LIVERPOOL REVIEWS AND IMPLEMENTATION GROUP (LRiG)

Lapatinib and trastuzumab in combination with an aromatase inhibitor for the first-line treatment of metastatic hormone receptor positive breast cancer which over-expresses HER2

This report was commissioned by the NIHR HTA Programme as project number 09/101/01

Completed September 15th 2010

CONTAINS IN CONFIDENCE DATA

Lapatinib and trastuzumab in combination with an aromatase inhibitor for the first-line treatment of metastatic hormone receptor positive breast cancer which over-expresses HER2

Produced by: Liverpool Reviews and Implementation Group (LRiG)

Authors: Nigel Fleeman, Research Fellow, LRiG, University of Liverpool

Adrian Bagust, Professor of Modelling in Health, LRiG, University of Liverpool

Angela Boland, Associate Director, LRiG, University of Liverpool

Rumona Dickson, Director, LRiG, University of Liverpool

Yenal Dundar, Research Fellow, LRiG, University of Liverpool

May Moonan, Clinical Lecturer in Public Health Medicine and Specialty Registrar in Public Health, University of Liverpool

James Oyee, Statistician, LRiG, University of Liverpool

Michaela Blundell, Statistician, LRiG, University of Liverpool

Helen Davis, Assistant Director, North West Medicines Information Centre

Anne Armstrong, Consultant Medical Oncologist, The Christie NHS Foundation Trust

Nicky Thorp, Consultant Clinical Oncologist, Clatterbridge Centre for Oncology NHS Foundation Trust

Correspondence to: Ms. Rumona Dickson, Director (LRiG), Liverpool Reviews and Implementation Group, University of Liverpool, Room 2.06, Whelan Building, The Quadrangle, Brownlow Hill, Liverpool L69 3GB

Date completed: 15/09/2010

Source of funding: This report was commissioned by the NIHR HTA Programme as project number 09/101/01

Declared competing interests of the authors:

In the past, Anne Armstrong has received consultancy fees, reimbursement for attending a conference and hospitality at a conference, from Roche

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:

Lapatinib and trastuzumab in combination with an aromatase inhibitor for the first-line treatment of metastatic hormone receptor positive breast cancer which over-expresses HER2. Fleeman N, Bagust A, Boland A, Dickson R, Dundar Y, Moonan M, Oyee J, Blundell M, Davis H, Armstrong A and Thorp N. Liverpool Reviews and Implementation Group, The University of Liverpool, 2010

Contributions of authors:

Nigel Fleeman	Project lead, involved in all aspects of the clinical review and report writing.
Adrian Bagust	Critical appraisal of manufacturers' economic models, development of de novo model and report writing.
Angela Boland	Involved in all aspects of the economics review and report writing.
Rumona Dickson	Support of review process and commented on draft versions of the final report.
Yenal Dundar	Conducted literature searches and contributed to the screening and selection process.
May Moonan	Prepared background section, contributed to analysis and interpretation of data and commented on draft versions of the final report.
James Oyee	Contributed to analysis and interpretation of data, prepared sections of the report relating to statistical issues and commented on draft versions of the final report.
Michaela Blundell	Additional statistical support.
Helen Davis	Contributed to background section and commented on draft versions of the final report.
Anne Armstrong	Contributed to analysis and interpretation of data and commented on draft versions of the final report.
Nicky Thorp	Contributed to background section, analysis and interpretation of data and commented on draft versions of the final report.

Table of contents

DEFINITION OF TERMS AND LIST OF ABBREVIATIONS 5
1 EXECUTIVE SUMMARY 7
1.1 Background 7
1.2 Objectives 8
1.3 Methods 8
1.4 Results 9
1.5 Discussion 14
1.6 Conclusions 15
2 BACKGROUND 16
2.1 Description of health problem 16
2.2 Description of technologies under assessment 19
2.3 Current service provision 23
3 DEFINITION OF THE DECISION PROBLEM 27
3.1 Decision problem 27
3.2 Overall aims and objectives of assessment 28
4 ASSESSMENT OF CLINICAL EFFECTIVENESS 29
4.1 Methods for reviewing effectiveness 29
4.2 Results 31
5 ASSESSMENT OF COST EFFECTIVENESS 57
5.1 Systematic review of existing cost-effectiveness evidence 57
5.2 Overview and critique of GlaxoSmithKline economic evaluation 59
5.3 Overview and critique of Roche economic evaluation 68
5.4 Independent economic assessment 78
5.5 Summary of cost-effectiveness evidence 101
6 DISCUSSION 104
7 CONCLUSIONS 106
7.1 Suggested research priorities 106
8 REFERENCES 107
9 APPENDICES 117

DEFINITION OF TERMS AND LIST OF ABBREVIATIONS

Technical terms and abbreviations are used throughout this report. The meaning is usually clear from the context, but a glossary is provided for the non-specialist reader.

Abbreviations:

AE	Adverse event
AG	Assessment Group
AI	aromatase inhibitor
ANA	anastrozole
BNF	British National Formulary
CBR	Clinical benefit rate
CEAC	Cost Effectiveness Acceptability Curve
CI	confidence interval
CR	complete response
CSR	Clinical Study Report
ECHO	echocardiogram
EGF30008	efficacy and safety of lapatinib combined with letrozole trial
eLEcTRA	efficacy and safety of letrozole combined with trastuzumab trial
EQ-5D	EuroQol five dimensions
ER	oestrogen receptor
ER+	oestrogen receptor-positive
EXE	exemestane
FACT	Functional Assessment of Cancer Therapy
FISH	fluorescent *in situ* hybridization
HER2	human epidermal growth factor receptor 2
HER2+	HER2-positive
HER2-	HER2-negative
HR	hazard ratio
HR+	hormone receptor-positive
HR-	hormone receptor-negative
ICER	incremental cost-effectiveness ratio
IHC	immunohistochemistry
IPCW	Inverse Probability Censoring Weight
ITT	intention to treat
IV	intravenous
LAP	lapatinib

LET	letrozole
LVEF	left ventricular ejection fraction
LYG	life year gain
MBC	metastatic breast cancer
MS	manufacturer's submission / manufacturers submissions
MUGA	multi gated acquisition scan
NICE	National Institute for Health and Clinical Excellence
OR	odds ratio
ORR	overall response rate
OS	overall survival
PFS	progression-free survival
PgR	progesterone receptor
PgR+	progesterone receptor-positive
PP	per protocol
PPS	post-progression survival
PR	partial response
PSA	probabilistic sensitivity analysis
PSS	Personal Social Services
PSSRU	Personal Social Services Research Unit
QALY	quality adjusted life year
QoL	quality of life
RCT	randomised controlled trial
RFS	recurrence-free survival
RPSFT	rank preserving structural failure time
RR	relative risk
SAE	serious adverse event
SD	stable disease
TAM	tamoxifen
TAnDEM	efficacy and safety of trastuzumab combined with anastrozole trial
TRA	trastuzumab
TTP	time to progression

Definitions of terms:

Biological therapy	Treatments that use natural substances from the body, or drugs made from these substances, to fight cancer or to lessen the side-effects that may be caused by some cancer treatments. An example includes trastuzumab
Chemotherapy	Treatment with drugs that kill cancer cells
Endocrine therapy	Treatment that adds, blocks, or removes hormones. Also commonly known as hormonal or anti-oestrogen therapy.
ER (oestrogen receptor)	Proteins that bind oestrogens
ER+ (oestrogen receptor-positive)	Cancer cells which may need estrogen to grow and can thus be treated with endocrine therapy. Cancer cells that are ER- (oestrogen receptor-negative)do not need estrogen to grow
HER2+ (HER2-positive)	over-expression of the HER2 receptor (HER2 receptors present in cancer cells)
HER2- (HER2-negative)	HER2 receptors are not present in cancer cells
Heterogeneity	In statistics this means that there is between-study variation. If heterogeneity exists the pooled effect size in a meta-analysis has no meaning as the presence of heterogeneity indicates that there is more than one true effect size in the studies being combined
Hormone receptor	A receptor that binds to a hormone
HR+ (hormone receptor-positive)	A tumour consisting of cells that express receptors for certain hormones, usually the oestrogen receptor (ER) i.e. oestrogen receptor-positive (ER+) or the progesterone receptor (PgR) i.e. progesterone receptor-positive (PgR+)
Meta-analysis	A quantitative method for combining the results of many studies into one set of conclusions
Oestrogen	A general term for female steroid sex hormones that are secreted by the ovary and responsible for typical female sexual characteristics
Oestrogen receptor-positive (ER+)	A tumour that contain oestrogen receptor-positive cells
Oestrogen receptor-negative (ER-)	Cells that do not have a protein to which oestrogen will bind
Quality adjusted life year(s) (QALYs)	An index of survival that is weighted or adjusted by a patient's quality of life during the survival period. QALYs are calculated by multiplying the number of life years by an appropriate utility or preference score
Receptor	A protein molecule embedded in a membrane, to which a signal molecule (ligand) such as a pharmaceutical drug may attach itself to and which usually initiates a cellular response (although some ligands merely block receptors without inducing any response)

1 EXECUTIVE SUMMARY

1.1 Background

Breast cancer is the uncontrolled, abnormal growth of malignant breast tissue affecting predominantly women. Metastatic breast cancer (MBC) is an advanced stage of the disease when the disease had spread beyond the original organ.

Hormone receptor status and human epidermal growth factor 2 (HER2) status are two predictive factors which are taken into consideration when estimating the prognosis of patients with breast cancer. Tumours which express either ER (ER+) or PgR (PgR+) are commonly referred to as being hormone receptor positive (HR+) and patients with HR+ breast cancer generally have an improved prognosis compared to those who are HR-. More recently it has been discovered that over expression of ErbB2 (i.e. the HER2 protein) and/or amplification of the *HER2* gene results in an abnormally high number of *HER2* genes per cancer cell which results in cancer cells growing and dividing more quickly. Thus HER2+ breast cancer is considered to be an aggressive disease and there is growing evidence that the prognosis of HER2+ patients is generally poor, whether they are HR+ or HR-.

Lapatinib (LAP; brand name: Tyverb®/Tykerb®), an orally active drug given once per day, inhibits the tyrosine kinase components of the epidermal growth factor receptors (ErbB1 and ErbB2), implicated in the growth of various tumours. Currently, LAP is recommended for the first-line treatment of breast cancer in England and Wales in combination with capecitabine in the context of clinical trials for women with advanced or metastatic HER2+ breast cancer. In June 2010, the European Medicines Agency (EMA) granted conditional approval for the use of LAP in combination with an aromatase inhibitor (AI) for the first-line treatment of post-menopausal women with HR+/HER2+ MBC.

Trastuzumab (TRA; brand name: Herceptin®) is a recombinant humanised IgG1 monoclonal antibody directed against HER2 which is administered by intravenous infusion. It has a number of licences including for the treatment of early HER2+ breast cancer. For MBC it is licensed in combination with paclitaxel or docetaxel for patients with HER2+ tumours who have not received chemotherapy for MBC and in whom anthracycline treatment is inappropriate; in combination with an AI for post-menopausal patients with HR+/HER2+ tumours not previously treated with TRA (the use of TRA for patients with early or advanced breast cancer was relatively rare at the time evidence was gathered to support this); and as monotherapy for patients with HER2+ tumours who have received at least two chemotherapy regimens (women with ER+ breast cancer should also have received endocrine therapy).

The aim of current treatments for MBC is to palliate symptoms, prolong survival and maintain a good quality of life (QoL) with minimal adverse events (AE). Choice of treatment depends on previous therapy, hormone receptor status, HER2 status and the extent of the disease. Currently NICE recommends that endocrine therapy (such as tamoxifen [TAM] or an AI) is offered as first-line treatment to the majority of women with ER+ advanced breast cancer. However, providing patients understand and are prepared to accept the toxicity of chemotherapy, this is recommended as first-line treatment when the ER+ MBC is life-threatening or requires early relief of symptoms because of significant visceral organ involvement. Thus in practice, for patients with HR+/HER2+, TRA is commonly given in combination with chemotherapy. A combination of about three chemotherapy drugs are frequently used together, but the choice and number of chemotherapeutic agents used is specific to the patient and decided by the lead clinician. Examples of chemotherapeutic agents commonly used include fluorouracil (5FU), methotrexate, cyclophosphamide and epirubicin.

While it is known there were 45,695 new cases of breast cancer in women in the UK in 2007, data on the number of women with HR+/HER2+ breast cancer and/or MBC are not routinely collected. However, the number of patients with HR+/HER2+MBC estimated to be suitable for treatment with either LAP or TRA in combination with an AI is around 50 patients a year.

1.2 Objectives

The remit of this appraisal is to review the clinical and cost-effectiveness evidence base for LAP+AI and TRA+AI within their licensed indications for the first-line treatment of patients who have HR+/HER2+ MBC.

1.3 Methods

Evidence for clinical effectiveness of LAP in combination with an AI (LAP+AI) and TRA in combination with an AI (TRA+AI) for the first-line treatment of HR+/HER2+ MBC was assessed by conducting a systematic review of published research evidence. The review was undertaken following the general principles published by the Centre for Reviews and Dissemination's guidance for undertaking reviews in healthcare.

Randomised controlled trials (RCTs) were identified by searching major electronic medical databases including MEDLINE, EMBASE and the Cochrane Library. The search strategy was broad and not limited to RCTs. Information on studies in progress, unpublished research or research reported in the grey literature were sought by searching a range of relevant databases including National Research Register and Controlled Clinical Trials. In addition, bibliographies of previous reviews and retrieved articles were searched for further studies. The same search strategies were used to identify economic evaluations and are presented in Appendix 1.

Further attempts to identify studies were made by contacting clinical experts and examining the reference lists of all retrieved articles. The manufacturers' submissions (MS) were assessed for unpublished data.

Two reviewers independently screened all titles and abstracts. Full paper manuscripts of any titles/abstracts that were considered relevant by either reviewer were obtained where possible. Data were extracted by one reviewer using a standardised data extraction form in Microsoft Word 2007 and checked independently by a second reviewer. Disagreements were resolved by discussion. The quality of the individual clinical-effectiveness studies was assessed according to criteria based on the CRD's guidance for undertaking reviews in healthcare. The assessment of risk of bias was conducted independently by both reviewers (MM, NF). Disagreements were resolved through discussion.

