£208,333	-
+ PFS2 utility 0.69 (TA496 final value)	NSAI	£40,049	2.283	Referent	£192,356
	RIBO+NSAI	£142,614	2.719	Dominated	ABE+NSAI dom.
	PAL+NSAI	£140,748	2.727	Dominated	ABE+NSAI dom.
	ABE+NSAI	£127,062	2.735	£192,356	-
+ ERG 2L drug use	NSAI	£47,230	2.318	Referent	£195,730
	PAL+NSAI	£146,607	2.738	Dominated	ABE+NSAI dom.
	RIBO+NSAI	£148,784	2.752	Dominated	ABE+NSAI dom.
	ABE+NSAI	£133,041	2.757	£195,730	-

SW = South West quadrant of the cost-effectiveness plane (ABE+NSAI less expensive and less effective than comparator).

Table 38 reports the results of the ERG's scenario analyses.

Table 38 ERG preferred assumptions - deterministic

ERG scenario		Treatment	Total Costs (£)	Total QALYs	Incremental ICER (£/QALY)	Pairwise ICERs ABE+NSAI vs. comparator
ERG preferred		NSAI	£45,359	2.283	Referent	£190,838
		RIBO+NSAI	£147,369	2.720	Dominated	ABE+NSAI dom.
		PAL+NSAI	£145,556	2.728	Dominated	ABE+NSAI dom.
		ABE+NSAI	£131,753	2.736	£190,838	-
1	**Not IC adjusted**	NSAI	£47,230	2.318	Referent	£195,730
		PAL+NSAI	£146,607	2.738	Dominated	ABE+NSAI Dominant
		RIBO+NSAI	£148,784	2.752	Dominated	ABE+NSAI Dominant
		ABE+NSAI	£133,041	2.757	£195,730	-
2	**IC and baseline adjusted**	NSAI	£48,905	2.426	Referent	£210,805
		ABE+NSAI	£134,855	2.833	£210,805	-
		RIBO+NSAI	£155,116	2.868	Dominated	£585,195
		PAL+NSAI	£153,993	2.868	£552,743	£552,743
3	**TTP1 - Joint model (M3)**	NSAI	£47,230	2.318	Referent	£156,923
		PAL+NSAI	£146,607	2.738	Dominated	ABE+NSAI Dominant
		RIBO+NSAI	£148,784	2.752	Dominated	ABE+NSAI Dominant
		ABE+NSAI	£132,721	2.863	£156,923	-
4	**TTP1 - Weibull**	NSAI	£47,409	2.341	Referent	£189,086
		PAL+NSAI	£149,458	2.778	Dominated	ABE+NSAI Dominant
		RIBO+NSAI	£151,926	2.793	Dominated	ABE+NSAI Dominant
		ABE+NSAI	£133,533	2.797	£189,086	-
5	**TTP1 - Gompertz**	NSAI	£47,663	2.378	Referent	£162,135
		PAL+NSAI	£156,569	2.897	Dominated	ABE+NSAI Dominant
		ABE+NSAI	£134,402	2.913	£162,135	-
		RIBO+NSAI	£159,783	2.920	£3,801,382	£3,801,382
6	**PFS1 HRs - CS Figure 10**	NSAI	£47,230	2.318	Referent	£195,730
		PAL+NSAI	£146,572	2.734	Dominated	ABE+NSAI Dominant
		RIBO+NSAI	£148,283	2.743	Dominated	ABE+NSAI Dominant
		ABE+NSAI	£133,041	2.757	£195,730	-
7	**PFS1 HRs - ABE+NSAI 0.5**	NSAI	£47,230	2.318	Referent	£180,970
		PAL+NSAI	£146,607	2.738	Dominated	ABE+NSAI Dominant
		RIBO+NSAI	£148,784	2.752	Dominated	ABE+NSAI Dominant
		ABE+NSAI	£131,626	2.785	£180,970	-
8	**PFS1 HRs - ABE+NSAI 0.55**	NSAI	£47,230	2.318	Referent	£202,367
		PAL+NSAI	£146,607	2.738	Dominated	ABE+NSAI Dominant
		ABE+NSAI	£133,047	2.742	£202,367	-
		RIBO+NSAI	£148,784	2.752	£1,613,579	£1,613,579
9		NSAI	£47,230	2.318	Referent	£266,681
		ABE+NSAI	£131,233	2.633	Ex dom.	-

ERG scenario		Treatment	Total Costs (£)	Total QALYs	Incremental ICER (£/QALY)	Pairwise ICERs ABE+NSAI vs. comparator
	PFS1 HRs - ABE+NSAI: 0.60	PAL+NSAI RIBO+NSAI	£146,607 £148,784	2.738 2.752	Ex dom. £234,092	£146,930 £147,704
10	**PF Deaths**	NSAI ABE+NSAI PAL+NSAI RIBO+NSAI	£47,230 £124,090 £146,607 £148,784	2.318 2.631 2.738 2.752	Referent Ex dom. Ex dom. £234,092	£245,883 - £210,358 £203,691
11	**PFS2 Weibull**	NSAI PAL+NSAI RIBO+NSAI ABE+NSAI	£46,834 £146,528 £148,682 £132,964	2.323 2.736 2.750 2.755	Referent Dominated Dominated £199,503	£199,503 ABE+NSAI Dominant ABE+NSAI Dominant -
12	**PFS2 Gompertz**	NSAI PAL+NSAI RIBO+NSAI ABE+NSAI	£45,399 £146,495 £148,695 £132,931	2.326 2.737 2.751 2.756	Referent Dominated Dominated £203,688	£203,688 ABE+NSAI Dominant ABE+NSAI Dominant -
13	**OS2 Log-logistic**	NSAI PAL+NSAI ABE+NSAI RIBO+NSAI	£67,348 £161,459 £147,619 £163,906	3.031 3.294 3.302 3.311	Referent Dominated £295,768 £1,917,513	£295,768 ABE+NSAI Dominant - £1,917,513
14	**OS2 Exponential + CONFIRM**	NSAI PAL+NSAI ABE+NSAI RIBO+NSAI	£66,219 £161,692 £147,847 £164,009	2.994 3.304 3.312 3.315	Referent Dominated £256,312 £7,229,037	£256,312 ABE+NSAI Dominant - £7,229,037
15	**BOLERO 2 PFS2 & OS2**	NSAI PAL+NSAI ABE+NSAI RIBO+NSAI	£59,501 £148,733 £135,128 £149,577	2.652 3.091 3.103 3.118	Referent Dominated £167,526 £992,631	£167,526 ABE+NSAI Dominant - £992,631
16	**OS/PFS surrogacy - 10%**	NSAI PAL+NSAI ABE+NSAI RIBO+NSAI	£47,230 £143,544 £129,380 £146,100	2.318 2.645 2.645 2.670	Referent Dominated £251,315 £665,323	£251,315 ABE+NSAI Dominant - £665,323
17	**OS/PFS surrogacy - 50%**	NSAI ABE+NSAI PAL+NSAI RIBO+NSAI	£47,230 £135,026 £149,108 £151,613	2.318 2.821 2.826 2.842	Referent £174,758 Ex dom. £761,947	£174,758 - £2,475,919 £761,947
18	**OS/PFS surrogacy - 100%**	NSAI PAL+NSAI ABE+NSAI RIBO+NSAI	£47,230 £151,244 £137,531 £153,909	2.318 3.016 3.022 3.035	Referent Dominated £128,251 £1,299,209	£128,251 ABE+NSAI Dominant - £1,299,209