The results of the data extraction and quality assessment for each study are presented in structured tables and as a narrative summary. The possible effects of study quality on the effectiveness data and review findings are discussed. It was intended by the Assessment Group (AG) that meta-analyses would be conducted in which direct evidence would be pooled using a standard meta-analysis and where a direct comparison between LAP+AI and TRA+AI was not possible, by indirect comparisons. However the AG considered it inappropriate to conduct either of the analyses, as discussed further below.

1.4 Results

1.4.1 Assessment of clinical effectiveness

Quantity and quality of research available

Once duplicates were removed, a total of 2069 references were identified. Two trials, EGF30008 and TAnDEM met the inclusion criteria and a further trial (eLEcTRA), reported only as a conference abstract, was also included following information passed on to the AG by Roche at the NICE consultation meeting in February, 2010. Thus three trials were included in the systematic review.

All three trials were multi-centre and multi-national trials (between 7 and 29 countries) enrolling post-menopausal patients receiving first-line treatment for MBC; all three trials included patients who had HR+/HER2+ MBC although EGF30008 and eLEcTRA also included patients who were HR+/HER2-.

Clinical endpoints including overall survival (OS), Progression-free survival (PFS) and time to treatment progression (TTP) that are commonly used in trials of breast cancer were utilised in at least one of the trials included in this appraisal. However, the only efficacy endpoints common to all three, and reported on by all three, were the secondary endpoints, clinical benefit rate (CBR) and overall response rate (ORR). All three trials also reported on AEs.

As patients in TAnDEM and eLEcTRA received second-line treatment once their disease had progressed, data on OS should be treated with caution as clearly this extra treatment could potentially impact on OS. It is not stated that patients received second-line treatment in EGF30008. Data on PFS, TTP, CBR and ORR in all trials should be treated with caution due to the way they were measured.

Overall, the risk of bias assessment conducted by the AG found EGF30008 and TAnDEM to be of a good standard. The eLEcTRA trial was deemed to be of poorer quality which may be a reflection of poor quality reporting rather than trial design as this trial was published as an abstract.

Arguably, the most significant difference between EGF30008 and the other two trials is choice of exclusion criteria. The EGF30008 trial excluded patients in which "the disease was considered by the investigator to be rapidly progressing or life threatening." The potential importance of this criterion became apparent when analysing median OS which was no greater than 23.9 months (unadjusted ITT population) in the TAnDEM trial compared to 33.3 months in the EGF30008 trial. If it is assumed, as NICE guidance on early breast cancer suggests, that there is a 'class-effect' and LET or ANA are equally effective then if the populations were similar, a similar median OS would be expected in the LET and ANA arms. Thus it was felt any comparisons made across trials would not be valid and the AG decided to focus on discussing the three trials individually.

Assessment of effectiveness

The findings from the three main trials examining the efficacy of LAP+LET (EGF30008), TRA+ANA (TAnDEM) and TRA+LET (eLEcTRA) all suggest that LAP+AI or TRA+AI result in improved outcomes when compared to AIs (LET or ANA). In the EGF30008 and TAnDEM trials, while these differences were not significant for OS, significantly different outcomes were reported for PFS and TTP. Large differences were reported between TRA+LET and LET patients in eLEcTRA; this trial lacked statistical power to adequately test for significant differences. In addition, both ORR and CBR appeared to be improved for patients taking LAP+AI or TRA+AI although the only statistically significant differences were found for TRA+ANA compared to ANA in TAnDEM. No new safety concerns were identified from the trials although both AEs and SAEs were more common in the LAP+LET and TRA+AI groups than in AIs alone. For LAP+LET, the most significant AE was diarrhoea experienced by around a third of all patients. The impact this may have on patient QoL is difficult to estimate as only the EGF30008 trial attempted to measure QoL and to date the findings have only been presented as a conference abstract. However, it would appear there are no statistically significant differences between patients in either treatment group. Indeed, the majority of cases of AEs (including diarrhoea) were of grade 1 or 2 severity. Nevertheless, diarrhoea did result in around 1% of all patients who received LAP+LET discontinuing their treatment as a result; all other patients were managed by dose reduction, dose interruption or supportive intervention without treatment dose

adjustments. For TRA+ANA patients, the most frequently reported AEs were fatigue, diarrhoea and vomiting experienced by around a fifth of all patients, of which the majority were grade 1 or 2 severity. Fatigue was also a problem for around a quarter of patients who received TRA+LET but infections, gastrointestinal disorders and musculoskeletal and connective tissue disorders were even more common; over half of TRA+LET patients experienced these latter two AEs. Around a third of LET patients also reported gastrointestinal disorders and musculoskeletal and connective tissue disorders.

The fact that patients were able to cross-over in the TAnDEM trial has added an extra problem in interpreting the OS findings, from this trial, namely, how much of the benefit in OS is attributable to the first-line treatment and how much of the benefit is attributable to subsequent treatment following disease progression? Post-hoc attempts were made, firstly comparing both the median OS between those receiving TRA+ANA with those who initially received ANA but did not cross-over to receive TRA, and those in the AI group who crossed over to receive TRA+ANA with those who did not. Secondly, a a statistical modelling approach known as rank preserving structural failure time (RPSFT) approach was also employed. Both approaches have their weaknesses and the AG believes that other different randomisation-based methods should ideally be used to compute and compare a range of OS estimates to assess sensitivity of treatment effects. Therefore, the AG believes the findings from the approaches attempting to adjust for cross-over should be treated with extreme caution.

Because the manufacturers also believed that direct comparison across trials would be too crude and simplistic, both manufacturers conducted adjusted indirect comparisons. However, the AG believes these indirect comparisons were not appropriate and must also be treated with caution for a number of important reasons. First and foremost, as the AG does not believe the patient populations to be sufficiently similar in the EGF30008 and TAnDEM trials, then these studies should not be compared with each other at all. In addition, trials which included patients of unknown HR+/HER2+ status were included in the network analyses submitted by the manufacturers. An important finding from the EGF30008 trial was that PFS in patients with HR+/HER2+ differed to that in patients of any HER2 status, particularly in the LET group.

Thus, overall, the AG believes comparisons across trials cannot be made and that only the individual findings from each trial should be considered.

1.4.2 Assessment of cost effectiveness

Cost-effectiveness review

The AG did not identify any relevant papers for inclusion in the cost effectiveness review of LAP+AI or TRA+AI in patients who are HR+/HER2+ with MBC. The manufacturer of TRA identified a poster which was presented at the ASCO 2010 conference; the study described compared LAP+LET vs TRA+ANA using an indirect comparisons analysis. The AG is of the opinion that the results of the indirect analysis performed by Hastings et al are unreliable as the studies which make up the evidence network are inappropriate. In addition, the AG notes that without access to more detailed information on costs, it is difficult to comment on the reliability of the cost-effectiveness results in this study.

Submitted economic evaluations by manufacturers

The two economic evaluations submitted by the manufacturers appear to meet the NICE reference case criteria. However, the AG is critical of the approaches used by the manufacturers to estimate OS in each of their models; the AG is of the opinion that projective modelling in this group of patients can lead to substantial bias in OS estimates. In addition, the AG also identified several costing inaccuracies and inconsistencies in both of the economic evaluations submitted.

For the direct comparisons, GlaxoSmithKline demonstrated that LAP+LET is not cost effective compared with LET and Roche demonstrated that TRA+ANA is not cost effective compared with ANA.

Both of the manufacturers undertook indirect comparisons analyses in order to be able to compare LAP+LET vs TRA+ANA. GlaxoSmithKline demonstrated that LAP+LET is cost effective compared with TRA+ANA. Roche demonstrated that TRA+ANA is cost effective compared with LAP+LET. The AG concludes that the indirect comparisons analyses conducted by the manufacturers are unreliable and that only the ICERs estimated from the direct comparisons are valid.

Roche makes the case for TRA+ANA to be considered as an end of life treatment for women with HR+/HER2+ MBC. The AG does not have sufficient information to verify whether all three NICE criteria for consideration of end of life treatments are met.

AG's cost-effectiveness results and sensitivity analysis

The AG reports the results of two separate *de novo* cost-effectiveness analyses using a common framework and common parameter values but employing effectiveness data drawn only from a single RCT (either EGF30008 or TAnDEM). The AG model employs outcome data derived from the relevant clinical trial in the form of Kaplan-Meier estimated survival values augmented by projected survival estimates calibrated against the observed data. The AG uses PFS and post-progression survival (PPS) estimates directly as the basis for calculating expected OS in each group of the RCT.

As the AG is of the opinion that the evidence base is too unstable to allow meaningful comparison of LAP+LET vs TRA+ANA, the only questions that may be addressed legitimately are:

- Can LAP+LET be considered a cost-effective treatment compared with LET alone?
- Can TRA+ANA be considered a cost-effective treatment compared with ANA alone?

Base case result: LAP+LET vs LET

The AG concludes that in HR+/HER2+ women with MBC, LAP+LET compared with LET is not cost effective. Using a time horizon of 20 years, the AG estimates an ICER which exceeds £220,000 per QALY gained for the comparison of LAP+LET vs LET; the incremental total costs and QALYs per patient treated are estimated as £25,209 and 0.114 respectively.

Base case result: TRA+ANA vs ANA

The AG concludes that in HR+/HER2+ women with MBC, TRA+ANA compared with ANA is not cost effective. Using a time horizon of 20 years, the AG estimates an ICER which exceeds £80,000 per QALY gained for the comparison of TRA+ANA vs ANA; the incremental total costs and QALYs per patient treated are estimated as £36,687 and 0.448 respectively.

LAP+AI vs TRA+AI

The AG emphasises again that the currently available clinical evidence base is too unstable to allow meaningful comparison of LAP+AI vs TRA+AI.

Sensitivity analyses undertaken by the AG

For the comparison of LAP+LET vs LET the univariate sensitivity analysis shows that the ICER is most sensitive to the choice of health state utility parameter values, the cost of LAP and is insensitive to most of the other variables. In all cases, the ICER remains above £137,000 per QALY gained. The PSA shows that there is no measureable probability of LAP+LET being cost effective at a willingness-to-pay threshold of £40,000 per QALY gained; to achieve a 50% probability of LAP+LET being cost effective, the willingness-to-pay threshold needs to increase to around £3,000,000 per QALY gained.

For the comparison of TRA+ANA vs ANA, the univariate sensitivity analysis shows that the ICER is most sensitive to the choice of health state utility parameter values, the cost of TRA and discounting rates only. In all cases, the ICER exceeds £65,000 per QALY gained. The PSA shows that there is no measureable probability of TRA+ANA being cost effective compared to ANA at a willingness-to-pay threshold of £50,000.

1.5 Discussion

Strengths, limitations of the analyses and uncertainties

Only three RCTs have been identified which present head-to-head comparisons of the interventions of interest to this appraisal. It was not possible to compare the data across the trials because of differences in the patient populations. From a health economics perspective, the AG agrees with both manufacturers that LAP+LET and TRA+ANA are not cost effective compared with AIs alone for women with HR+/HER2+MBC. The ICERs estimated by the AG for LAP+LET vs LET and TRA+ANA vs ANA are higher than those estimated by the manufacturers.

The AG believes that the results of any indirect comparisons analyses of LAP+LET vs TRA+ANA are unreliable due to heterogeneous patient populations. In addition, to complete the evidence network in the indirect comparisons analyses presented in the submitted MS, the manufacturers had to use trials with mixed HER2- and HER2+ populations. The AG is of the opinion that use of clinical effectiveness evidence from a mixed population adds to the uncertainty regarding the results of the indirect analyses conducted by the manufacturers. Consequently, the AG did not address the cost effectiveness of LAP+LET vs TRA+ANA as there were insufficient comparative clinical data available to allow estimation of meaningful ICERs.

Generalisability of the findings

None of the patients in EGF30008or TAnDEM have received prior treatment with TRA; this is not surprising as, at the time the trials were recruiting, the use of TRA for patients with early or advanced breast cancer was relatively rare. This contrasts very much with what happens in clinical practice in the NHS today. Now, when a patient is diagnosed with early HER2+ breast cancer, TRA is the standard treatment of choice and in reality it is likely that only *de novo* patients with HR+/HER2+ MBC will be eligible for TRA+AI as per the wording of the recently awarded EMA licence. Patients who have been treated with TRA previously are eligible for treatment with LAP+AI; however, it is uncertain whether the clinical effectiveness of LAP+AI is the same for patients who are and who are not TRA-naive.

1.6 Conclusions

Clinical effectiveness evidence from two RCTs demonstrates that LAP+LET or TRA+ANA improves median PFS and/or TTP compared with AI monotherapy in patients who are HR+/HER2+ MBC. To date, the trials do not show a statistically significant benefit in terms of OS for patients taking LAP+LET vs AI monotherapy or TRA+ANA vs AI monotherapy. The results of the economic evaluations conducted by the manufacturers, and confirmed by the AG, demonstrate that LAP+LET is not cost effective compared with AI monotherapy, nor is TRA+ANA cost effective compared with AI monotherapy.

Due to differences in the patient populations of EGF30008 and TAnDEM, the AG believes the indirect comparisons analyses conducted by the manufacturers are inappropriate and for the same reason chooses not to compare LAP+LET with TRA+ANA in an economic evaluation.

As the results of the EGF30008 trial appears to demonstrate that there are large differences in PFS for HR/HER2+ and HR+/HER2-patients receiving both LAP+LET and, in particular, LET, further research may be warranted to compare the clinical effectiveness of AIs alone in patients with HER2+ and HER2- breast cancer.

Most patients who present for HR+/HER2+ MBC are likely to have been previously treated for early breast cancer and very probably with regimens including TRA (unlike at the time the pivotal trials in this appraisal were conducted). Further research may be required into treating MBC in the HR+/HER2+ population who are not TRA (or LAP) naive. In addition, future research should consider adjusting for cross-over *a priori.*

2 BACKGROUND

2.1 Description of health problem

Breast cancer is the uncontrolled, abnormal growth of malignant breast tissue affecting predominantly women. Though frequently referred to as a homogenous disease, breast cancer has been recognised as a biologically heterogeneous disease[1] with several sub-groups including those with different stages and types of the disease. Metastatic breast cancer (MBC) is an advanced stage of the disease when the disease has spread beyond the original organ. Common sites of metastasis include bone, liver, lung and brain.

2.1.1 Aetiology

After gender, the strongest risk factor for breast cancer is age. The incidence of breast cancer increases with age, doubling every 10 years until menopause, after which the rate of increase slows. Breast cancer is rare under the age of 20.

Genetic and hormonal risk factors have also been identified in the aetiology of breast cancer,[2, 3] and women with a family history of breast cancer have an increased risk of developing the disease.[4] Mutations in some genes can increase the risk of developing breast cancer. *BRCA1* gene mutations account for 2% of breast cancers, where the risk is as high as 85% by the age of 35 years.[5] *BRCA2* mutations account for up to 1% of breast cancers, with a 60% chance of breast cancer. Many breast cancer tumours are stimulated to grow and change by female sex hormones, oestrogen receptors (ER) and progesterone receptors (PgR). *BRCA2* tumours characteristically express ER and PgR. Other gene mutations contributing less frequently to familial breast cancer include mutations in the *PTEN*, *MSH1*, *MSH2* and *p53* genes.

Higher concentrations of some endogenous hormones appear to increase breast cancer risk.[6] Early age at menarche, late natural menopause, later age at first full-term pregnancy and never breastfeeding are all associated with an increased risk of breast cancer[6] whilst childbearing and higher numbers of full-term pregnancies increase protection.[6] Use of exogenous hormones such as oral contraception, oestrogen replacement therapy and combined endocrine therapy increase the risk of breast cancer as do other factors such as breast density (a risk factor independent of endogenous hormones), a body mass index (BMI) of 25+ in post-menopausal women, moderate to heavy alcohol intake and a sedentary lifestyle.[6]

2.1.2 Pathology and prognosis

There are several prognostic factors which are taken into account by clinicians when deciding on treatment options and making a clinical prognosis.[7] These include age, tumour size, histological type, nuclear grade, histological grade, number of metastatic axillary lymph nodes, and clinical stage. Patients with stage IV disease are classified as having MBC according to the tumour/nodes/metastasis (TNM) staging system developed and maintained by the American Joint Committee on Cancer[8] and the Union International Contre le Cancer.[9]

Hormone receptor status and human epidermal growth factor 2 (HER2) status are two other predictive factors which are taken into consideration in estimating the prognosis of patients with breast cancer. As noted above, many breast cancer tumours are stimulated to grow and change by ER and PgR. Tumours which express either ER (ER+) or PgR (PgR+) are commonly referred to as being hormone receptor positive (HR+) and patients with HR+ breast cancer generally have an improved prognosis compared to those who are HR-. More recently it has been discovered that over expression of ErbB2 (i.e. the HER2 protein) and/or amplification of the *HER2* gene results in an abnormally high number of *HER2* genes per cancer cell which results in cancer cells growing and dividing more quickly. Thus HER2+ breast cancer is considered to be an aggressive disease and there is growing evidence that the prognosis of HER2+ patients is generally poor, whether they are HR- or HR+. It should be emphasised that prior to this understanding of the role of HER2, trials did not routinely present data on this sub-group of patients.

Both HR+ tumours and HER2+ tumours are determined by immunohistochemistry (IHC).[5] Fluorescent *in situ* hybridisation (FISH) can also be used to measure HER2 expression by measuring the number of gene copies present. An IHC score of 3+ or a FISH amplification of 2.1 or greater confirms a HER2+ status. An IHC of 2+ is usually confirmed by FISH.[5] Biological markers such as HER2 are also used as a predictor of prognosis and as a guide to therapy.

In England and Wales, 80%, 72% and 64% of people diagnosed with breast cancer live for at least 5, 10 and 20 years after diagnosis, respectively.[10] Although therapeutic innovations have provided modest improvements in survival rates over the past two decades, MBC remains an incurable disease and the aim of treatment is to prolong Progression-free survival (PFS) and palliation.[11] Following a diagnosis of MBC, the average length of survival has been reported to be 12 months for those receiving no treatment,[12] compared to 18-24 months for those receiving chemotherapy, a figure reduced by up to 50% for patients who are HER2+.[13]

2.1.3 Epidemiology

Breast cancer is the most common cancer in the UK with 45,972 new cases diagnosed in 2007, 99% (45,695) being in women.[14] Accounting for almost a third (31%) of all new cases of cancer in women in the UK, the lifetime risk of breast cancer for a woman is 1 in 9.[14]

There is little regional variation in breast cancer rates in the UK,[15] although there appears to be geographical variation within Europe. Breast cancer is one of the few cancers to show a clear trend of increasing rates from most to least deprived groups [14] with rates in the most deprived groups around 20% lower than in the most affluent.[14] The European age-standardised incidence rate (EASR) for women has increased by 5% from 114 per 100,000 in 1998 to 120 per 100,000 in 2007, with the number of cases rising from 40,377 to 45,695, an increase of 13%. The EASR has been projected to increase from 119 per 100,000 in 2000-04 to 124 per 100,000 in 2020-24, with the average number of new cases per year rising from 41,900 to 55,700 over the same time period. Analysis of breast cancer survival by level of deprivation has however consistently shown higher survival for more affluent women.[16]

UK data on breast cancer by stage of disease is not routinely collected and so neither the incidence nor the prevalence of MBC in the UK is known. However, a study of five cancer registries in 2004 estimated that the proportion of all new breast cancer patients with MBC was around 5%[17] and prevalence is thought to be relatively higher because some women live with the disease for many years[18] (although as noted above, the average life-expectancy in the UK is thought to be around 18-24 months).

Approximately 60% and 80% of all breast cancers have been estimated to be HR+ in pre-menopausal and post-menopausal women respectively. Since the introduction of HER2 testing in the UK in October 2006, up to 25% of women diagnosed with breast cancer have been reported to be HER2+.[19]

In women with MBC, NICE[20] states that up to 30% have HER2+ tumours, of which approximately 50% are also HR+. As mentioned previously, there are limited data on how many women are diagnosed with *de novo* MBC, but it has been estimated that around 30% of women with earlier stages of breast cancer will eventually be diagnosed with MBC.[21]

Thus, assuming that 5% of all women diagnosed with breast cancer have MBC, of which 30% are HER2+ and 50% of these are HR+, in the UK approximately 350 patients each year are diagnosed with HR+/HER2+ MBC. An independent estimate from 2008 data derived from the IMS Oncology Analyzer obtained by GlaxoSmithKline[22] has estimated there may be around 500 new cases of HR+/HER2+ MBC each year in the UK.

2.1.4 Impact of health problem

The impact of a diagnosis of MBC breast cancer on a patient is both physiological and psychological [23], affecting not only the patients but also their families and wider social network. Physical ill-health can stem from both the disease and disease treatment. NICE CG81[24] gives guidance for the management of complications such as lymphoedema, fatigue and metastases. Adequate rehabilitation is vital as women may be less productive after treatment for the disease.[25] The psychological impact on the patient can be debilitating, including depression and fear of loss of autonomy,[26] sexuality and body image.[27]

2.2 Description of technologies under assessment

Lapatinib

Lapatinib (LAP; brand name: Tyverb®/Tykerb®) inhibits the tyrosine kinase components of the epidermal growth factor receptors (ErbB1 and ErbB2), implicated in the growth of various tumours.[28] LAP belongs to a group of medicines called protein kinase inhibitors which work by blocking enzymes known as protein kinases. Protein kinases can be found in some receptors on the surface of cancer cells including HER2. HER2, a receptor for epidermal growth factor, is involved in stimulating the cells to divide uncontrollably. By blocking these receptors, LAP helps to control cell division.

The most common side effects of LAP are loss of appetite, diarrhoea, nausea, vomiting, rash and fatigue. Monitoring of left ventricular function and for pulmonary toxicity should be carried out regularly. Monitoring of liver function should be performed before treatment and at monthly intervals.[29] The manufacturer has advised caution in the use of LAP in patients with moderate to severe hepatic impairment and severe renal impairment. Pregnancy should be avoided and breastfeeding discontinued during treatment with LAP.

Lapatinib is an orally active drug given once per day and is available as 250mg tablets.

Currently, LAP is recommended for the first-line treatment of breast cancer in England and Wales in combination with capecitabine in the context of clinical trials for women with advanced or metastatic HER2+ breast cancer.[30]

In June 2010, the European Medicines Agency (EMA) granted conditional approval for the use of LAP in combination with an aromatase inhibitor (AI) for the first-line treatment of post-menopausal women with HR+/HER2+ MBC.[31 32]

Trastuzumab

Trastuzumab (TRA; brand name: Herceptin®) is a recombinant humanised IgG1 monoclonal antibody directed against HER2. It is administered by intravenous infusion (IV), the regimen and dose dependent on several clinical factors including the patient's weight and other medications and stage of disease. Common infusion regimens include once a week for advanced breast cancer and every three weeks for early breast cancer, with the infusion taking approximately 30 to 90 minutes each time.

The most common side effects of TRA are fatigue and diarrhoea. Recent clinical trial data suggest that patients require a left ventricular ejection fraction (LVEF) of >55% for treatment with TRA and the Summary of Product Characteristics for TRA states that cardiac monitoring is required every 12 weeks during treatment. However, the optimal frequency of cardiac monitoring in the clinical practice setting is not universally agreed.[33]

TRA should be used with caution in patients with symptomatic heart failure, history of hypertension, coronary artery disease and uncontrolled arrhythmias. Pregnancy should be avoided during treatment and breast-feeding should be avoided during treatment and for six months after.[29]

TRA is currently licensed in the UK for:

1. The treatment of early breast cancer which over-expresses HER2.

2. In combination with paclitaxel or docetaxel, for MBC in patients with HER2+ tumours who have not received chemotherapy for MBC and in whom anthracycline treatment is inappropriate.

3. In combination with an AI, for MBC in post-menopausal patients with HR+/HER2+ tumours not previously treated with TRA. (Fig. 1)

4. As monotherapy for MBC in patients with HER2+ tumours who have received at least two chemotherapy regimens including, where appropriate, an anthracycline and a taxane; women with ER+ breast cancer should also have received endocrine therapy.[29]

The AG contacted the European Medicines Agency for clarification about point 3 above as the interpretation of this licence varied amongst NHS clinicians. It was not clear whether TRA was indicated for a woman who had been given TRA during the treatment of early breast cancer, who subsequently progressed to MBC. The EMA responded by stating that TRA was licensed for use in MBC in TRA-naive patients. Of note, at the time the trials were recruiting, the use of TRA for patients with early or advanced breast cancer was relatively rare.

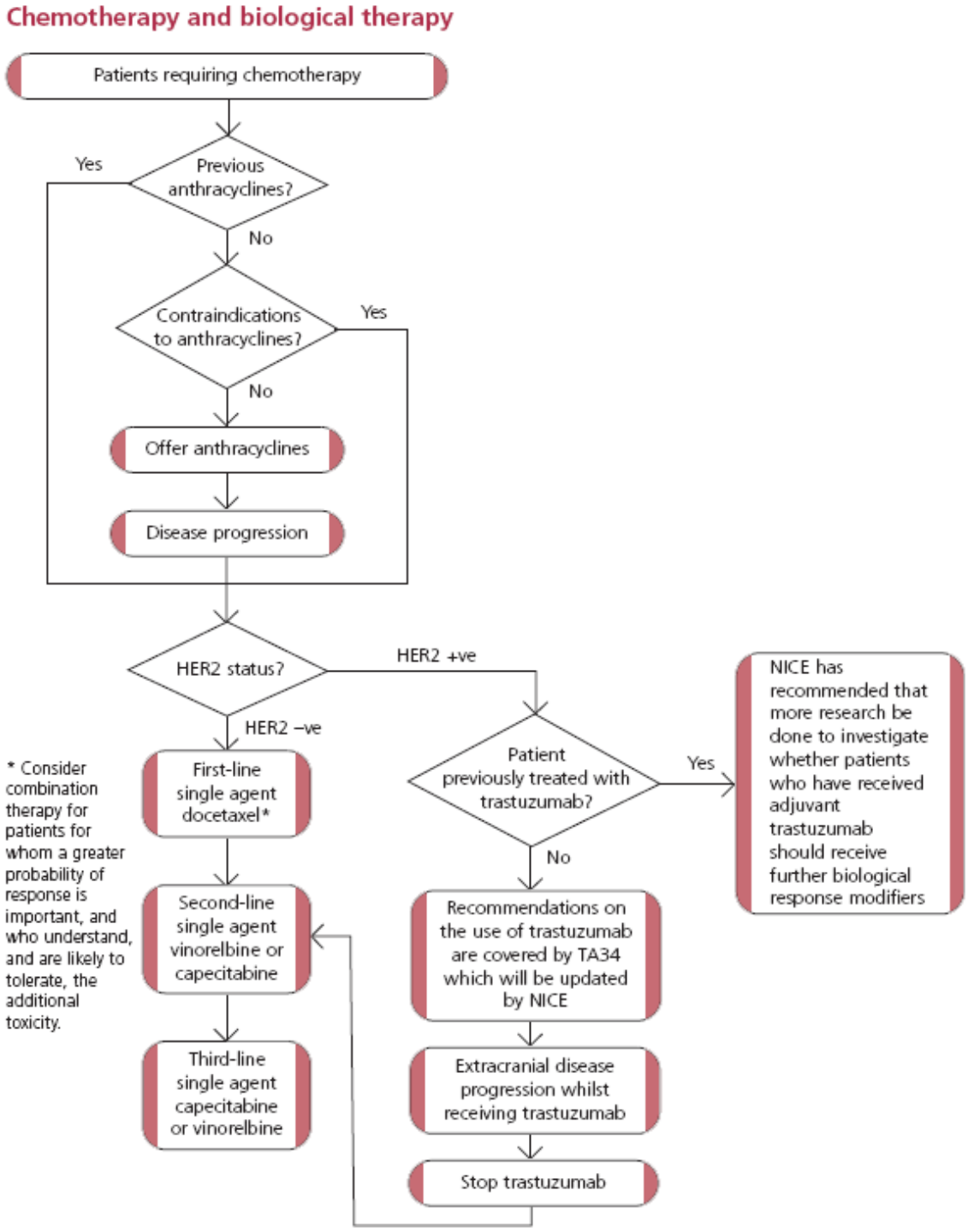

Figure 1 Pathway recommended by NICE for offering trastuzumab

Aromatase inhibitors

Aromatase inhibitors are not *per se* one of the technologies under assessment in this appraisal. However, they are being assessed in combination with LAP and TRA and are one of the comparators. NICE issued guidance regarding the use of AIs in 2006.[34] During this appraisal, the Appraisal Committee "agreed that there is insufficient evidence to conclude that any one aromatase inhibitor (used within the licensed indications) or treatment strategy is more clinically effective than another." As such, the AIs considered in this technology appraisal are assumed to be equally clinically effective. However, in practice only LET and ANA are commonly used as first-line treatments for women with MBC (and they may also be offered as a second-line treatment), with exemestane (EXE) being mostly used as second-line.