ERG scenario		Treatment	Total Costs (£)	Total QALYs	Incremental ICER (£/QALY)	Pairwise ICERs ABE+NSAI vs. comparator
19	**TTD1 lognormal**	NSAI PAL+NSAI ABE+NSAI RIBO+NSAI	£47,230 £151,014 £138,461 £153,613	2.318 3.016 3.022 3.035	Referent Dominated £129,571 £1,202,036	£129,571 ABE+NSAI Dominant - £1,202,036
20	**TTD1 Gompertz**	NSAI PAL+NSAI ABE+NSAI RIBO+NSAI	£47,229 £150,300 £140,684 £153,066	2.318 3.016 3.022 3.035	Referent Dominated £132,731 £982,245	£132,731 ABE+NSAI Dominant - £982,245
21	**TTD1 exponential**	NSAI PAL+NSAI ABE+NSAI RIBO+NSAI	£47,227 £135,423 £127,478 £138,554	2.318 3.016 3.022 3.035	Referent Dominated £113,977 £878,702	£113,977 ABE+NSAI Dominant - £878,702
22	**TTD2 log-logistic**	NSAI PAL+NSAI ABE+NSAI RIBO+NSAI	£47,230 £151,244 £137,531 £153,909	2.318 3.016 3.022 3.035	Referent Dominated £128,251 £1,299,209	£128,251 ABE+NSAI Dominant - £1,299,209
23	**TTD2 Gompertz**	NSAI PAL+NSAI ABE+NSAI RIBO+NSAI	£47,230 £151,244 £137,531 £153,909	2.318 3.016 3.022 3.035	Referent Dominated £128,251 £1,299,209	£128,251 ABE+NSAI Dominant - £1,299,209
24	**TTD3 - 10%**	NSAI PAL+NSAI RIBO+NSAI ABE+NSAI	£44,723 £144,090 £146,254 £130,571	2.318 2.738 2.752 2.757	Referent Dominated Dominated £195,815	£195,815 ABE+NSAI Dominant ABE+NSAI dom. -
25	**TTD3 - 50%**	NSAI PAL+NSAI RIBO+NSAI ABE+NSAI	£48,437 £147,818 £150,002 £134,231	2.318 2.738 2.752 2.757	Referent Dominated Dominated £195,689	£195,689 ABE+NSAI Dominant ABE+NSAI Dominant -
26	**AE rates diarrhoea**	NSAI PAL+NSAI RIBO+NSAI ABE+NSAI	£48,437 £147,818 £150,002 £134,231	2.318 2.738 2.752 2.757	Referent Dominated Dominated £195,689	£195,689 ABE+NSAI Dominant ABE+NSAI Dominant -
27	**AE rates leukopenia**	NSAI PAL+NSAI RIBO+NSAI ABE+NSAI	£48,437 £147,818 £150,002 £134,234	2.318 2.738 2.752 2.757	Referent Dominated Dominated £195,696	£195,696 ABE+NSAI Dominant ABE+NSAI Dominant -
28		NSAI	£48,437	2.318	Referent	£195,707

ERG scenario		Treatment	Total Costs (£)	Total QALYs	Incremental ICER (£/QALY)	Pairwise ICERs ABE+NSAI vs. comparator
	AE rates neutropenia	PAL+NSAI	£147,818	2.738	Dominated	ABE+NSAI Dominant
		RIBO+NSAI	£150,002	2.752	Dominated	ABE+NSAI Dominant
		ABE+NSAI	£134,238	2.757	£195,707	-
29	**Utility PFS1 0.69**	NSAI	£48,437	2.264	Referent	£213,952
		PAL+NSAI	£147,818	2.648	Dominated	ABE+NSAI Dominant
		RIBO+NSAI	£150,002	2.661	Dominated	ABE+NSAI Dominant
		ABE+NSAI	£134,231	2.665	£213,952	-
30	**Utility PFS1 0.774**	NSAI	£48,437	2.398	Referent	£173,864
		PAL+NSAI	£147,818	2.870	Dominated	ABE+NSAI Dominant
		RIBO+NSAI	£150,002	2.886	Dominated	ABE+NSAI Dominant
		ABE+NSAI	£134,231	2.891	£173,864	-
31	**Utility PFS2 (ET/targeted) 0.505**	NSAI	£48,437	2.185	Referent	£169,191
		PAL+NSAI	£147,818	2.672	Dominated	ABE+NSAI Dominant
		RIBO+NSAI	£150,002	2.688	Dominated	ABE+NSAI Dominant
		ABE+NSAI	£134,231	2.692	£169,191	-
32	**Utility PFS2 (ET/targeted) 0.724**	NSAI	£48,437	2.343	Referent	£201,489
		PAL+NSAI	£147,818	2.750	Dominated	ABE+NSAI Dominant
		RIBO+NSAI	£150,002	2.764	Dominated	ABE+NSAI Dominant
		ABE+NSAI	£134,231	2.769	£201,489	-
33	**Utility PFS2 (chemotherapy) 0.505**	NSAI	£48,437	2.301	Referent	£191,792
		PAL+NSAI	£147,818	2.729	Dominated	ABE+NSAI Dominant
		RIBO+NSAI	£150,002	2.744	Dominated	ABE+NSAI Dominant
		ABE+NSAI	£134,231	2.748	£191,792	-
34	**Utility PFS2 (chemo) 0.724**	NSAI	£48,437	2.354	Referent	£204,158
		PAL+NSAI	£147,818	2.755	Dominated	ABE+NSAI Dominant
		RIBO+NSAI	£150,002	2.769	Dominated	ABE+NSAI Dominant
		ABE+NSAI	£134,231	2.774	£204,158	-
35	**Second and third line therapies**	NSAI	£41,154	2.283	Referent	£191,941
		RIBO+NSAI	£143,612	2.720	Dominated	ABE+NSAI Dominant
		PAL+NSAI	£141,758	2.728	Dominated	ABE+NSAI Dominant
		ABE+NSAI	£128,047	2.736	£191,941	-
36	**Hospitalisation for diarrhoea**	NSAI	£48,482	2.318	Referent	£196,371
		PAL+NSAI	£147,871	2.738	Dominated	ABE+NSAI Dominant
		RIBO+NSAI	£150,092	2.752	Dominated	ABE+NSAI Dominant
		ABE+NSAI	£134,574	2.757	£196,371	-

5 END OF LIFE

The CS does not present a justification for NICE's end of life criteria to be applied.

6 INNOVATION

The company provides a justification for abemaciclib to be considered a treatment innovation on the following basis:

- Abemaciclib delays disease progression and thus the need for cytotoxic chemotherapy to be given. Expert clinical opinion to the ERG is that the increase in PFS is clinically meaningful.
- Abemaciclib has a favourable safety profile which permits continuous dosing. The CS notes that palbociclib and ribociclib are associated with higher levels of neutropenia which requires regular blood count monitoring and treatment gaps at the end of each 21 day cycle. Expert clinical advice to the ERG is that reduced neutropenia-associated myelosupression would be a minor advantage when choosing a between abemaciclib and palbociclib / ribociclib.

7 DISCUSSION

7.1 Summary of clinical effectiveness issues

The MONARCH 3 trial showed a gain of [redacted] months in median PFS for the combination of abemaciclib and NSAI compared to NSAI alone. This is regarded to be a clinically meaningful benefit and is in-keeping with PFS gains for the other CDK 4/6 inhibitors ribociclib (and NSAI) (median difference 9.3 months[19, 29]) and palbociclib (and NSAI) (median difference 13.1 months[21, 49]). The indirect comparison of these treatments showed no statistically significant differences between them.

Abemaciclib can therefore be considered similar in effects to existing NICE recommended treatments in delaying cancer progression, one of the key treatment goals for patients with advanced breast cancer. The effect of abemaciclib on overall survival is currently unclear, as the duration of follow-up is not yet long enough to have measured the required number of events (deaths) needed for the analysis (the estimated study completion date is April 2020). A similar lack of follow-up of survival also applies to the palbociclib and ribociclib pivotal phase III trials. Thus, the clinical effectiveness of these CDK 4/6 inhibitors in terms of overall

survival is uncertain, hence the need for the alternative approach to economic modelling used by the company (which used a fixed-pay model to include subsequent treatment lines).

Abemaciclib can be considered to have a reasonable safety profile. Notably, grade 3/4 diarrhoea was higher for patients taking abemaciclib than it was in the trials of palbociclib and ribociclib. Incidence of neutropenia and leukopenia was high for all three of the CDK4/6 inhibitors, but particularly so for palbociclib and ribociclib. Diarrhoea can impair quality of life, though is commonly short-lived and can be managed.

MONARCH 3 was a multi-national trial with only a small number of patients from the UK participating. Whilst the patient population in the trial may be generalisable to the UK, it should be noted that around 40% of patients in the trial presented with de novo advanced breast cancer. This is a higher percentage than is commonly experienced in the UK (incidence in the range 10%-15%). This was also the case in the comparator trials of palbociclib and ribociclib. In NICE TA495 it was noted that a difference in treatment effect between patients with recurrent advanced breast cancer and patients with newly diagnosed advanced breast cancer would be unlikely. One of the expert clinical advisors to the ERG noted that patients with de novo advanced breast cancer could be considered to have biologically different disease (due to absence of prior hormonal therapy). Whether this would modify treatment effects is unclear. [redacted]

7.2 Summary of cost effectiveness issues

The company's base case results (all drugs at list price) suggests that ABE+NSAI is marginally more effective and less expensive than the comparators PAL+NSAI and RIBO+NSAI. Compared with NSAI monotherapy, ABE+NSAI had an estimated ICER of around £250,000 per QALY gained. This result was quite consistent across the company's scenario analyses, and our results were similar, for our preferred set of assumptions and across a range of scenario analyses. The absolute difference in QALYs between the CDK 4/6 inhibitors was very small, and the ranking of abemaciclib, ribociclib and palbociclib did change between scenarios. However, as the company note, the lower costs of abemaciclib are driven by a shorter time on treatment with ABE+NSAI. We note that this difference is based on weak evidence, as hazard ratios between treatments were estimated from

reported median time to discontinuation. Another aspect of the economic analysis that was subject to uncertainty and may not be fully represented in the model is adverse events: the assumed QALY loss with the included events was low, due to small disabilities and durations assumed. Exploration of uncertainty around the model results was hampered by model run time.