Aromotase inhibitors are a form of endocrine therapy and act predominantly by blocking the conversion of androgens to oestrogens in the peripheral tissues. As such, they are classified as antiestrogen therapies. Aromatase inhibitors are classified into irreversible steroidal inhibitors (e.g. EXE) and non-steroidal inhibitors (e.g. ANA and LET), the latter inhibiting the enzyme by reversible competition.

Leterozole (Femara®) is indicated as adjuvant treatment of HR+ early breast cancer in post-menopausal women, advanced breast cancer in post-menopausal women (including those in whom other endocrine therapy has failed), early invasive breast cancer in post-menopausal women after standard adjuvant TAM therapy and pre-operative treatment in post-menopausal women with localised hormone-receptor-positive breast cancer to allow subsequent breast conserving surgery.

Cautions and contraindication include the avoidance of use during pregnancy and breast feeding. Avoidance has also been advised in severe hepatic impairment whilst caution has been advised if creatinine clearance is less than 10 mL/minute.

Anastrozole (Arimidex®) is indicated as adjuvant treatment of ER+ early invasive breast cancer in post-menopausal women, adjuvant treatment of ER+ early breast cancer in post-menopausal women following 2–3 years of TAM therapy and in advanced breast cancer in post-menopausal women which is ER+ or responsive to TAM.[29]

Caution has been advised for the use of ANA in patients susceptible to osteoporosis; bone mineral density should be measured before treatment and at regular intervals during treatment. Anastrazole is contraindicated in pre-menopausal women. Its use should also be avoided in patients with moderate to severe hepatic impairment and renal impairment where creatinine clearance is less than 20 mL/minute. As with LET, it should also be avoided in pregnancy and breast-feeding.

Common side effects of ANA include hot flushes, vaginal dryness, vaginal bleeding, hair thinning, anorexia, nausea, vomiting, diarrhoea, headache, arthralgia, bone fractures and rash (including Stevens-Johnson syndrome).[29]

Exemestane (Aromasin®) is indicated as adjuvant treatment of ER+ early breast cancer in post-menopausal women following two to three years of TAM therapy and in advanced breast cancer in post-menopausal women where endocrine therapy has failed.

As with other AIs, EXE is contraindicated in pre-menopausal women and should be avoided in pregnant and breast feeding women. Caution in its use is advised in patients with renal and hepatic impairment. Common side effects include nausea, vomiting, abdominal pain, dyspepsia, constipation, anorexia, dizziness, fatigue, headache, depression, insomnia, hot flushes, sweating, alopecia and rash.

2.3 Current service provision

The aim of current treatments for MBC is to palliate symptoms, prolong survival and maintain a good quality of life (QoL) with minimal adverse events (AE). Choice of treatment depends on previous therapy, hormone receptor status, HER2 status and the extent of the disease (Figure 2).

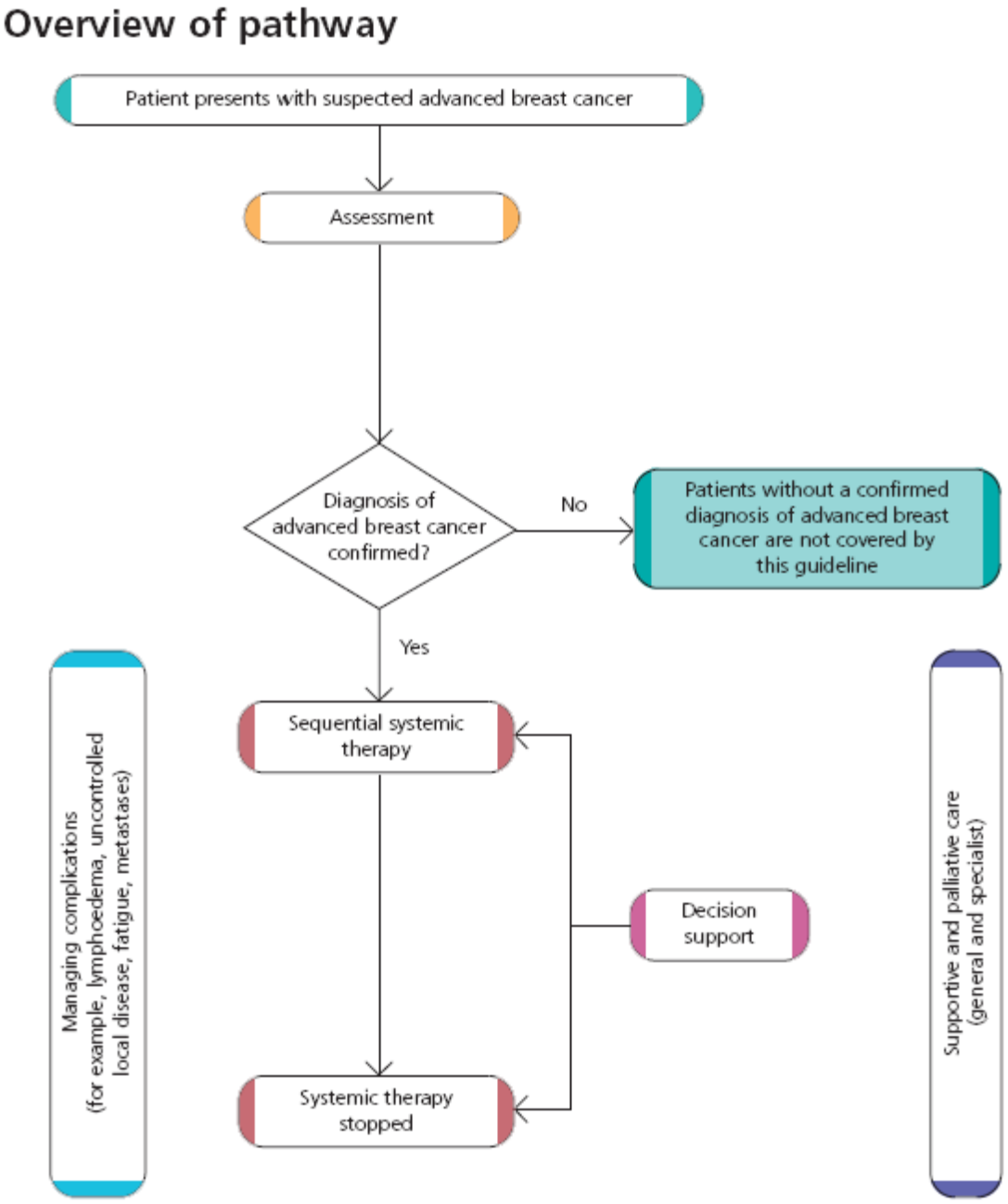

Figure 2 Pathway recommended by NICE for treatment for patients with advanced breast cancer

NICE[24] recommends that endocrine therapy (such as TAM or an AI) is offered as a first-line treatment to the majority of women with ER+ advanced breast cancer. However, providing patients understand and are prepared to accept the toxicity of chemotherapy, this is also recommended as first-line treatment when the ER+ MBC is life-threatening or requires early relief of symptoms because of significant visceral organ involvement (Figure 3).

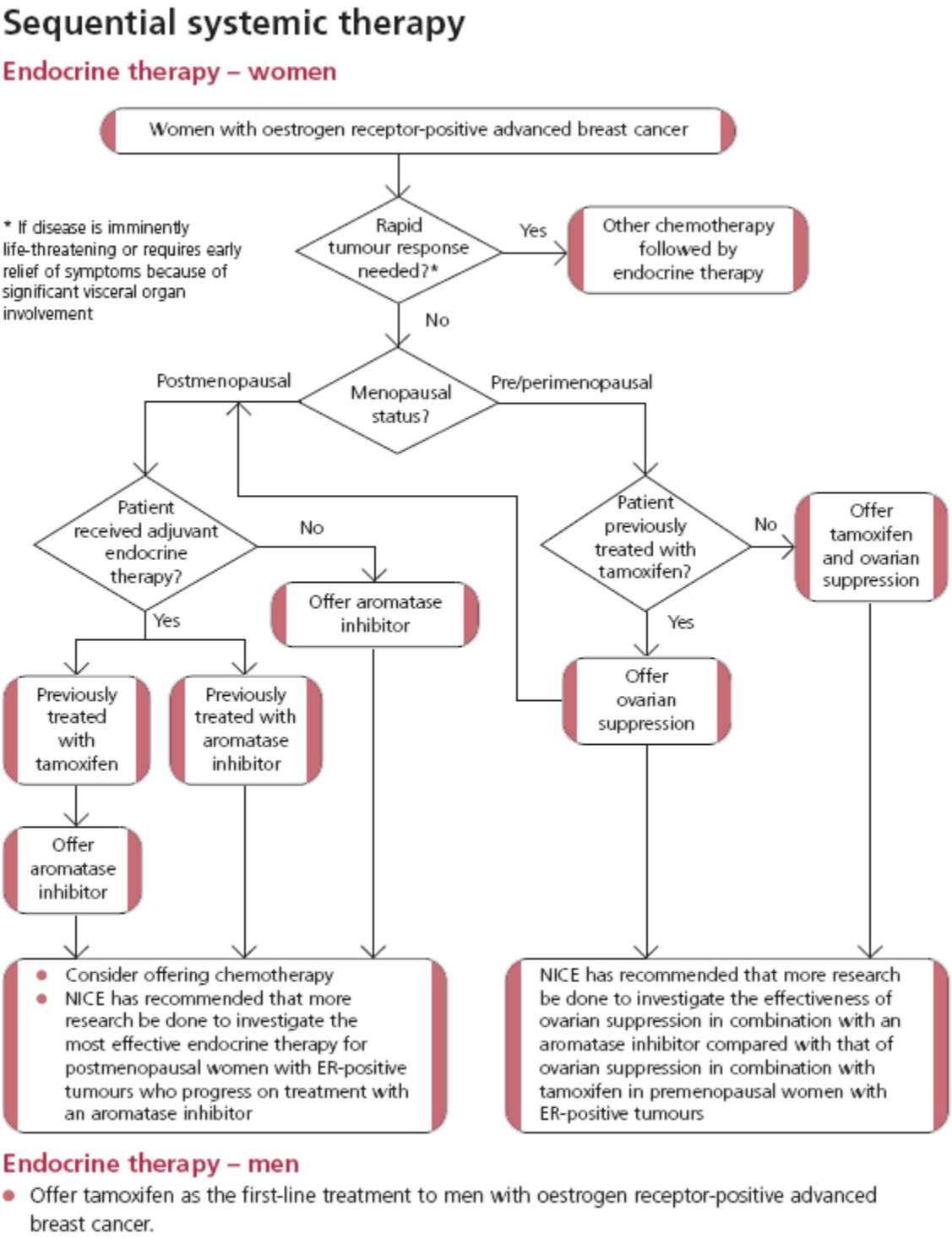

Figure 3 Treatment pathway recommended by NICE for endocrine therapy and chemotherapy in HR+ breast cancer

For patients who are receiving treatment with TRA for advanced breast cancer, NICE recommends[13] that treatment with TRA is discontinued at the time of disease progression outside the central nervous system but that TRA is continued if disease progression is within the central nervous system alone.

In practice, for patients with HR+/HER2+, TRA is commonly given in combination with chemotherapy. A combination of about three chemotherapy drugs are frequently used together, but the choice and number of chemotherapeutic agents used is specific to the patient and decided by the lead clinician. Examples of chemotherapeutic agents commonly used include fluorouracil (5FU), methotrexate, cyclophosphamide and epirubicin.

However, variation in management of patients by age has also been reported.[35, 36] Variation in practice regarding continued use of TRA at the time of disease progression also exists,[37] partly due to uncertainty about mechanisms of resistance and whether this is partial or absolute.

As already noted, TRA, in combination with an AI, has been licensed for the treatment of post-menopausal patients with HR+/HER2+ MBC, not previously treated with TRA.[38] Given the growing number of patients who are treated with TRA in the early breast cancer setting, the number of patients estimated to be suitable for treatment with either LAP or TRA in combination with an AI is around 50 patients a year according to estimates from the manufacturers of LAP[22] and TRA.[39]

3 DEFINITION OF THE DECISION PROBLEM

3.1 Decision problem

3.1.1 Interventions

The following two interventions are being considered:

- Lapatinib (LAP) + aromatase inhibitor (AI)
- Trastuzumab (TRA) + AI

3.1.2 Population including sub-groups

The population of interest is patients with MBC receiving first -line treatment who must:

- have hormone receptor-positive (HR+) tumours and
- have tumours over-expressing ErbB2 receptor, i.e. be HER2+

3.1.3 Relevant comparators

For LAP+AI, the relevant comparators are:

- AIs alone
- TRA+AI

For TRA+AI, the relevant comparators are:

- AIs alone
- LAP+AI

3.1.4 Outcomes

The NICE scope identified the following relevant outcomes:

- overall survival (OS)
- progression-free survival (PFS)
- time to progression (TTP)
- response rate, which (although not specified in the scope) may further be broken down to:
    - overall response rate (ORR)
    - complete response (CR)
    - partial response (PR)
- adverse events (AEs)
- clinical benefit rate (CBR)
- health-related quality of life (QoL)

3.1.5 Key issues

It is important to note the following criteria were to be fulfilled *a priori*:

- Only trials that measure effectiveness in the population of interest were to be included in the systematic review, i.e. women must have MBC, have tumours which are HR+/HER2+ and had no prior treatment for MBC
    - Women were to be considered to have HR+ breast cancer if they have oestrogen receptor-positive (ER+) or progesterone-positive (PgR+) tumours
- Where head to head comparisons do not exist, indirect comparisons were to be attempted
- Cost effectiveness of treatments was to be expressed in terms of incremental cost per quality adjusted life years (QALY) gained
- The time horizon for estimating clinical and cost effectiveness was to be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared
- Costs were to be considered from an NHS and Personal Social Services perspective (PSS)

3.2 Overall aims and objectives of assessment

The remit of this appraisal is to review the clinical and cost-effectiveness evidence base for LAP+AI and TRA+AI within their licensed indications for the first-line treatment of patients who have HR+/HER2+ MBC.

4 ASSESSMENT OF CLINICAL EFFECTIVENESS

4.1 Methods for reviewing effectiveness

Evidence for the clinical effectiveness of LAP in combination with an AI (LAP+AI) and TRA in combination with an AI (TRA+AI) for the first-line treatment of patients with HR+/HER2+ MBC was assessed by conducting a systematic review of published research evidence. The review was undertaken following the general principles published by the Centre for Reviews and Dissemination's guidance for undertaking reviews in healthcare.[40]

4.1.1 Identification of studies

Randomised controlled trials (RCTs) were identified by searching major electronic medical databases including MEDLINE, EMBASE and the Cochrane Library. The search strategy was broad and not limited to RCTs. Information on studies in progress, unpublished research or research reported in the grey literature were sought by searching a range of relevant databases including National Research Register and Controlled Clinical Trials. In addition, bibliographies of previous reviews and retrieved articles were searched for further studies. The search strategy used for MEDLINE is presented in Appendix 1. The same search strategies were used to identify economic evaluations.

Further attempts to identify studies were made by contacting clinical experts and examining the reference lists of all retrieved articles. The manufacturers' submissions (MS) were assessed for unpublished data.

4.1.2 Inclusion and exclusion criteria

Two reviewers (NF/MM) independently screened all titles and abstracts. Full paper manuscripts of any titles/abstracts that were considered relevant by either reviewer were obtained where possible. The relevance of each study was assessed (NF/MM) according to the criteria in Table 1. Studies that did not meet the criteria were excluded and their bibliographic details were listed alongside reasons for their exclusion. These are listed in Appendix 2. Any discrepancies were resolved by consensus.

Table 1 Inclusion criteria (clinical effectiveness) based on the decision problem

Study design	Randomised controlled trials
Population(s)	Post-menopausal women with HR+/HER2+ MBC, who have not previously received treatment for metastatic disease and for whom treatment with an AI is suitable. The following broad sub-groups are considered if data permit: • patients based on disease characteristics such as tumour burden • number of metastatic sites • disease free interval (length of time prior to onset of metastatic disease)
Intervention(s)	Lapatinib (Tyverb®/Tykerb®) in combination with an aromatase inhibitor; Trastuzumab (Herceptin®) in combination with an aromatase inhibitor.
Comparators	The two interventions will be compared with each other; The interventions will also be compared with AIs*
Outcomes	The outcome measures to be considered include: • Overall survival (OS) • Progression-free survival (PFS) • Time to progression (TTP) • Overall response rate (ORR) • Clinical benefit rate (CBR) • Adverse events (AEs) • Quality of life (QoL)
HER2+=over-expresses HER2; HR=hormone receptor-positive; MBC=metastatic breast cancer	

4.1.3 Data abstraction strategy

Data were extracted by one reviewer (MM) using a standardised data extraction form in Microsoft Word 2007 and checked independently by a second reviewer (NF). Disagreements were resolved by discussion.

4.1.4 Critical appraisal strategy

The quality of the individual clinical-effectiveness studies was assessed according to criteria based on the CRD's guidance for undertaking reviews in healthcare.[40] The assessment of risk of bias was conducted independently by both reviewers (MM, NF). Disagreements were resolved through discussion.

4.1.5 Methods of data synthesis

The results of the data extraction and quality assessment for each study are presented in structured tables and as a narrative summary. The possible effects of study quality on the effectiveness data and review findings are discussed.

It was intended by the Assessment Group (AG) that meta-analyses would be conducted where direct evidence could be pooled using a standard meta-analysis[41] and where a direct comparison between LAP+AI and TRA+AI were not possible, by indirect comparisons analyses.[42] However the AG considered it inappropriate to conduct either of the analyses, as discussed further in the next section.

4.2 Results

4.2.1 Quantity and quality of research available

Identification of studies

Once duplicates were removed, a total of 2069 references were identified (Figure 4); a scan of the titles and abstracts resulted in eleven potential records.[43-53] Four of these citations[50-53] reporting on two trials (EGF30008[51] and TAnDEM[52]) met the inclusion criteria and a further trial (eLEcTRA[54]), reported only as a conference abstract, was also suitable for inclusion following information passed on to the AG by Roche at the NICE consultation meeting in February, 2010. Thus three trials were included in the systematic review. Additional data on these trials were submitted to NICE from the manufacturer of LAP (GlaxoSmithKline[22]) and the manufacturer of TRA (Roche[39]) including included the relevant clinical study reports for the EGF30008[51] and TAnDEM trials.[52] Of the seven excluded citations, three[47-49] were excluded either because they did not examine LAP or TRA in combination with an AI or because it was a conference report in relation to TAnDEM.[52] Four citations[43-46] were excluded because they could not be obtained. Each was a Physician Data Query (identified through the Cochrane Clinical Trials library) relating to the three included trials, and indexed prior to the final study publication dates. In their submissions, both Roche and GlaxoSmithKline identified additional studies which they utilised as indirect evidence. The majority of these trials had also been identified by the AG's search but as none were limited to the HR+/HER2+ MBC population (or at least did not include sub-group analysis on the HR+/HER2+ population), they did not meet the review inclusion criteria. Reasons outlining all of the excluded citations, including those identified and included by the manufacturers, are given in Appendix 2.

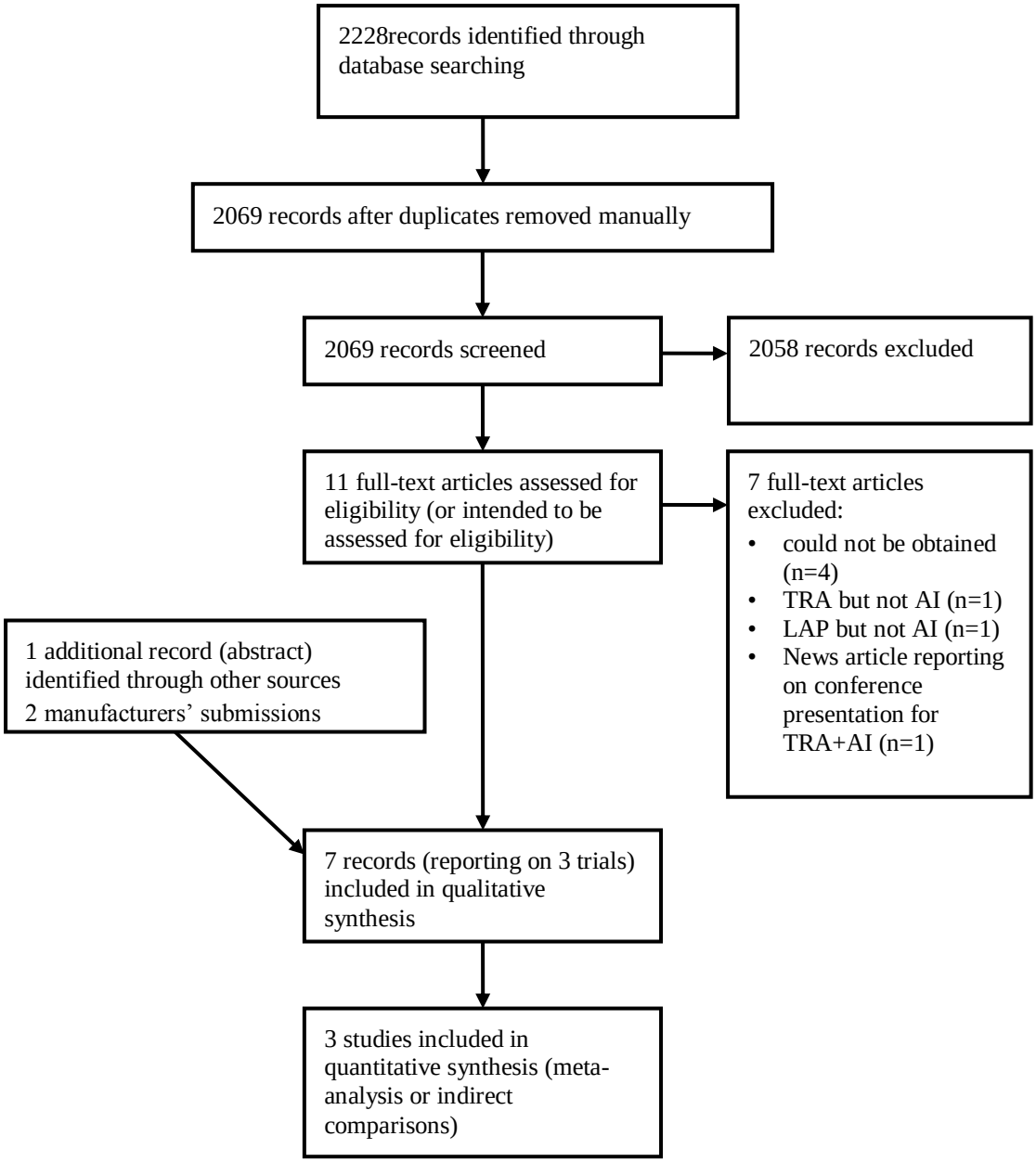

Figure 4: Identification of eligible studies

Included trials

Key characteristics of the included trials are summarised in Table 2.

Table 2 Included studies

Study and principal citation	Type of study and years of recruitment	Population	Interventions, dose and duration	Size of study	Notes
EFG30008[51]	Double-blind multicentre trial conducted internationally: 212 sites in 29 countries,2003-2006	First-line post-menopausal HR+/HER2+ MBC	LAP+LET vs LET+placebo LAP=1500 mg/day (oral) LET= 2.5mg/day (oral) placebo= pill (oral) Treatment was planned to continue until disease progression or study withdrawal	n=219[a] LAP+LET=111 LET=108	The trial was funded by GlaxoSmithKline and excluded patients with extensive symptomatic visceral disease including hepatic involvement and pulmonary lymphangitic spread of tumor, or the disease was considered by the investigator to be rapidly progressing or life threatening. It is not stated if second-line treatment was permitted following disease progression
TAnDEM[52]	Open-label multicentre trial conducted internationally: 77 sites in 22 countries (including 8 sites in the UK), 2001 - 2004	First-line post-menopausal HR+/HER2+ MBC	TRA+ANA vs ANA TRA= 4mg/kg loading dose (IV) followed by 2mg/kg/week (IV) or 8mg/kg on day 1 followed by 6mg/kg 3 weekly ANA=1mg/day (oral) Treatment was planned to continue until disease progression	n=208 TRA+ANA=103 ANA=104	TAnDEM[52] The trial was funded and conducted by Roche and permitted patients in the ANA group to cross-over to TRA+ANA following disease progression and patients in both groups were permitted chemotherapy following disease progression, i.e. patients were permitted second-line treatment. A greater proportion of patients in the ANA group received second-line treatment
eLEcTRA[54]	Open-label multicentre trial conducted internationally: 32 sites in 7 countries, 2003-2007	First-line post-menopausal HR+/HER2+ MBC	TRA+LET vs LET TRA= 4mg/kg loading dose (IV) followed by 2mg/kg/week (IV) or 8mg/kg on day 1 followed by 6mg/kg 3 weekly ANA=1mg/day (oral) Treatment was planned to continue until disease progression	n=57[b] TRA+LET=26 LET=31[b]	The trial was funded by Novartis, with Roche described as a collaborator, and halted prematurely due to slow recruitment. Patients were permitted to receive second-line TRA following disease progression. A greater proportion of patients in the LET group received second-line treatment

[a] Also included another 1059 patients who were HR+/HER2- MBC and received either LAP+LET or LET
[b] Also included another 35 patients who were HR+/HER2- MBC and received LET

AI=aromatase inhibitor; ANA= anastrozole; HER2+=over-expresses the HER2 receptor; HR+=hormone receptor positive; IV=intravenous; LAP=lapatinib; LET=leterozole; MBC=metastatic breast cancer; TRA=trastuzumab; Tx=treatment of interest

All three trials (EFG30008,[51] TAnDEM[52] and eLEcTRA[54]) were multi-centre and multi-national trials (between 7 and 29 countries) enrolling post-menopausal patients receiving first-line treatment for MBC; all three trials included patients who had HR+/HER2+ MBC although EGF30008[51] and eLEcTRA[54] also included patients who were HR+/HER2-. The trials were designed to evaluate the efficacy and safety of the addition of LAP+LET to LET (EFG30008[51]), TRA+ANA to ANA (TAnDEM[52]) and TRA+LET to LET (eLeCTRA[54]). In all trials, treatment was administered at licensed doses in which treatment was planned until disease progression at which point patients in the TAnDEM[52] and eLEcTRA trials[54] received second-line therapy which included chemotherapy; for patients in the ANA group, TRA+ANA was also a second-line treatment option. It is not stated whether patients in EGF30008[51] received any second-line therapy.