8 REFERENCES

1. International Agency for Research on Cancer WHO. GLOBOCAN 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012. Available at: http://globocan.iarc.fr/Pages/fact_sheets_population.aspx Accessed: 23.02.18. 2012.
2. Cancer Research UK. Breast Cancer Statistics. Breast cancer incidence (invasive) Available at: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer#heading-Zero Accessed: 23.02.18. 2014.
3. NICE. Resource Impact Template: Palbociclib with an aromatase inhibitor for previously untreated, hormone receptor-positive, HER2-negative, locally advanced or metastatic breast cancer (TA495) And Ribociclib with an aromatase inhibitor for previously untreated, hormone receptor-positive, HER2-negative, locally advanced or metastatic breast cancer (TA496). 2017.
4. Office for National Statistics. Population Estimates (Mid 2016). Secondary Population Estimates (Mid 2016). 2017. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates.
5. Office for National Statistics. Cancer registration statistics, England. 2016: Final. Secondary Cancer registration statistics, England. 2016: Final. 2018. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancerregistrationstatisticsengland/final2016.
6. NICE. Abemaciclib with an aromatase inhibitor for untreated advanced hormone receptor-positive, HER2-negative breast cancer: Final Scope [ID1227] Available at: https://www.nice.org.uk/guidance/indevelopment/gid-ta10262/documents. 2018.
7. Cancer Research UK. Breast cancer incidence statistics. http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/incidence-invasive. Accessed 29.05.2018. 2014.
8. Cardoso F, Costa A, Senkus E, et al. 3rd ESO-ESMO International Consensus Guidelines for Advanced Breast Cancer (ABC 3). Annals of oncology : official journal of the European Society for Medical Oncology 2017;**28**(12):3111.
9. Breast Cancer Care. Facts and Statistics 2018. Available at: https://www.breastcancercare.org.uk/about-us/media/press-pack-breast-cancer-awareness-month/facts-statistics Accessed: 05.03.18. 2018.
10. Walker MS, Hasan M, Yim YM, et al. Retrospective study of the effect of disease progression on patient reported outcomes in HER-2 negative metastatic breast cancer patients. Health and quality of life outcomes 2011;**9**:46.
11. NICE. Ribociclib with an aromatase inhibitor for previously untreated, hormone receptor-positive, HER2-negative, locally advanced or metastatic breast cancer [TA496]. Available at: https://www.nice.org.uk/guidance/ta496 (Accessed 27.02.18). 2017.
12. NICE. Palbociclib with an aromatase inhibitor for previously untreated, hormone receptor-positive, HER2-negative, locally advanced or metastatic breast cancer [TA495]. Available at: https://www.nice.org.uk/guidance/ta495 (Accessed 27.02.18). 2017.
13. Goetz MP, Toi M, Campone M, et al. MONARCH 3: Abemaciclib As Initial Therapy for Advanced Breast Cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2017;**35**(32):3638-46.
14. EMA. Verzenios (abemaciclib) - Summary of Product Characteristics.
15. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). European journal of cancer (Oxford, England : 1990) 2009;**45**(2):228-47.
16. Sledge GW, Jr., Toi M, Neven P, et al. MONARCH 2: Abemaciclib in Combination With Fulvestrant in Women With HR+/HER2- Advanced Breast Cancer Who Had

Progressed While Receiving Endocrine Therapy. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2017;**35**(25):2875-84.
17. Padhiar J, Eaton J. Network meta-analysis of interventions for advanced or metastatic breast cancer patients comparable to MONARCH 2. Final report. Abingdon: ICON, 2018.
18. Ward K, Quigley J, O'Rourke J. Clinical efficacy and safety of treatments in advanced or metastatic breast cancer of relevance to the MONARCH 2 trial: a systematic literature review. Report with updated results. Update 2. Abingdon: ICON, 2018.
19. Hortobagyi GN, Stemmer SM, Burris HA, et al. Ribociclib as First-Line Therapy for HR-Positive, Advanced Breast Cancer. New England Journal of Medicine 2016;**375**(18):1738-48.
20. Finn RS, Crown JP, Lang I, et al. The cyclin-dependent kinase 4/6 inhibitor palbociclib in combination with letrozole versus letrozole alone as first-line treatment of oestrogen receptor-positive, HER2-negative, advanced breast cancer (PALOMA-1/TRIO-18): a randomised phase 2 study. The Lancet Oncology 2015;**16**(1):25-35.
21. Finn RS, Martin M, Rugo HS, et al. Palbociclib and Letrozole in Advanced Breast Cancer. New England Journal of Medicine 2016;**375**(20):1925-36.
22. Robertson JF, Llombart-Cussac A, Rolski J, 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.
23. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]: The Cochrane Collaboration, 2011.
24. Dias S, Welton NJ, Sutton AJ, et al. NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials. Available at: https://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0088912/pdf/PubMedHealth_PMH0088912.pdf. 2011.
25. Dias S, Welton NJ, Sutton AJ, et al. NICE DSU Technical Support Document 3: Heterogeneity: subgroups, meta-regression, bias and bias-adjustment. Availalble at: https://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0089907/pdf/PubMedHealth_PMH0089907.pdf. 2011.
26. Dias S, Welton NJ, Sutton AJ, et al. NICE DSU Technical Support Document 4: Inconsistency in networks of evidence based on randomised controlled trials, 2011. Available at: http://nicedsu.org.uk/wp-content/uploads/2016/03/TSD4-Inconsistency.final_.15April2014.pdf. 2011.
27. Woods BS, Hawkins N, Scott DA. Network meta-analysis on the log-hazard scale, combining count and hazard ratio statistics accounting for multi-arm trials: a tutorial. BMC medical research methodology 2010;**10**:54.
28. Guyot P, Ades AE, Ouwens MJ, et al. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC medical research methodology 2012;**12**:9.
29. Hortobagyi GN, Stemmer SM, Burris HA, et al. Updated results from MONALEESA-2, a phase III trial of first-line ribociclib plus letrozole versus placebo plus letrozole in hormone receptor-positive, HER2-negative advanced breast cancer. Annals of oncology : official journal of the European Society for Medical Oncology 2018;**29**(7):1541-47.
30. Jansen JP. Network meta-analysis of survival data with fractional polynomials. BMC medical research methodology 2011;**11**(1):61.
31. National Institute for Health and Care Excellence. Cabozantinib for previously treated advanced renal cell carcinoma. Technology appraisal guidance (TA463). London: NICE, 2017.

32. National Institute for Health and Care Excellence. Tivozanib for treating advanced renal cell carcinoma (TA512). . London, 2018.
33. MRC Biostatistics Unit. DIC: Deviance Information Criteria. Secondary DIC: Deviance Information Criteria 2017. https://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-dic/.
34. Cocks K, King MT, Velikova G, et al. Evidence-based guidelines for determination of sample size and interpretation of the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2011;**29**(1):89-96.
35. Das R, Cope S, Ouwens M, et al. Economic evaluation of fulvestrant 500 mg versus generic nonsteroidal aromatase inhibitors in patients with advanced breast cancer in the United Kingdom. Clinical therapeutics 2013;**35**(3):246-60.e5.
36. Polanyi Z, Dale P, Taylor M, et al. A Cost Effectiveness Analysis of Everolimus Plus Exemestane Compared to Chemotherapy Agents for the Treatment of ER+ HER2- Metastastic Breast Cancer in the United Kingdom. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research 2014;**17**(7):A634.
37. Lloyd A, Nafees B, Narewska J, et al. Health state utilities for metastatic breast cancer. British Journal of Cancer 2006;**95**(6):683-90.
38. Woods B, Sideris E, Palmer S, et al. NICE DSU Technical Support Document 19. Partitioned Survival Analysis for Decision Modelling in Health Care: A Critical Review.[Available from http://www.nicedsu.org.uk]. 2017.
39. Pennington P. Ribociclib in combination with an aromatase inhibitor for previously untreated advanced or metastatic hormone receptor-positive, HER2-negative breast cancer: a review of model structure, inputs and assumptions. Report by the Decision Support Unit, 2017.
40. Riemsma R, Büyükkaramikli N, de Groot S, et al. Ribociclib in combination with an aromatase inhibitor for previously untreated advanced or metastatic hormone receptor-positive, HER2-negative breast cancer; ERG report for TA496, 2017.
41. Milla-Santos A, Milla L, Rallo L, et al. Phase III randomized trial of toremifene vs tamoxifen in hormonodependant advanced breast cancer. Breast cancer research and treatment 2001;**65**(2):119-24.
42. Li N, Hao Y, Xie J, et al. Everolimus-Based Therapy versus Chemotherapy among Patients with HR+/HER2− Metastatic Breast Cancer: Comparative Effectiveness from a Chart Review Study. International Journal of Breast Cancer 2015;**2015**:9.
43. Piccart M, Hortobagyi GN, Campone M, et al. Everolimus plus exemestane for hormone-receptor-positive, human epidermal growth factor receptor-2-negative advanced breast cancer: overall survival results from BOLERO-2. Annals of oncology 2014;**25**(12):2357-62.
44. Yardley DA, Noguchi S, Pritchard KI, et al. Everolimus plus exemestane in postmenopausal patients with HR+ breast cancer: BOLERO-2 final progression-free survival analysis. Advances in therapy 2013;**30**(10):870-84.
45. Di Leo A, Jerusalem G, Petruzelka L, et al. Final overall survival: fulvestrant 500 mg vs 250 mg in the randomized CONFIRM trial. Journal of the National Cancer Institute 2014;**106**(1):djt337.
46. Di Leo A, Jerusalem G, Torres R, et al. First-line vs second-line fulvestrant for hormone receptor-positive advanced breast cancer: A post-hoc analysis of the CONFIRM study. The Breast 2018;**38**:144-49.
47. Li N, Hao Y, Xie J, et al. Everolimus-Based Therapy versus Chemotherapy among Patients with HR. International journal of breast cancer 2015;**2015**.