Clinical endpoints including OS, PFS and TTP that are commonly used in trials of breast cancer were utilised in at least one of the trials included in this appraisal. However, the only efficacy endpoints common to all three, and reported on by all three, were the secondary endpoints, CBR and ORR. All three trials also reported on AEs. The eLEcTRA trial[54] intended to report on OS but to date, no findings for OS have been reported, possibly because this trial was halted prematurely due to slow recruitment. Only EGF30008[51] reported on QoL.

As patients in TAnDEM[52] and eLEcTRA[54] received second-line treatment once their disease had progressed, data on OS should be treated with caution as clearly this extra treatment could potentially impact on OS. It is not stated that patients received second-line treatment in EGF30008.[51]

Data on PFS, TTP, CBR and ORR in all trials should be treated with caution due to the way they were measured. This is discussed in more detail below.

Overall, the risk of bias assessment conducted by the AG (Table 3) found EGF30008[51]to be of a good standard. Some imbalances in baseline characteristics between the groups in the HR+/HER2+ population were noted (see Table 32 in Appendix 3) which did not exist between groups in the population as a whole. The imbalances were not however deemed to be of clinical significance by the study authors or the AG. However, while the study was a double-blind study, because of the significantly increased incidences of diarrhoea and rash in the LAP+LET group (see below), the effectiveness of blinding may be questioned.

Table 3 Risk of bias table for included studies

	EGF30008	TAnDEM	eLEcTRA
Was the method used to assign participants to the treatment groups really random?	✓	✓	**?**
Was the allocation of treatment concealed?	✓	NA (open label)	NA (open label)
Was the number of participants who were randomised stated?	✓	✓	✓
Were details of baseline comparability presented in terms of prognostic factors?	✓	✓	✓
Was baseline comparability achieved in terms of prognostic factors?	✓	✓	✕
Were the eligibility criteria for study entry specified?	✓	✓	✓
Were any co-interventions identified that may influence the outcomes for each group?	✓	✓	✓
Were the outcome assessors blinded to the treatment allocation?	✓/✕	✓/✕	✕
Were the individuals who administered the intervention blinded to the treatment allocation?	✓/✕	✕	✕
Were the participants who received the intervention blinded to the treatment allocation?	✓	✕	✕
Was the success of the blinding procedure assessed?	✕	NA (open label)	NA (open label)
Were at least 80% of the participants originally included in the randomisation process followed up in the final analysis?	✓	✓	✕
Were the reasons for withdrawals stated?	✓	✓	Trial was stopped prematurely
Is there any evidence to suggest that the authors measured more outcomes than they reported?	✓/✕[a]	✓	✓
Was an intention to treat analysis included?	✓	✓	✕

✓ yes (item properly addressed)
✕ no (item not properly addressed)
✓/✕ partially (item partially addressed)
? unclear or not enough information

[a] data on TTP were only included in the MS from GlaxoSmithKline[22] which also included data on QoL outcomes which were previously reported separately in conference abstracts

TAnDEM[52] was similarly considered to be of good standard, the weakest aspect being the lack of blinding. As LAP is administered orally, the EGF30008 trial[51] was able to blind treatment by also administering a placebo pill with LET. To have blinded treatment for TRA, however, a placebo IV therapy would have been required which may arguably have been difficult to justify from an ethical perspective. Generally, baseline characteristics were well balanced across the treatment arms although there were a small number of imbalances (see Table 33 in Appendix 3). Clinical advice received by the AG was that these were not a cause for concern in terms of biasing any results.

The eLEcTRA trial[54] was deemed to be of poorer quality compared to EGF30008[48] and TAnDEM.[52] This may a reflection of poor quality reporting rather than trial design as this trial was only published as an abstract, with limited additional data subsequently available from Roche.[39] However, the fact that the trial ended prematurely due to slow recruitment did affect quality. Firstly, slow recruitment is attributed in the Roche submission[39] to the fact that investigators believed TRA+LET was superior to LET (although no evidence is presented to support this claim) and investigators were reluctant to continue randomising patients into the LET group. This could have introduced selection bias. Secondly, because fewer than 25% of the intended patients were recruited, the trial lacked statistical power and finally, there were large differences in baseline comparability (see Table 34 in Appendix 3).

Comparing baseline characteristics across the three trials was problematic because of differences in how measures were defined and/or reported. However, it was noticeable that the median age of patients in TAnDEM[52] differs to that of the other two trials, the median age being around 55 compared to around 60 in the other two trials. There also appears to be more patients with soft tissue metastases in TAnDEM[52] than EGF30008.[51]

Arguably, the most significant difference between EGF30008[48] and the other two trials is choice of exclusion criteria. According to the manufacturer's submission,[22] the EGF30008[51] trial excluded patients in which "the disease was considered by the investigator to be rapidly progressing or life threatening" (MS, pg 32).[22] The potential importance of this criterion became apparent when analysing median OS which was reported to be no greater than 23.9 months (unadjusted ITT population) in the TAnDEM trial[52] compared to 33.3 months in the EGF30008 trial.[51] If it is assumed that there is a 'class-effect'(and certainly for early breast cancer where NICE guidance[34] exists on the use of AIs, it is indeed assumed that LET and ANA are equally effective) then if the populations were truly similar, a similar median OS would be expected for patients in the LET and ANA arms of the different trials.

The generalisability of the trials to the UK population may also be questioned as it is unclear whether patients at imminent risk of death, as some patients were in the TAnDEM trial,[52] would be eligible for treatment with LAP+AI or TRA+AI. However, clinical advice received by the AG was that some clinicians would offer TRA or LAP with an AI if patients were deemed to be too unfit for chemotherapy even if they were at risk of imminent death. Equally, those who were not at risk of imminent of death, as in the EGF30008 trial,[51] are also be offered TRA or LAP with an AI. Hence, both study populations appear generalisable to the UK.

In summary, whilst study designs appear appropriate for the comparison of LAP+AI vs AI or TRA+AI vs AI , key differences in the trials led the AG to the conclusion that it would not be appropriate to pool data or make meaningful comparisons, directly or indirectly, across the two completed trials. This decision was primarily based on differences in patient populations – the key factor being the exclusion of patients in whom the disease was considered by the investigator to be rapidly progressing or life threatening as in EGF30008[48] (but not the other trials). As eLEcTRA[54] was halted prematurely some data were not available/reported, and data which were reported should be treated with extreme caution. For these reasons, the AG decided to focus on discussing the trials individually.

4.2.2 Assessment of effectiveness

EGF30008: LAP+LET vs LET

Patients were recruited into EGF30008[51] between December 2003 and December 2006, during which time there were **** amendments to the original protocol. One amendment led to increased target enrolment from 760 to 1280 subjects in order to ensure adequate statistical power in the HR+/HER2 sub-group (October 2005). The decision to focus on the HR+/HER2 sub-group was made as a result of pre-clinical and clinical studies suggesting LAP modulates its effect in breast cancer primarily via ErbB2.[55, 56] Another significant amendment was the definition of the HR+/HER2+ population as the primary population of interest, at which time the primary endpoint was changed from TTP to PFS (October 2007); PFS was defined as the time from randomisation until the earliest date of disease progression or death due to any cause, if sooner. The decision to change the primary endpoint was made because PFS includes deaths (and is thus a better correlate with OS) unlike TTP and is thus preferred by the FDA[57] and the Committee for Medicinal Products for Human Use.[58] All EGF30008[48] amendments were carried out prior to un-blinding and are less likely to increase risk of bias than if amendments had occurred after blinding.

As well as protocol amendments, there were also reported cases of protocol violations. Although the proportion of subjects with protocol violations was ****** ******Confidential information removed (** subjects [*%] and ** subjects [*%] in the LAP+LET and LET groups, respectively), the specific

violations varied. For example, there were more cases in the ***** group with violations such as "failure to take study drug" or "received incorrect treatment assignment" than in the ***** group. The most frequent protocol violation in either group was ***** as per protocol (PP) (* subjects [*%] in ***** and * subjects [*%] in *****). It should be noted these relate to all patients in the study, not just patients who are HR+/HER2-.

The findings from the EGF30008 trial[51] are summarised in Table 4 where it can be seen data were available for the HR+/HER2+ population as well as the wider population of patients recruited, which included patients who were HR+/HER2-. The wider population in the study was referred to as the intention to treat (ITT) population.

Table 4 Summary of efficacy results from the EGF30008 trial

	HR+/HER2+ population[a]			**All patients, i.e. including those who are HR+/HER2-[b]**		
	LAP+LET (n=111)	**LET (n=108)**	**HR (95% CI) OR (95% CI) p-value**	**LAP+LET (n=644)**	**LET (n=642)**	**HR (95% CI) OR (95% CI) p-value**
OS (months)[c]	32.3	33.3	HR=0.74 (0.5 to 1.1) p=0.113	not reported	not reported	not reported
PFS (months)[c]	8.2	3.0	HR=0.71 (0.53 to 0.96) p=0.019 Cox regression analysis (adjusting for known baseline prognostic factors) HR=0.65 (0.47 to 0.89) p=0.008	11.9	10.8	HR=0.86 (0.76 to 0.98) p=026
TTP (months)[c]	8.2[d]	3.0[d]	HR=0.71 (0.53 to 0.96) p=0.019	not reported	not reported	not reported
ORR[e]	28%	15%	OR=0.4 (0.2 to 0.9) p=0 .021	33%	32%	OR not reported p=0.726
- CR	5%	4%		5%	4%	
- PR	23%	11%		28%	27%	
SD≥ 6 months[e]	20%	14%	not reported	26%	25%	not reported
CBR[f]	48%	29%	OR=0.4 (0.2 to 0.8) p=0.003	58%	56%	OR not reported p=0.761

CBR=clinical benefit rate; CI=confidence interval; CR=complete response; ORR=overall response rate; OR=odds ratio; OS=overall survival; PFS=progression-free survival; PR=partial response; SD=stable disease; TTP=time to progression

a median follow-up of 1.8 years
b median follow-up of 2 years
c median (95% CIs were not presented)
d TTP data only presented in the GlaxoSmithKline submission
e data only presented as percentages
f CBR=CR, PR or SD≥ 6 months

No significant differences were reported in terms of OS between the groups, although there was a possible trend in favour of LAP+LET compared with LET.[22] A pre-planned analysis within known prognostic factor sub-populations reported consistently improved OS with LAP+LET compared to

LET in the following groups: ECOG performance status score <1 and patients with fewer than three metastatic sites; site of disease (non-visceral/visceral) did not significantly affect OS.[22]

For the HR+/HER2+ population, EGF30008[51] reported significant improvements in PFS in the LAP+LET group when compared to the LET group.[51] When adjusted for baseline prognostic factors, the stepwise Cox regression analysis for PFS confirmed the benefit of LAP+LET compared to LET.[53] A pre-planned analysis within known prognostic factor sub-populations reported consistently improved PFS with LAP+LET compared to LET in the following groups: patients with an ECOG performance status score >0, patients without bone as the only site of metastasis, patients with and without liver metastases, patients with fewer than three metastatic sites and patients having received prior endocrine therapy for <6 months.[53] Significant differences were also reported for differences in PFS in the ITT population[51] but here the differences between the groups were less pronounced. In particular, it was noticeable that the PFS was greater in the ITT population compared to the HR+/HER2+ population, particularly for patients receiving LET (the difference in PFS between the ITT and HR/HER2+ populations here was 7.8 months compared to 3.7 months between the same two populations amongst patients receiving LAP+LET).

Because only one subject of the HR+/HER2+ population died from a cause other than breast cancer in this study, the TTP findings were almost identical to those reported for PFS.[22] In the same population, ORR was significantly improved for patients treated with LAP+LET compared to LET as was CBR. However, the differences in ORR and CBR between treatment groups were not significant in the ITT population.

Assessment of disease progression is liable to subjectivity so introducing observation bias which needs to be considered when interpreting PFS, TPP, ORR and CBR. Blinded independent review has been recommended in order to circumvent such problems.[59, 60] In the EGF30008 trial,[51] investigator assessment and a blinded Independent Radiological Review Committee (IRC) were employed. Comparison of PFS assessment results reported a **% concordance between the assessments that were made by the IRC and the investigators (in the ITT population, this rose to **%). The main reasons for differences as noted in the clinical study review (CSR) were primarily due to differences in the censoring methods used by the investigator and the IRC, thus the PFS assessed by IRC was **** compared to the investigator assessments. The differences were however constant which reduces the risk of bias.

Patients who received LAP+LET were more likely to experience AEs, with nearly all patients in the HR+/HER2+ population experiencing an AE compared to around three quarters of patients who received LET (Table 5). Serious adverse events (SAEs), however, were relatively rare in both groups.

Only three patient deaths were attributed to treatment, one of these taking LAP+LET in the HR+/HER2+ population.

Table 5 Summary of adverse events from the EGF30008 trial

	HR+/HER2+ population		All patients, i.e. including those who are HR+/HER2-	
	LAP+LET (n=111)	LET (n=108)	LAP+LET (n=644)	LET (n=642)
AEs	96%[a, b]	77%[a, b]	not reported	not reported
SAEs	not reported	not reported	8%[b,c]	4%[b,c]
Discontinued treatment due to AE	not reported	not reported	2%[b,c]	1%[b,c]
Treatment related deaths	1 (<1%)	0	1 (<1%)	2 (<1%)

AE=adverse event; SAE=serious adverse event

a Data taken from Schwarzberg et al 2010[53] and so only available for HR+/HER2+
b data only presented as percentages
c Data only available for all patients, i.e. including those with HR+/HER2not reported MBC and for discontinuing treatment, this is given only for diarrhoea, it is not known if other AEs resulted in discontinuation of treatment but it is assumed not

In patients with HR+/HER2+ MBC and in all patients as a whole (i.e. including HR+/HER2- MBC), the most common AEs were diarrhoea, rash, nausea, arthralgia, and fatigue, of which the majority were grade 1 or 2 (Table 6 and Table 7). In particular, incidences of diarrhoea, rash and nausea were significantly greater in patients receiving LAP+LET. It was reported by Johnston et al 2009[51] that 15% of all 60 patients with grade 3 or grade 4 diarrhoea discontinued LAP+LET as a result, i.e. around 1% of all patients. For the remainder of patients, diarrhoea was managed by dose reduction (19%), dose interruption (36%), or supportive intervention without treatment dose adjustments (31%).

Table 6 Most common adverse events recorded in the HR+/HER2+ population in EGF30008*

Adverse events*	LAP+LET (n=111)					LET (n=108)				
	Grade 1	Grade 2	Grade 3	Grade 4	All	Grade 1	Grade 2	Grade 3	Grade 4	All
Diarrhoea	38%	23%	7%	0	68%	8%	<1%	0	0	8%
Rash	30%	16%	0	0	46%	8%	<1%	0	0	8%
Nausea	20%	7%	0	0	27%	15%	2%	<1%	0	18%
Fatigue	12%	6%	4%	0	22%	8%	6%	0	0	14%
Arthralgia	10%	4%	4%	0	18%	15%	4%	<1%	0	20%
Back pain	8%	7%	2%	0	17%	4%	5%	<1%	0	9%
Vomiting	12%	4%	<1%	0	17%	6%	<1%	0	0	7%
Headache	8%	6%	0	0	14%	7%	4%	<1%	0	11%
Asthenia	7%	5%	2%	0	14%	8%	2%	0	0	9%
Pruritus	9%	4%	0	0	13%	2%	2%	<1%	0	5%
Dizziness	8%	4%	0	0	12%	8%	0	0	0	8%
Cough	8%	3%	0	0	11%	7%	3%	0	0	9%
Alopecia	11%	0	0	0	11%	4%	0	0	0	4%
Musculoskeletal pain	4%	4%	<1%	0	10%	3%	2%	0	0	5%
Epistaxis	7%	2%	<1%	0	10%	<1%	<1%	0	0	2%
Dyspnea	4%	4%	0	<1%	9%	4%	3%	4%	0	10%
Hot flush	5%	<1%	0	0	6%	9%	3%	0	0	12%
Alanine Aminotransferase increase	7%	3%	<1%	0	11%	4%	<1%	<1%	0	6%
Aspartate Aminotransferase increase	6%	3%	<1%	0	10%	3%	0	2%	0	5%

* Events reported in ≥10% of patients in any group taken from Schwarzberg et al 2010;[53] discrepancies between values in the total column and the addition of the incidence rates reported for grades 1, 2, 3, and 4 are a result of mathematical rounding

Table 7 Most common adverse events recorded in all patients in EGF30008

Adverse events	LAP+LET (n=654)					LET (n=624)				
	Grade 1	Grade 2	Grade 3	Grade 4	All	Grade 1	Grade 2	Grade 3	Grade 4	All
Diarrhoea*	32%	22%	9%	<1%	64%	15%	4%	1%	0	20%
Rash *	28%	15%	1%	0	45%	11%	2%	0	0	13%
Nausea *	22%	8%	1%	0	31%	14%	6%	1%	0	21%
Arthralgia	12%	6%	1%	0	19%	16%	6%	1%	0	23%
Fatigue	12%	7%	2%	0	20%	10%	7%	0	0	17%
Back pain	8%	6%	2%	0	16%	7%	7%	2%	<1%	16%
Vomiting *	10%	6%	1%	<1%	17%	7%	3%	1%	<1%	11%
Headache	10%	4%	<1%	0	14%	8%	4%	0	0	13%
Cough	9%	3%	<1%	0	12%	12%	2%	0	0	14%
Hot flush *	8%	2%	0	0	11%	10%	4%	0	0	15%
Asthenia	8%	3%	1%	0	12%	7%	3%	1%	0	11%
Pain in extremity	6%	4%	<1%	0	10%	7%	4%	1%	0	11%
Dyspnea	5%	4%	1%	<1%	10%	6%	4%	1%	<1%	12%
Pruritus *	8%	4%	<1%	0	12%	7%	2%	0	0	9%
Alopecia *	13%	<1%	<1%	0	13%	7%	<1%	0	0	7%
Constipation	8%	1%	0	0	9%	8%	3%	<1%	0	11%
Anorexia	8%	2%	1%	0	11%	5%	3%	<1%	0	9%
Dry skin	11%	2%	<1%	0	13%	4%	<1%	0	0	4%
Epistaxis	10%	1%	<1%	0	11%	1%	0	<1%	0	2%
Nail disorder	9%	2%	<1%	0	11%	1%	0	0	0	1%

* A statistically significant ($p<0.05$) effect was reported between treatment groups for the total incidence of these adverse events in Johnston et al 2009[51]

An additional 8 months of data beyond trial reporting have been collected (through to 3 February 2009) and presented in the GlaxoSmithKline submission.[22] These data remain consistent with the initial study results although more patients in the LET group reported AEs than before: 629 (96%) patients reported an AE in the LAP+LET group compared to 537 (86%) in the LET group.

Overall, therefore, no new safety issues were identified, the safety profile of LAP+LET being consistent with the safety profiles of both drugs when given as single agents and with safety data from previously reported LAP studies.

Finally, QoL was also assessed in the EGF30008 trial[51] utilising the functional assessment of cancer therapy-breast (FACT-B) questionnaire.[22] Within the HR+/HER2+ population, QoL scores and changes from baseline were reported to be generally stable over time for subjects who stayed in the study in both the LAP+LET and LET groups, suggesting maintenance of QoL. The Quality-Adjusted Time Without Symptoms and Toxicity (Q-TWIST) difference between treatment groups for the HR+/HER2+ population ranged from 8 to 9.5 weeks, favouring LAP+LET over LET for all hypothetical utility levels, although none of the findings were reported to be statistically significant.[61]

TAnDEM: TRA+ANA vs ANA

Between March 2001 and May 2006, TAnDEM[52] enrolled 207 HR+/HER2+ patients, of which 103 were randomly assigned to TRA+ANA and 104 to ANA. According to the CSR, prior to any patient being recruited into the trial, there had been **** amendments to the protocol ******. Progression-free survival was defined as the time between random assignment and the date of progressive disease (PD), clinical or radiographic, or death. Time to treatment progression was defined as time between random assignment and. There were also *** amendments following recruitment of the first patient. The second of these amendments, ****, is perhaps the most significant as this allowed for cross-over of patients from ANA to TRA+ANA following disease progression thus impacting on the size of the OS results. No statistical methods were described **** to address this issue of cross-over *a priori*, the trial simply being separated into two treatment phases: main and extension. The main phase was defined as the first 24 months of treatment or until disease progression, and the extension phase was defined as the treatment period after 24 months or the treatment period after disease progression, whichever came earliest. Patients had a safety follow up assessment 28 days after their last dose of treatment. Subsequently, post-hoc analyses were performed by Roche which attempted to take into account the effects of cross-over as described further below.

The manner in which the protocol is implemented in a clinical trial should be clear to all principal investigators to ensure that the same systems and procedures are in place across all centres to reduce protocol violations. This appeared not to be the case in the TAnDEM trial[52] where a few major protocol violations were identified causing the exclusion of one patient from full analysis and 15 from the PP analysis of efficacy. These major protocol violations were: HER2 over expression/amplification not documented; no protocol-specified tumour (no metastatic disease); poor study drug compliance and prior chemotherapy.

The findings from TAnDEM[52] are summarised in Table 8. No significant differences in terms of OS were found between the groups. However, it should be noted that 70% of the patients randomised to ANA subsequently received TRA+ANA in the extension phase of the trial and this may have impacted on the findings. In addition, around a third (31%) of ANA patients went on to receive chemotherapy compared to a minority of patients who had been randomised to TRA+ANA (8%). With some legitimacy, the manufacturer of TRA argued that this could impact on the size of the OS estimates because, in this situation, the ITT results will be significantly compromised and will either under or over-estimate the treatment effect between groups. Roche contended that this would under-estimate the treatment effect and show a reduced incremental gain from TRA+ANA over ANA. Thus unplanned exploratory post hoc analyses were performed by Roche to investigate the impact of this cross-over from the control group of the trial on OS.

Table 8 Summary of efficacy results from the TAnDEM trial[a]

	TRA+ANA (n=103)	ANA (n=104)	HR (95% CI) p-value
OS (months)[a] unadjusted ITT population[b]	28.5 (22.8 to 42.4)	23.9 (18.2 to 37.4)	HR =0.84 (0.59 to 1.20) p=0.325
OS (months)[a] centrally confirmed HR status[b]	34.1 (23.9 to 52.0)	28.6 (17.4 to 40.0)	HR=0.85 (no CIs) p=0.451
OS (months)[a] adjusted for cross-over by RPSFT (ITT population?)	28.52[c]	21.98[c]	HR =0.73 (0.51 to 1.04) p=not reported
OS (months)[d] PP analysis (patients who did not cross-over)	28.5 (22.8 to 42.4)	17.2[e]	p=0.218[f] p=0.048[g]
PFS (months)[a] ITT population[b]	4.8 (3.7 to 7.0)	2.4 (2.0 to 4.6)	HR=0.63 (0.47-0.84) p=0.0016
PFS (months)[a] centrally confirmed HR status[b]	5.6 (3.8 to 8.3)	3.8 (2.0 to 6.3)	HR=0.62 (no CIs) p=0.006
PFS (months)[a] updated[d]	5.8 (4.6 to 8.3)	2.9 (2.1 to 4.5)	HR= 0.55 (0.41 to 0.74) p<0.0001
TTP (months)[a] ITT population[b]	4.8 (3.7 to 7.7)	2.4 (2.0 to 4.6)	HR not reported p=0.0007
TTP (months)[a] centrally confirmed HR status[b]	5.6 (3.8 to 8.3)	3.9 (2.1 to 6.3)	HR=0.62 (no CIs) p=0.0007
	TRA+ANA (n=74)	**ANA (n=73)**	**OR (95% CI) p-value**
ORR centrally confirmed hormone receptor status[b]	20%	5%	OR not reported p=0.018
- CR	0	0	
- PR	20%	5%	
SD≥ 6 months	38%	38%	not reported
CBR[h] ITT population[b]	43% (33% to 53%)	28% (20% to 38%)	OR not reported p=0.026

CBR=clinical benefit rate; CI=confidence interval; CR=complete response; OR=odds ratio; ORR=overall response rate; OS=overall survival; PFS=progression-free survival; PR=partial response; RPSFT= rank preserving structural failure time; SD=stable disease; TTP=time to progression

[a] median (95% CI)
[b] the ITT data constituted data from local investigator assessments, the centrally confirmed data was that confirmed by a blinded Response Evaluation Committee
[c] The RPSFT adjustment was made only in the submission by Roche;[39] No CIs were presented for median OS
[d] The updated PFS was data from a later cut-off point; it is unclear whether this was centrally confirmed
[e] n=31; no CIs presented for OS
[f] Log-rank Test
[g] Wilcoxon test
[h] CBR=CR, PR or SD≥ 6 months

Currently there is no uniform agreement about which is the best method(s) to use for adjusting for cross-over. In the published paper,[52] an attempt to highlight the impact of cross-over on OS was explored using a PP analysis approach in which the median OS for patients receiving TRA+ANA (28.5 months, n=103) from randomisation was compared to the sub-group of patients who initially received ANA and did not cross-over to receive TRA (17.2 months, n=31). By log rank testing, there was no significant difference in the OS analysis (p=0.218). However, due to the small number of patients with long survival times available for analysis, the Wilcoxon test was also used as this gives more weight to early time points than the log rank test. The analysis using the Wilcoxon test resulted

in a modestly statistically significant difference (p=0.048). Similarly, comparing patients in the ANA group who crossed over to receive TRA+ANA (n=73) with those who did not (n=31) resulted in an OS estimate of 25.1 months and 17.2 months respectively. These differences were reported to be not statistically significant using the log rank test (p=0.358) but were statistically significant using the Wilcoxon test (p=0.040). The AG notes that where there is a relatively large proportion of patients who cross-over, these PP approaches are prone to selection bias.

In their submission, the cross-over adjustment employed by Roche[39] was based on a rank preserving structural failure time (RPSFT) approach initially proposed by Robins and Tsiatis[62] and later modified by Mark and Robins.[63] Using the RPFST approach (ITT population), median OS in the TRA+ANA group becomes 28.52 months and the median OS in the ANA group becomes 21.98 months. The RPSFT method is an accelerated failure time model, a form of randomisation-based analysis that more effectively preserves the integrity of randomisation than do PP analyses. There are a number of key assumptions to the RPSFT approach including:

- subjects who cross-over are similar to those who do not with regard to important prognostic factors
- no treatment interaction occurs
- the distribution of subjects who do not experience an event is identical between the randomised treatment groups

As such, the validity of the RPSFT method has been questioned when imbalances occur post-randomisation, e.g. when there is an unequal distribution of patients receiving second-line treatment across the arms.[64]

The use of the RPSFT approach in the current appraisal was justified by the manufacturer since it has also been used for two other NICE appraisals: sunitinib for the treatment of gastrointestinal stromal tumours[65] and everolimus for the second-line treatment of advanced renal cell carcinoma.[66] However, in the former appraisal, only 7% of patients crossed-over to receive sunitinib whereas in the latter, 81% of patients crossed over to receive everolimus. In the sunitinib appraisal,[65] because so few patients crossed-over from the control arm to sunitinib, the Appraisal Committee had confidence in the results from the RPSFT as well as the PP analysis which was also performed. For everolimus,[66] two different methods were employed to adjust for cross-over by the manufacturer, the Inverse Probability Censoring Weight approach[67] and the RPSFT approach. Because 81% of people had crossed over to receive everolimus, the Appraisal Committee agreed that it was appropriate to adjust the results utilising statistical methods such as these to control for cross-over. However, as both methods gave different OS estimates,[66] it was unclear which method, if either, was most suitable.