48. Finn RS, Crown JP, Ettl J, et al. Efficacy and safety of palbociclib in combination with letrozole as first-line treatment of ER-positive, HER2-negative, advanced breast cancer: expanded analyses of subgroups from the randomized pivotal trial PALOMA-1/TRIO-18. Breast cancer research : BCR 2016;**18**(1):67.
49. Rugo H, Finn R, Dieras V, et al. Abstract P5-21-03: Palbociclib (PAL) + letrozole (LET) as first-line therapy in estrogen receptor–positive (ER+)/human epidermal growth factor receptor 2–negative (HER2−) advanced breast cancer (ABC): Efficacy and safety updates with longer follow-up across patient subgroups. Cancer Research 2018;**78**(4 Supplement):P5-21-03-P5-21-03.
50. Updated results from MONALEESA-2, a phase 3 trial of first-line ribociclib + letrozole in hormone receptor-positive (HR+), HER2-negative (HER2–), advanced breast cancer (ABC). San Antonio Breast Cancer Symposium; 2017.
51. Fleeman N, Stainthorpe A, Bagust A, et al. Palbociclib in combination with an aromatase inhibitor for previously untreated metastatic, hormone receptor-positive, HER2-negative breast cancer. ERG report for TA495: University of Liverpool, 2016.
52. NICE. NICE position statement on the EQ-5D-5L. Available at: https://euroqol.org/nice-position-statement-on-the-eq-5d-5l/. 2017.
53. Peasgood T, Ward SE, Brazier J. Health-state utility values in breast cancer. Expert review of pharmacoeconomics & outcomes research 2010;**10**(5):553-66.
54. Hudgens S. A Method of Deriving Estimates of Health State Utilities for Patients With Metastatic Breast Cancer. Journal of Clinical Pathways 2016;**2**(5):39-45.
55. Swinburn P, Lloyd A, Nathan P, et al. Elicitation of health state utilities in metastatic renal cell carcinoma. Current medical research and opinion 2010;**26**(5):1091-6.
56. NICE. Pixantrone monotherapy for treating multiply relapsed or refractory aggressive non-Hodgkin's B‑cell lymphoma. Available at: https://www.nice.org.uk/guidance/ta306. 2014.
57. NICE. Fulvestrant for untreated locally advanced or metastatic oestrogen-receptor positive breast cancer [TA503]. Available at: https://www.nice.org.uk/guidance/ta503 (Accessed 27.02.18). 2018.
58. Beuselinck B, Wynendaele W, Dirix L, et al. Weekly paclitaxel versus weekly docetaxel in elderly and frail patients with metastatic breast carcinoma having failed previous anthracyclines: a randomised phase II study of the Belgian Society of Medical Oncology. Cancer Research 2009;**69**(2 Supplement):6115.
59. Kaufman PA, Awada A, Twelves C, et al. Phase III open-label randomized study of eribulin mesylate versus capecitabine in patients with locally advanced or metastatic breast cancer previously treated with an anthracycline and a taxane. Journal of Clinical Oncology 2015;**33**(6):594-601.
60. Stenbygaard LE, Herrstedt J, Thomsen JF, et al. Toremifene and tamoxifen in advanced breast cancer - a double-blind cross-over trial. Breast cancer research and treatment 1993;**25**(1):57-63.
61. NICE. Everolimus with exemestane for treating advanced breast cancer after endocrine therapy [TA421]. Available at: https://www.nice.org.uk/guidance/ta421 (Accessed 27.02.18). 2016.
62. NICE. Fulvestrant for the treatment of locally advanced or metastatic breast cancer [TA239]. Available at: https://www.nice.org.uk/guidance/TA239 (Accessed 09.03.18). 2011.
63. Leo A, Jerusalem G, Petruzelka L, et al. Results of the CONFIRM phase III trial comparing fulvestrant 250 mg with fulvestrant 500 mg in postmenopausal women with estrogen receptor-positive advanced breast cancer. Journal of clinical oncology 2010; 28(30). http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/798/CN-00761798/frame.html.
64. Baselga J, Campone M, Piccart M, et al. Everolimus in postmenopausal hormone-receptor-positive advanced breast cancer. New England journal of medicine 2012;

366(6). http://onlinelibrary.wiley.com/o/cochrane/clcentral/articles/298/CN-00804298/frame.html

9 APPENDICES

9.1 ERG critical appraisal of the first-line treatment NMA

For a description and detailed critique of this NMA see section 3.1.7 of this report.

Checklist	**Response yes/no**
Does the CS present an NMA?	Yes
Are the NMA results used to support the evidence for the clinical effectiveness of the intervention	Yes
Are the NMA results used to support the evidence for the cost-effectiveness of the intervention	Yes
Homogeneity	
1. Is homogeneity considered?	Yes (CS section B.2.9.3 and Appendix D.1.5).
2. Are the studies homogenous in terms of patient characteristics and study design?	Unclear. The CS identifies some areas of heterogeneity (CS section B.2.9.3) but the effect of these on results is unclear.
3. Is the method used to determine the presence of statistical heterogeneity adequate? (e.g. Chi-squared test, I-squared statistic)	Not reported
4. If the homogeneity assumption is not satisfied, is clinical or methodological homogeneity across trials in each set involved in the indirect comparison investigated by an adequate method? (e.g. sub group analysis, sensitivity analysis, meta-regression)	No. The CS states methods such as meta-regression were not considered feasible due to limited study availability.
Similarity	
1. Is the assumption of similarity stated?	No
2. Have they justified their assumption?	N/A
Consistency	
1. Does the analysis explicitly assess consistency?	No The CS notes that none of the comparisons in which direct and indirect evidence is available involved scoped comparators (section B.2.9.5). The ERG notes indirect and direct evidence is available for some of the non-scoped comparators included in the OS, ORR and CR (but not PFS) networks. Clarification response A16 states that a consistency assessment was undertaken but results were not presented as the only closed loops involved

		comparisons not relevant to this appraisal.
	2. Does the method described include a description of the analyses/ models/ handling of potential bias/ inconsistency/ analysis framework?	No
	3. Are patient or trial characteristics compared between direct and indirect evidence trials?	No
	4. If Q3 is yes, and inconsistency is reported, is this accounted for by not combining the direct and indirect evidence?	N/A

Criterion	**ERG assessment**
ITC purpose	
1. Are the NMA results used to support the evidence for the clinical effectiveness of the intervention?	Yes, for the indirect comparison of abemaciclib vs ribociclib and vs palbociclib via a common comparator (NSAI), although the results for the indirect comparisons of abemaciclib vs ribociclib and palbociclib were not presented (provided in clarification question response A12).
2. Are the NMA results used to support the evidence for the cost-effectiveness of the intervention?	Yes. The NMA results for the outcomes of PFS and OS are used to inform the economic model.
Evidence selection	
3. Are inclusion/exclusion criteria adequately reported?	Yes, CS Appendix D.1.2. These are broader than the NICE scope, to permit inclusion of non-scoped comparators, with the aim of including more data in the network.
4. Is quality of the included studies assessed?	Yes, CS Appendix Table 25 provides tabulated risk of bias assessments of all 18 included trials, using the NICE recommended criteria. The CS states that all studies were judged to be good quality with acceptable risk of bias. High risk of bias was judged for blinding as several trials were open-label. The ERG notes that the risk of bias was judged unclear in many studies for some items, including adequate randomisation, concealment of allocation, attrition, and use of ITT analysis / appropriate methods for handling missing data.
Methods – statistical model	
5. Is the statistical model described?	Yes, CS Appendix D.1.5. Further clarification on some procedures was requested by the ERG.
6. Has the choice of outcome measure used in the analysis been justified?	Yes, see 'Feasibility assessment' heading in CS Appendix D.1.3. The outcomes considered were chosen due to their relevance to the MONARCH 3 therapy setting and for the cost-effectiveness model. Outcomes included PFS, OS, ORR, CBR, and CR. PFS and OS are used in the economic model.
7. Has a structure of the network been provided?	Yes, network diagrams are provided for all outcomes, in CS Appendix D.1.3 (Figure 2 to 7).

8. Is homogeneity considered?	Yes. The CS provides a discussion of characteristics where studies were similar, and where there were some differences between studies (CS section B.2.9.3 and Appendix D.1.5).
9. Are the studies homogenous in terms of patient characteristics and study design?	Unclear. The CS identifies some areas of heterogeneity (CS section B.2.9.3) but the effect of these on results is unclear.
10. If the homogeneity assumption is not satisfied, is clinical or methodological homogeneity across trials in each set involved in the indirect comparison investigated by an adequate method? (e.g. sub group analysis, sensitivity analysis, meta-regression)	No. The CS reports that meta-regression was not considered feasible due to limited study availability. The ERG agrees with this as generally a minimum of 10 studies are required to perform meta-regression.
11. Is the assumption of similarity stated?	No.
12. Is any of the programming code used in the statistical programme provided (for potential verification)?	No. Requested by the ERG and provided in clarification question response A18.
Sensitivity analysis	
13. Does the study report sensitivity analyses?	No (stated in CS B.2.9.4).
Results	
14. Are the results of the ITC presented?	Yes, in CS section B.2.9.2. The ERG notes that results are presented for each treatment relative to the reference treatment (letrozole/anastrozole), rather than relative to the scoped comparators (ribociclib and palbociclib). These were requested from the company by the ERG and provided in clarification question response A12.
15. Does the study describe an assessment of the model fit?	Yes. CS Appendix D.1.5 describes use of the Deviance Information Criterion to assess fit of random effects and fixed effect models. The ERG requested the DIC values from the company and these were provided in clarification question response A17.
16. Has there been any discussion around the model uncertainty?	Yes in various sections. The CS notes the immaturity of the OS data and the lack of evidence to support the proportional hazards assumption (CS p.94) as potential limitations in the analysis, also the low event counts for CR (CS p. 63) and, heterogeneity in DFI and the proportion of patients with visceral metastases and the site of disease (CS p.156).
17. Are the point estimates of the relative treatment effects accompanied by some measure of variance such as confidence intervals?	Yes, credible intervals are given to accompany the point estimates.
Discussion - overall results	
18. Does the study discuss both conceptual and statistical heterogeneity?	Yes – conceptual (clinical) heterogeneity is discussed. Statistical heterogeneity is not discussed (NB. this only applies to pairwise

	comparisons in network loops where comparisons include both direct and indirect evidence).
Discussion - validity	
19. Are the results from the indirect/NMA compared, where possible, to those just using direct evidence?	No. None of the comparisons in which direct and indirect evidence is available involved scoped comparators. Indirect and direct evidence is available for some of the non-scoped comparators included in the network.

9.2 ERG critical appraisal of the second-line treatment NMA

Information used to complete this checklist is taken from a confidential separate report of the second-line treatment NMA,[17] and a separate confidential report of the associated SLR of second-line treatments in advanced or metastatic breast cancer of relevance to the MONARCH 2 trial,[18] provided to the ERG by the company in response to a clarification question (A21).

Checklist	**Response yes/no**
Does the MS present an NMA?	Yes
Are the NMA results used to support the evidence for the clinical effectiveness of the intervention	No
Are the NMA results used to support the evidence for the cost-effectiveness of the intervention	Yes (not directly for abemaciclib, but used to provide comparative evidence of second-line endocrine treatments for the economic model)
Homogeneity	
1. Is homogeneity considered?	Yes (NMA report sections 3.9 and 4.2)
2. Are the studies homogenous in terms of patient characteristics and study design?	No. Although there were similarities for a number of characteristics, the company states 'Ultimately, the comparability of MONARCH 2 to the identified studies is questionable' (p. 31)
3. Is the method used to determine the presence of statistical heterogeneity adequate? (e.g. Chi-squared test, I-squared statistic)	Not reported
4. If the homogeneity assumption is not satisfied, is clinical or methodological homogeneity across trials in each set involved in the indirect comparison investigated by an adequate method? (e.g. sub group analysis, sensitivity analysis, meta-regression)	No. The NMA report states methods such as meta-regression were not considered feasible due to limited study data availability. Only one sensitivity analysis was considered feasible.
Similarity	
1. Is the assumption of similarity stated?	No
2. Have they justified their assumption?	N/A
Consistency	
1. Does the analysis explicitly assess consistency?	Yes (NMA report section 3.8 and section 5)
2. Does the method described include a description of the analyses/ models/ handling of potential bias/ inconsistency/ analysis framework?	Yes

	3. Are patient or trial characteristics compared between direct and indirect evidence trials?	No
	4. If Q3 is yes, and inconsistency is reported, is this accounted for by not combining the direct and indirect evidence?	N/A

Criterion	ERG assessment
NMA purpose	
1. Are the NMA results used to support the evidence for the clinical effectiveness of the intervention?	No
2. Are the NMA results used to support the evidence for the cost-effectiveness of the intervention?	Yes, NMA results are used to provide comparative evidence of second-line treatments for the economic model.
Evidence selection	
3. Are inclusion/exclusion criteria adequately reported?	Yes, NMA report section 2.1. These are broader than the inclusion criteria for the MONARCH 2 trial[16] as a low volume of matching studies was anticipated.
4. Is quality of the included studies assessed?	Yes, SLR report section 4.7, using NICE recommended criteria. The SLR report states that all studies were assessed as being of good quality with an acceptable risk of bias, but notes in many studies an unclear risk of bias was assigned across multiple domains due to lack of reporting.
Methods – statistical model	
5. Is the statistical model described?	Yes, NMA report section 3 and Appendix A.
6. Has the choice of outcome measure used in the analysis been justified?	Yes, NMA report section 3.2, based on a feasibility assessment (NMA report section 3.1) and economic model requirements. Of the four outcomes included the ERG notes that only PFS and OS are used in the economic model.
7. Has a structure of the network been provided?	Yes for all outcomes, NMA report Figures 4.1, 4.3, 4.4, 4.6, 4.8
8. Is homogeneity considered?	Yes, NMA report sections 3.9 and 4.2.
9. Are the studies homogenous in terms of patient characteristics and study design?	No. Although there were similarities for a number of characteristics (age, post-menopausal status and cancer performance status), the report states 'ultimately, the comparability of MONARCH 2 to the identified studies is questionable' (NMA report page 30). The MONARCH 2 trial assessed a very specific population (HR+/HER2-, ≤ 1 prior endocrine therapy and no prior chemotherapy permitted in the advanced setting), whereas the other studies allowed prior chemotherapy in the advanced setting and some trials allowed for more than one prior endocrine therapy in the advanced setting. The proportion of patients with visceral metastases ranged from 13.5% to 100% where reported, although definitions varied (and was

	often not reported). HR+/HER2- status differed or was unknown across a number of trials.
10. If the homogeneity assumption is not satisfied, is clinical or methodological homogeneity across trials in each set involved in the indirect comparison investigated by an adequate method? (e.g. sub group analysis, sensitivity analysis, meta-regression)	No. The NMA report states methods such as meta-regression were not considered feasible due to limited study data availability. Only one sensitivity analysis was considered feasible.
11. Is the assumption of similarity stated?	No.
12. Is any of the programming code used in the statistical programme provided (for potential verification)?	Yes, NMA report Appendix E.
Sensitivity analysis	
13. Does the study report sensitivity analyses?	Yes, NMA report section 3.10. Only one sensitivity analysis was considered feasible.
Results	
14. Are the results of the NMA presented?	Yes, NMA report section 4.
15. Does the study describe an assessment of the model fit?	Yes, NMA report section 3.6.3 and Appendix A. Deviance Information Criterion (DIC) is used to assess fit of random effects and fixed effect models and for consistency / inconsistency models.
16. Has there been any discussion around the model uncertainty?	Yes, NMA report section 6. The immaturity of OS data was noted.
17. Are the point estimates of the relative treatment effects accompanied by some measure of variance such as confidence intervals?	Yes, credible intervals are given to accompany the point estimates.
Discussion - overall results	
18. Does the study discuss both conceptual and statistical heterogeneity?	Conceptual (clinical) heterogeneity is discussed. Statistical heterogeneity is not discussed
Discussion - validity	
19. Are the results from the indirect/NMA compared, where possible, to those just using direct evidence?	No. However, an inconsistency assessment was conducted.

The second-line treatment NMA was conducted to inform cost-effectiveness modelling of second-line treatments for advanced breast cancer in the “fixed pay-off” sub-model (see section 4.3.4.3 of this report for a description of how the NMA informs modelling of second-line treatment).

An SLR was conducted to identify relevant trials.[18] The search was run in December 2015, and updated in March 2017 and January 2018. This appears to be the same search that was

run for the assessment of clinical effectiveness of abemaciclib as a first-line treatment for advanced breast cancer, reported in the CS (see section 3.1.1 of this report).

9.2.1 Eligibility criteria

The aim was to set criteria to include studies similar to the MONARCH 2 trial.[16] However, the criteria were set to be broader than the population in MONARCH 2 as it was anticipated there would be a low volume of relevant evidence given that MONARCH 2 included patients with specific characteristics (women with advanced HR+, HER2-, breast cancer which had progressed on endocrine therapy, who had not received chemotherapy for advanced breast cancer).

- Intervention: abemaciclib as monotherapy or combination therapy
- Population: women with advanced breast cancer including
    - Trials where ≥50% of the trial population were HR+
    - Trials in which HER2 status of patients was not stated
    - Trials with patients who had received prior chemotherapy or >1 prior endocrine therapy in the advanced setting
- Comparators: endocrine monotherapy, chemotherapy monotherapy, targeted therapy monotherapy, combination chemotherapy, combination endocrine and targeted therapy and combination chemotherapy and targeted therapy.
- Outcomes: survival (OS and PFS), disease free-survival, response (CR, PR, SD), ORR, duration of response, CBR, disease control rate, grade 3 and 4 adverse events, and HRQoL.

A total of 29 trials met the inclusion criteria for the SLR. Of these, nine were unable to be included in the NMA because they did not include an endocrine therapy comparison and therefore could not be connected to the networks. All of these nine trials included chemotherapy treatments (e.g. paclitaxel, gemcitabine, capecitabine), thus the NMA does not compare endocrine therapy with chemotherapy treatments (as noted in CS Figure 2). In addition, one eligible trial of endocrine therapy was excluded as it could not be connected to the outcome networks.[60]

A total of 19 trials were included in the NMA as a whole, with the number of trials included in each outcome network varying (see section 9.2.2 below). The CS reports that the following treatments were not considered clinically relevant to the MONARCH 2 trial-aligned population because they are considered older therapies not commonly used, or not licensed doses:

- Letrozole 0.5mg
- Megestrol 160mg
- Megestrol 800mg
- Toremifene

The CS therefore only reports results for what it considers to be relevant treatments:

- Abemaciclib and fulvestrant (ABE+FUL)
- Anastrozole 1mg (ANAS 1)
- Anastrozole 10mg (ANAS 10)
- Letrozole 2.5mg (LTZ 2.5)
- Exemestane (EXE)
- Everolimus + exemestane (EVE+EXE)
- Fulvestrant 250mg (FUL 250)
- Fulvestrant 500mg
- Palbociclib + fulvestrant 500mg (PAL+FUL)
- Tamoxifen (TMX)

The ERG notes that some, but not all, of these treatments are included in the company's economic model (see 4.3.4.3 of this report), and that not all are recommended or have been appraised by NICE:

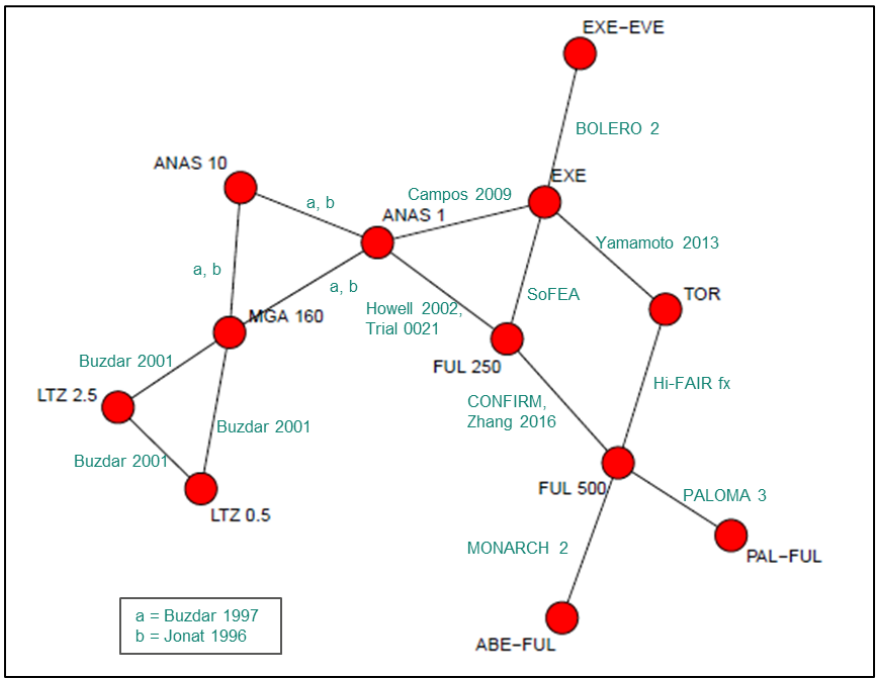
- Abemaciclib + fulvestrant has not yet been appraised by NICE; guidance is expected to be issued in summer 2019 (NICE ID1339). It is not included in the company's economic model as a second-line treatment.
- Anastrozole and letrozole are not included in the company's economic model as second-line treatments.
- Exemestane monotherapy does not appear to have been appraised by NICE in this indication. It is included in the company's economic model as a second-line treatment.
- Exemestane + everolimus is recommended by NICE (TA421[61]). It is included in the company's economic model as a second-line treatment.
- Fulvestrant 500 mg is not recommended by NICE as a second-line treatment for advanced breast cancer (NICE TA239[62]). It is used as a reference treatment in the NMA (chosen because it was the comparator arm in the MONARCH 2 trial). Fulvestrant 500mg (but not 250mg) is included in the company's economic model as a second-line treatment.

- Palbociclib + fulvestrant has not yet been appraised by NICE in this indication (appraisal currently suspended – NICE ID916). It is not included in the company's economic model as a second-line treatment.
- Although tamoxifen was eligible to be included in the NMA, no PFS or OS data were available from the single trial identified that included this treatment (NMA report Table D.1).[60] Tamoxifen is included in the company's economic model as a second-line treatment. We discussed earlier in this report (section 4.3.4.3) how the clinical effectiveness of tamoxifen as a second-line treatment has been estimated for the model.
- Although trials of chemotherapy could not be connected in the NMA, the company's economic model does include chemotherapy as a second-line treatment [specifically, capecitabine, paclitaxel and docetaxel (CS Table 35)]. The clinical effectiveness data for chemotherapy is from a retrospective chart review of 137 postmenopausal HR+/HER2- metastatic breast cancer women in community-based oncology practices in the US (CS Table 20).[42] The specific chemotherapies administered to patients in this study is not reported in the study publication. The CS did not provide a rationale for using this study in preference to any others, though did state that the study had been used to estimate the efficacy outcomes of chemotherapy in the NICE TA496 (ribociclib) (CS section B.3.2.2). We noted concerns about this study earlier in this report (section 4.3.4.3.1).

In summary, the treatments included in this NMA comprise a range of endocrine therapies, though not all of them have been recommended/appraised by NICE. The NMA does not include comparisons between endocrine therapy and chemotherapy. The treatments that are included in the economic model are exemestane + everolimus, exemestane monotherapy, fulvestrant 500 mg, tamoxifen, and chemotherapy (capecitabine, paclitaxel and docetaxel). The only results from the second-line treatment NMA that are used in the economic model are for the comparison of exemestane monotherapy with fulvestrant and the comparison of exemestane + everolimus with fulvestrant.

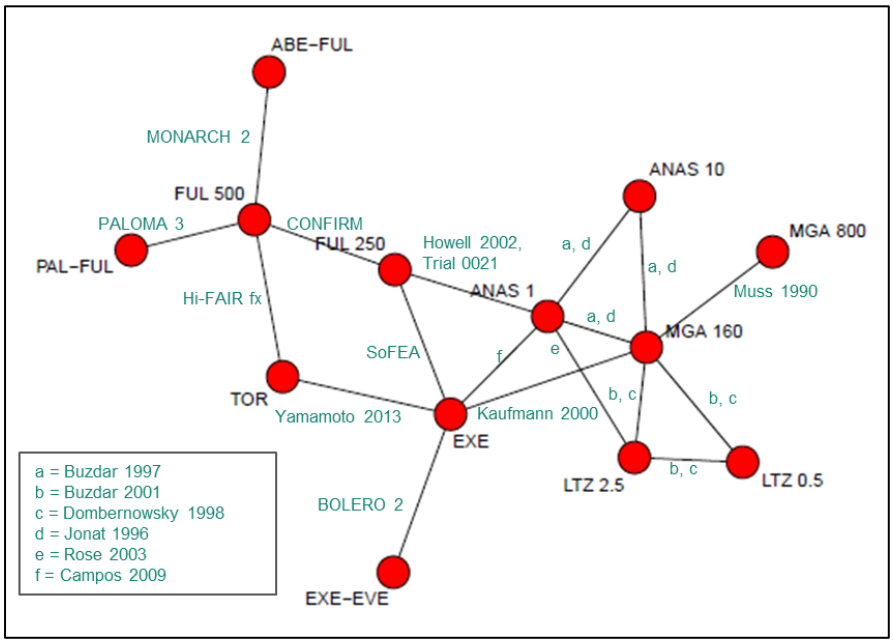
9.2.2 Evidence networks

A feasibility assessment was conducted to assess whether it was possible to construct networks for outcome measures. The following outcomes were considered relevant and feasible: PFS, OS, ORR, and CBR. Only PFS and OS are used in the economic model and therefore we focus on these outcomes in this ERG report. Network diagrams for PFS (n=14 trials) and OS (n=17 trials) are shown in Figure 6 and Figure 7 respectively.



(reproduced from Figure 4.1)[17]

Figure 6 Network diagram for PFS, second-line treatment NMA network



(reproduced from Figure 4.4)[17]

Figure 7 Network diagram for OS, second-line treatment NMA network

Fulvestrant 500mg is the reference treatment and connects abemaciclib + fulvestrant to the network. All treatments are compared pairwise to fulvestrant 500mg; there are no other treatment comparisons presented in the NMA report (though a probabilistic ranking of

treatments based on the odds of an event is given for the response outcomes of ORR and CBR in Appendix G).[17] The networks comprise comparisons that are informed by both direct and indirect evidence (closed loops) as well as comparisons only informed by indirect evidence.

9.2.3 Statistical methods

The statistical approach used is similar to that used to conduct the first-line treatment NMA (as described in more detail in section 3.1.7 of this report). In brief:

- A Bayesian generalised linear model is used, based on NICE DSU guidelines.[24]
- Fixed and random effects modelling is undertaken with selection of model according to best fit (based on DIC values). Both random effects and fixed effects model are presented for PFS, but only fixed effects results are presented for OS as there was evidence of the prior around the random effects standard deviation dominating the posterior estimates (it is not stated why). Given the observed clinical heterogeneity in the networks (see section 9.2.4 below) the ERG considers the random effects model would have been more appropriate in principle.
- Vague prior distributions were chosen for treatment and study-specific term, in accordance with DSU methodological guidance.[24]
- OpenBUGS software was used to run the analysis (the code is provided in Appendix E the NMA report). A Markov chain Monte Carlo simulator was run for 50,000 burn-in simulations with a further 100,000 simulations for convergence to the posterior distribution (Brooks-Gelman-Rubin plots).

The ERG notes that OS data are immature (median OS not reached in at least one arm) in eight of the trials, including the MONARCH 2 trial. (The final OS analysis of this trial will be conducted at 441 OS events. The estimated study completion date is February 2020.[16] However, none of the remaining seven trials included comparisons that were used in the economic model.

An inconsistency assessment was performed to determine the level of consistency between direct and indirect evidence in the NMA networks, based on the approach recommended by the NICE DSU.[26] For PFS and OS both the total residual deviance and DIC values remained similar (<5 point difference) between consistency and inconsistency models, indicating no inconsistency.

The validity of the assumption of proportional hazards of survival data was tested using the same methods as used in the first-line treatment NMA (i.e. log cumulative hazard plots, Schoenfeld residual plots, weighted residual test based on standardised Schoenfeld residuals). The NMA report states that the assumption held across the majority of trials. Where there was evidence of non-proportional hazards the potential reasons were suggested to be high levels of censoring in the tails, interval censoring for PFS and immature survival data. The ERG's interpretation is that proportional hazards do not hold for all of the trials in the NMA, with the assumption less likely to hold for OS than PFS. An NMA approach that allows for time-varying hazards should have been considered as an alternative to the approach used. The immaturity of the survival data is a particular limitation and creates significant uncertainty in the results of the OS network.

9.2.4 Heterogeneity assessment

The NMA report provides a discussion of clinical heterogeneity amongst the set of trials included in the NMA. This was based on a comparison of baseline trial characteristics, and expert clinical opinion on potential treatment effect modifiers. Tabulated study characteristics are presented in the accompanying SLR report.[18]

The NMA report identifies three areas of potential clinical heterogeneity:

- Proportion of patients with visceral involvement, ranging from 13.5% to 100%, where reported.
- Number of prior treatments for advanced breast cancer. The MONARCH 2 trial only permitted patients to have received one (or fewer) endocrine therapies, and no chemotherapy for advanced breast cancer. All of the other trials (where stated) permitted prior use of chemotherapy for advanced breast cancer and some permitted more than one endocrine therapy.
- HR/HER2 status. The majority of trials reported that HR+ patients were eligible for inclusion, however, the majority (n=14/19) of the trials did not specify HER2 status in the eligibility criteria (Table 3.1 of the NMA report[17]).

The second-line treatment NMA report states that it was not possible to conduct meta-regression to address heterogeneity due to limited study data available. The ERG concurs that this would not have been feasible. A sensitivity analysis was performed for PFS using a sub-group of patients who had not received prior chemotherapy corresponding to the ITT

population of MONARCH 2. There was only one trial reported to have provided data for this subgroup, the PALOMA 3 trial which compared palbociclib + fulvestrant vs fulvestrant. The comparison of these two trials (i.e. ABE-FUL vs PAL-FUL) was not included in the economic model.

The NMA report judges that the comparability of MONARCH 2 to the identified trials is questionable, due to the specific eligibility characteristics of MONARCH 2. However, if the inclusion criteria for the NMA had been restricted to fully match the MONARCH 2 trial there would have been very few eligible trials included. The ERG agrees with these observations. We also consider that there is a higher degree of clinical heterogeneity in the second-line treatment NMA than in the first-line NMA (section 3.1.7 of this report).

The NMA report does not state whether any statistical heterogeneity tests were performed for head-to-head pairwise comparisons.

9.2.5 Risk of bias

The NMA report does not comment on the risk of bias in the included trials.[17] The accompanying SLR report[18] provides an assessment of bias using NICE's recommended criteria. The report states that all studies were assessed as being of good quality with an acceptable risk of bias (bias that would not have a large impact on study outcomes). Across trials, the risk of bias was largely assessed as being of either low risk or unclear risk over each of criteria. The ERG notes that one of the included trials (Hi-FAIR) is missing from the risk of bias assessment.

The ERG has not performed an independent risk of bias assessment of these trials, but notes that there were very few trials (n=4) judged at high risk of bias on any one criterion. In terms of risk of selection bias, over half the trials were judged unclear for randomisation and concealment of allocation procedures (n=12 and n=11 respectively). In contrast, the majority of trials (n=17/19) were judged to have equivalent trial arms at baseline suggesting that the risk of selection bias may be low (for measured trial characteristics at least). The risk of bias associated with lack of blinding was judged low in over half the trials (n=11). The risk of bias from unexpected imbalances in drop-outs between groups was unclear in the majority of trials (n=13), as was the case for bias from missing data (n=16). The risk of bias from selective reporting of outcomes was generally low (n=12). Overall, the risk of bias was largely assessed as either low or unclear over each of the criteria.

9.2.6 Results

Brief results are presented here, for PFS and OS outcomes only.

[redacted]

[redacted]

9.2.7 Summary of the ERG's appraisal of the second-line treatment NMA

The ERG's main comments on the second-line treatment NMA are:

- The search strategy used in the SLR of second-line treatments appears to be similar to that used to identify studies for the SLR of abemaciclib as a first-line treatment for advanced breast cancer. As stated earlier in this report (section 3.1.2) the ERG considers the search strategies are fit for purpose. The ERG has not formally critically appraised the second-line treatment SLR but it is unlikely that there is a risk of bias in the identification, selection and critical appraisal of the included trials.[18]
- A range of endocrine therapies are included in the NMA. The only results from the NMA that are used in the economic model are for the comparison of exemestane monotherapy with fulvestrant and exemestane + everolimus vs fulvestrant.
- The included trials appear to be clinically heterogeneous, as acknowledged by the company. The comparability of the MONARCH 2 trial to the comparator trials is questionable due to its specific patient inclusion criteria.
- Reporting limitations means that in many studies an unclear risk of bias was assigned across multiple domains due to lack of reporting.

- The NMA methods are similar to those used for the first-line treatment NMA (i.e. based on NICE DSU technical support document 2). These are appropriate.
- However, proportional hazards do not appear to hold for all the trials included for both OS and PFS, indicating that a NMA approach that allows for time-varying hazards should have been considered as an alternative.
- OS data are immature in eight trials, including the MONARCH 2 trial. The results of the OS network should therefore be interpreted with caution.

9.3 Graphs of survival extrapolations used in model

Figure 8 Time to first progression: parametric survival estimated from MONARCH 3 (interval-censored adjusted)
Source: Company model with log-normal, log-logistic and gamma curves digitised from CS Figures 19 and 20

Figure 9 Time to first progression: company base case and NMA estimate for abemaciclib (interval-censored adjusted)
Source: Company model with log-normal, log-logistic and gamma curves digitised from CS Figures 19 and 20

Figure 10 Pre-progression death rates
Source: Company model with ERG corrections to calculation of rates for palbociclib and ribocicl

9.4 Comparison of baseline characteristics of trials used in the company's economic analysis for post-progression survival: MONARCH, BOLERO-2 and CONFIRM

As discussed earlier in this report, the post-progression survival data from the MONARCH 3 trial are immature, therefore clinical-effectiveness data were used from similar, progressed patient populations from alternative trials. Patients in the placebo + fulvestrant arm of the MONARCH 2 trial are assumed to represent patients progressing after treatment from the MONARCH 3 trial. The MONARCH 2 trial[16] inclusion criteria require patients to have progressed on one prior endocrine therapy. The OS data from the MONARCH 2 trial are also immature, so OS data from the CONFIRM trial[45, 63] are also used to inform longer-term estimates.

The ERG has explored the plausibility of the assumption that patients in the MONARCH 2 and CONFIRM trials are representative of patients progressing from MONARCH 3. We also explored this assumption in relation to the BOLERO-2 trial since this is also used in the CS to estimate post-progression survival in a scenario analysis.

The MONARCH 3 trial considered patients eligible for inclusion if they had received no systematic therapy for advanced disease, their cancer had progressed after at least 12 months following the completion of (neo)adjuvant endocrine therapy, or if they presented with de novo advanced breast cancer. This is a key difference from the other three trials, which permitted inclusion of patients who had progressed during (neo)adjuvant therapy, or less than 12 months after adjuvant treatment. These trials also permitted patients for inclusion if they progressed whilst receiving endocrine therapy for advanced breast cancer. Thus, they comprised mixed populations of patients who had progressed from the (neo)adjuvant setting (thus they were now receiving their first treatment for advanced breast cancer) or who had progressed from the advanced breast cancer setting (thus were now receiving their second treatment for advanced breast cancer). Only the patients in the latter sub-group can be considered comparable to the patients in the MONARCH 3 trial. This sub-group varied in size considerably between the trials:

- MONARCH 2 - patients receiving most recent endocrine therapy for metastatic cancer: n=256 (38%)
- BOLERO-2 - purpose of most recent treatment: treatment of advanced or metastatic cancer: n=586 (81%)

- CONFIRM - progression after first-line treatment for advanced breast cancer (>12 months after adjuvant endocrine treatment) / progression after first-line treatment for advanced breast cancer (de novo advanced breast cancer): n=343 (47%)

Thus, the MONARCH 2 trial has the lowest proportion of patients who had progressed from the advanced breast cancer setting (i.e. comparable to the patients in MONARCH 3 who progressed on treatment). The CS restricts the analysis of post-progression survival to this sub-group.

Table 39 provides a comparison of baseline characteristics of the four trials, in terms of demographic details, disease characteristics and prior treatments received. Note that these characteristics apply to the ITT populations and not for the relevant subgroups noted above. Also note that many details for the CONFIRM trial were not reported, including HER2 status, limiting our interpretation of its comparability to MONARCH 3. The trials appear generally comparable (where reported) in terms of median age, ECOG performance status, HER2 receptor status, PgR receptor status, and percentage of patients with visceral metastases (except CONFIRM where this slightly higher). There was some variation in race (with a higher percentage of white patients in BOLERO-2 compared to the MONARCH trials) and in region (a higher percentage of patients from North America and lower percentage of patients in Europe and Asia in BOLERO-2). None of the trials had quite as high a percentage of patients with measurable disease as MONARCH 3.

A recent publication of the CONFIRM trial[46] reports a post-hoc comparison of results for the sub-group of patients treated with fulvestrant first-line for advanced breast cancer (n=387) and the sub-group being treated second-line for advanced breast cancer (i.e. the sub-group of relevance to this appraisal as discussed above, n=343). A comparison of baseline characteristics between these two sub-groups showed that they were generally similar, with some exceptions relating to previous treatment with aromatase inhibitors, adjuvant antioestrogen therapy, prior chemotherapy, and bone only disease (higher in the first-line treatment sub-group). The ERG notes that the median age in the second-line treatment sub-group was 63 years (vs 58-59 years in the first-line treatment group) which is closer to the median age of patients in MONARCH 3 (63 years) that the other trials.

Table 39 Comparison of baseline characteristics of trials used in the company's economic analysis for post-progression survival: MONARCH, BOLERO-2 and CONFIRM

Baseline characteristic	First-line ABC treatment trial	Second-line ABC treatment trials		
	MONARCH 3 (n=493)[13] ABE+NSAI vs placebo+NSAI	MONARCH 2 (n=669)[16] ABE+FUL vs placebo+FUL	BOLERO-2 (n=724)[64] EVE+EXE vs placebo+EXE	CONFIRM (N=736)[45, 63] FUL 500 vs FUL 250
Age, years				
Median (range)	63 (32-88)	59 - 62 (32-91)	61 - 62 (28-93)	61
Race, n (%)				
White	288 (58%)	373 (56%)	74% - 78%	NR
Asian	148 (30%)	214 (32%)	19% - 20%	NR
Other	18 (4%)	42 (6%)	2%-3%	NR
Region, n (%)				
Europe	[redacted]	279 (42%)	275 (38%)	NR
Asia	[redacted]	212 (32%)	137 (19%)	NR
North America	[redacted]	178 (27%)	274 (38%)	NR
Other	[redacted]	0	38 (5%)	NR
ECOG performance status				
0	296 (60%)	400 (60%)	59%-60%	NR
1	197 (40%)	263 (39%)	35%-36%	NR
Receptor status, n (%)				
PgR+	382 (77%)	510 (76%)	523 (72%)	507 (69%)
PgR-	106 (22%)	140 (21%)	184 (25%)	188 (26%)
Missing / unknown	5 (1%)	19 (3%)	17 (3%)	41 (5%)
HER2 receptor status				
Negative	[redacted]	100%[a]	100%[a]	NR
Metastatic site, n (%)				
Visceral	261 (53%)	373 (56%)	406 (56%)	471 (64%)
Bone only	109 (22%)	180 (27%)	NR	162 (22%)
Other	123 (25%)	113 (17%)	NR	NR
No. of organ sites, n (%)				
1	143 (29%)	264 (40%)	60%[b]	NR
2	118 (24%)	202 (30%)	36%[b]	NR
≥3	229 (46%)	200 (30%)	2%[b]	NR
Prior (neo)adjuvant chemotherapy, n (%)				
Yes	191 (39%)	401 (60%)	306 (42%)	NR
Prior (neo)adjuvant endocrine therapy, n (%)				
Yes	230 (47%)	NR	NR	475 (65%)
Prior endocrine therapy for advanced breast cancer, n (%)				

Baseline characteristic	First-line ABC treatment trial	Second-line ABC treatment trials		
	MONARCH 3 (n=493)[13] ABE+NSAI vs placebo+NSAI	MONARCH 2 (n=669)[16] ABE+FUL vs placebo+FUL	BOLERO-2 (n=724)[64] EVE+EXE vs placebo+EXE	CONFIRM (N=736)[45, 63] FUL 500 vs FUL 250
Yes	0	256 (38%)	NR	353 (48%)
Prior chemotherapy for advanced breast cancer, n (%)				
Yes	0	0	186 (26%)	NR
Measurable disease, n (%)				
Yes	397 (81%)	482 (72%)	500 (69%)	501 (68%)

NR= Not reported; N/A = Not applicable; AI = Aromatase inhibitor; ABC = advanced breast cancer
NB. Where numbers do not sum to the total number randomised / percentages do not sum to 100 this is due to missing data, or rounding. Some numbers / percentages have been calculated by the ERG (rather than as originally reported in trial publications).
[a] Number not explicitly stated but study publication says the eligible women were HER2;
[b] Defined as number of metastatic sites in the trial publication
[c] includes no previous chemotherapy (n=232) and chemotherapy only in the (neo)adjuvant therapy setting (n=306).

Overall, the MONARCH 2 trial appears to be the most comparable to MONARCH 3 in terms of patient demographic and disease characteristics. However, only 38% of patients in MONARCH 2 are representative of the patients in MONARCH 3 (i.e. patients who had progressed from the advanced breast cancer setting). The baseline characteristics of this sub-group are not presented.