A recent paper by Morden et al[64] explored various approaches to adjusting for cross-over using a simulation exercise. Methods tested included PP approaches and accelerated failure time model

methods. The authors found that that when there is cross-over from the control group, commonly adopted approaches such as censoring at the time of cross-over, or considering treatment as a time-dependent covariate, may be associated with biased estimates of the true treatment effect where the reasons for the cross-over are strongly related to their underlying prognosis. Where patients who cross-over are excluded from the analysis altogether (i.e. a PP analysis), biases were reported to be small in situations with a low proportion of switchers (as was the case, for example, with sunitinib). However, as the number of patients who switch increases, the risk of bias was also reported to increase.

Specifically, with regard to accelerated failure time model methods, three methods were considered by Morden et al[64] - the RPSFT developed by Robins and Tsiatis,[62] the iterative parameter estimation algorithm approach[68] (which is a modification of the RPSFT method in which the test-based estimation is replaced with a likelihood-based analysis) and a parametric randomisation-based method (which as an extension to the previous two methods).[69] The findings from their simulation exercise suggested that the RPSFT[62] and the iterative parameter estimation algorithm[68] gave estimates close to the true treatment effect whereas the parametric randomisation-based method[69] over-estimated the true treatment effect . The iterative parameter estimation algorithm[68] appeared to be the most accurate method when the proportion of patients who crossed-over was relatively high.

In TAnDEM,[52] the AG questions whether the key assumptions underlying the RPSFT method hold given only patients who have progressive disease were eligible to cross-over. The AG also notes that the proportion of patients who crossed-over was relatively high, being around 70%, which as Morden et al[64] report, increases the likelihood of bias. However, the AG does agree that attempts to adjust for cross-over are worthwhile; ideally different randomisation-based methods should be used to compute and compare a range of OS estimates to assess sensitivity of treatment effects, the applicability of each individual method employed depending on the trial circumstances and characteristics; in the MS such sensitivity is not investigated. It should be noted that in order to undertake such analyses, individual patient data are required. Such data were not available to the AG and thus the AG was unable to employ any of the aforementioned approaches. Thus the AG has utilised its own method for adjusting for cross-over for the purposes of conducting its economic analysis. This is described further in section 5.4.3.

Assessment of disease progression (and therefore PFS, TTP, ORR and CBR) may be prone to subjectivity and thus to observation bias. Three universally accepted methods[59, 60] and procedures for assessing disease progression were however employed in the TAnDEM trial:[52] an investigator assessment (ITT), and a centrally confirmed assessment by a Response Evaluation Committee (REC) and, in situations where the investigator assessment was different from the REC assessment, an independent oncologist was appointed to make a reconciliation assessment. Results from both

methods demonstrated a statistically significant PFS in favour of TRA+ANA (Table 8). Patients in the TRA+ANA group experienced significant improvement in PFS and TTP. Significant differences were also reported in terms of ORR and CBR in the ITT population although interestingly, no complete response (CR) was recorded for any patient, the difference occurring as a result of improvements in partial response (PR) in the TRA+ANA group.

Patients who received TRA+ANA were more likely to experience AEs, with nearly 90% experiencing an AE compared to 65% of patients who received ANA (Table 9). Serious adverse events were also more common in the TRA+ANA group, nearly 25% experiencing an SAE compared to less than 10% of patients receiving ANA. There were no treatment-related deaths in either group.

Table 9 Summary of adverse events from the TAnDEM trial

	TRA+ANA (n=103)	**ANA (n=104)**
AEs	87%	65%
SAEs	23%	6%
Discontinued treatment due to AEs	9%	1%
Treatment related deaths	0	0
AE=adverse event; SAE=serious adverse event		

The most frequently reported AEs in both groups were: fatigue, diarrhoea, vomiting and arthralgia of which the majority were grade 1 or 2 (Table 10). Adverse events were more common in the TRA+ANA group than the ANA group although it should also be noted that duration of treatment was longer in the TRA+ANA group and that the open label design of the study meant that AEs in the ANA group were reported only until the patients crossed over to TRA+ANA.

Table 10 Most common adverse events recorded in patients in the TAnDEM trial[a]

	TRA+ANA (n=103)				ANA (n=104)			
Adverse events[a]	**Grade 1 / 2**	**Grade 3**	**Grade 4**	**All**	**Grade 1 / 2**	**Grade 3**	**Grade 4**	**All**
Fatigue	20%	1%	0	21%	10%	0	0	10%
Diarrhea	19%	1%	0	20%	8%	0	0	8%
Vomiting	18%	3%	0	21%	4%	1%	0	5%
Arthralgia	15%	0%	0	15%	9%	1%	0	10%
Pyrexia	17%	0%	0	18%	7%	0	0	7%
Back pain	13%	2%	0	15%	5%	2%	0	7%
Dyspnea	11%	1%	1%	13%	9%	0	0	9%
Nausea	16%	1%	0	17%	5%	0	0	5%
Cough	14%	0	0	14%	6%	0	0	6%
Headache	14%	0	0	14%	6%	0	0	6%
Nasopharyngitis	17%	0	0	17%	2%	0	0	2%
Bone pain	9%	2%	0	11%	6%	0	0	6%
Constipation	12%	0	0	12%	5%	0	0	5%
Chills	14%	1%	0	15%	0	0	0	0
Hypertension	5%	2%	0	7%	0	4%	0	4%

[a] Most common AEs are those of any grade occurring at an incidence rate of >10% in either treatment group and/or those of grade 3 or 4 occurring at a frequency of >2% in either treatment group
NB. AEs reported in the ANA group were only recorded prior to cross-over

Overall, therefore, no new safety issues were identified; the safety profile of TRA+ANA being consistent with the safety profiles of both drugs when given as single agents and with safety data from previously reported TRA studies.

eLEcTRA:TRA+LET vs LET

The eLEcTRA trial[54] planned to enrol 370 patients with HR+ MBC but between 2003 and 2007, only enrolled 92 patients, at which point the study was halted due to slow recruitment. The slow recruitment is attributed in the Roche submission[39] to the fact that investigators believed TRA+LET was superior to LET (although no evidence is presented to support this claim).When the trial was halted, patients who were HR+/HER2+ had been randomly assigned to TRA+LET (n=26) or LET (n=31) and patients who were HR+/HER2- had been assigned to receive LET (n=35).

The findings from eLEcTRA[54] are summarised Table 11where a large difference in TTP was observed between the two treatment groups although this difference was not statistically significant (HR=0.67; p=0.23). Interestingly, significant differences were however reported for differences in TTP between the two cohorts of patients that received LET (median: 15.2 months vs. 3.3 months for HR+/HER2+ MBC vs. HR+/HER2- MBC respectively; HR=0.71; p=0.03). Large differences were also observed between TRA+LET and LET for ORR and CBR but again these differences were not statistically significant (p=0.3124 and p=0.0636 respectively).

Table 11 Summary of efficacy results from the eLEcTRA trial

	TRA+LET (n=26)	LET (n=31)
TTP (months)[a]	14.1	3.3
ORR	27%	13%
CBR[b]	65%	39%
[a] median (no 95% CIs presented) [b] CBR not defined		
CBR=clinical benefit rate; CI=confidence interval; ORR=overall response rate; PR=partial response; TTP=time to progression		

Patients who received TRA+LET were slightly more likely to experience SAEs and/or "clinically significant AEs" (which were not defined) (Table 12). The most common AEs for patients in either group were musculoskeletal and connective tissue disorders and gastrointestinal disorders, while infections were also relatively common in both groups, particularly the TRA+LET group (Table 13). Fatigue was a problem particular to TRA+LET patients and around 15% experienced hepatobiliary disorders whereas no patient in the LET group experienced these AEs.

Table 12 Summary of adverse events from the eLEcTRA trial

	TRA+LET (n=26)	LET (n=31)
SAEs	27%	23%
SAEs and/or clinically significant AEs [a]	39%	36%
Discontinued treatment due to SAEs and/or clinically significant AEs [a]	4%	0
Death during treatment	0	3.2%
AE=adverse event; SAE=serious adverse event		
[a] clinically significant AEs not defined		

Notwithstanding the caveats raised by the AG in comparing data across trials, it is also impossible to compare the AE profiles of TRA+LET in eLeCTRA[54]with TRA+ANA in TAnDEM[52] or LET in EGF30008[51] because of the different ways in which AEs have been categorised, with the possible exceptions of fatigue and headaches. However, comparisons with eLEcTRA[54] are arguably still inappropriate given the small number of patients in this trial and the fact that the trial was halted early.

Table 13 Adverse events recorded in patients in the eLEcTRA trial

Adverse events	TRA+LET (n=26)	LET (n=31)
Musculoskeletal and connective tissue disorders	57.7%	38.7%
Gastrointestinal disorders	57.7%	32.2%
Infections	30.8%	16.1%
Fatigue	26.9%	0
Metabolism disorders	20%	3.2%
Headache	19.2%	9.7%
Hepatobiliary disorders	15.4%	0
Bone fractures	7.7%	6.5%
Psychiatric disorders	3.8%	16.1%
Hot flushes	7.7%	3.2%
Cardiac events	7.7%	9.7%

As noted above, there are known concerns about cardiac safety associated with TRA but there were fewer cardiac events recorded in the TRA+LET group compared to the LET group.

Overall, therefore, no new safety issues were identified, the safety profile of TRA+LET being consistent with the safety profiles of both drugs when given as single agents and with safety data from previously reported TRA studies.

While there were three trials[51, 52, 54] identified that compared the interventions of interest with a comparator of interest in the relevant population, only two of these trials[51, 52] were completed as intended. These two trials were primarily sponsored by the manufacturers of LAP (GlaxoSmithKline) and TRA (Roche) and it was from these two manufacturers that the MS[22, 39] were received. In both of the MS, the manufacturers reported and appraised each of the pivotal trials individually, , an approach also undertaken by the AG. Unlike the AG, however, Roche also performed a meta-analysis and both manufacturers also conducted indirect comparisons analyses in order to compare to LAP+AI to TRA+ AI.

Meta-analysis (Roche)

The fixed effect standard meta-analyses undertaken by Roche[39] examined PFS and were conducted for ANA vs TAM (two trials[70, 71]) and ANA vs megestrol acetate (two trials[72, 73]). There were insufficient trials to conduct meta-analyses for any other comparisons, such as LAP+LET vs LET or TRA+AI vs AI.

For the meta-analysis, forest plots for HR for individual studies and pooled studies were presented. The I^2 statistic was calculated to assess the potential heterogeneity between studies. The studies[70-73] included in these meta-analyses appeared to be associated with statistical and clinical heterogeneity. No significant differences were found for PFS between treatment groups in either meta-analysis.

Given that ANA was being compared to TAM or megestrol acetate, and given it was unclear how many, if any, patients were HR+/HER2+, the AG believes that the relevance of these analyses to the current appraisal are limited. They were however relevant to the Roche submission[39] because the results from these meta-analyses were used in their indirect comparisons analyses.

Indirect comparisons analyses

Both manufacturers performed indirect comparisons analyses, although different approaches were employed, as summarised in Table 14.

A complex network meta-analysis using the methods described by Puhan et al[74] was planned by GlaxoSmithKline[22] but was not possible due to the lack of data for the outcomes of interest: OS and PFS/TTP. Thus adjusted indirect comparisons analyses were performed for single outcomes as available using the methods and principles as described by Bucher et al[75] and incorporated data from five studies; EGF30008[51] and TAnDEM[52] were included as well as one study comparing LET (2.5mg/day) to tamoxifen (TAM) (20mg/day)[76] and two studies comparing ANA (1mg/day) to TAM (20mg/day).[70, 71] The eLEcTRA[54] study was not included in the GlaxoSmithKline analyses as it was only published as an abstract and the AG agrees with the manufacturer's argument that a lack of sufficient data from this trial justifies its exclusion.

The findings for both OS and PFS/TTP are summarised in Table 35 and Table 36 in Appendix 4. The results suggest there are no significant differences between any of the interventions for OS. Both LAP+LET and TRA+LET result in significantly improved outcomes for PFS/TTP when compared to ANA, LET and TAM. For reasons discussed below, the AG believes these findings should be treated with extreme caution.

Roche[39] employed an indirect network meta-analyses based on a Bayesian approach[77] in which a number of different analyses were performed for OS (base case of 12 trials,[51, 52, 70-73, 78-83]) and PFS (base case of seven trials,[51, 52, 70-72, 79, 82]). A number of assumptions were made and tested by sensitivity analyses. These included an assumption that PFS=TTP (which enabled four additional trials[54, 78, 80, 81] to be considered) and that OS findings for TAnDEM[52] based on the RPSFT adjustment should be used in the base case. In addition, for every outcome, the assumption that AIs hold a 'class effect' (i.e. LET=ANA, as suggested by clinical experts and as found in a head to head trial of second-line ANA vs LET[83]) was tested. This assumption related to the mixed HER2 status population (i.e. the population in which the proportion of patients with HER2+ breast cancer was unknown, as in the aforementioned ANA vs LET trial[83]). The mixed HER2 population was chosen because the HR+/HER2+ population was too specific to allow the inclusion of any trials other than EGF30008,[51] TAnDEM[52] and eLEcTRa.[54]

The findings from four of the main analyses presented by the manufacturer, in which a 'class-effect' was tested and then assumed as a result, are summarised in Table 37 to Table 44 in Appendix 4. In the final analyses, in which a class-effect was assumed for AIs, the evidence was derived from EGF30008[51] and TAnDEM[52] for OS and PFS and from EGF30008,[51] TAnDEM[52] and eLEcTRa[54] for PFS/TTP. These suggest there are no significant differences between LAP+LET and TRA+ANA for OS, PFS or PFS/TTP. For reasons discussed below, the AG believes these findings should be treated with extreme caution.

Aside from the fact that EGF30008[51] and TAnDEM[52] were too dissimilar in terms of patient populations, the AG believes that both the manufacturers' indirect comparisons analyses had one other major limitation, namely that the basic requirement for indirect comparisons with regard to exchangeability of relative treatment effect between trials in the two MS could not be assumed. This is a limitation recognised by the manufacturers themselves[22, 39] and is amplified when patient population characteristics are considered. Crucially, it was unknown if patients with HR+/HER2+ MBC were included in the trials and if so, how many patients were in fact HR+/HER2+. Only three trials[51, 52, 54] presented data for patients with HR+/HER2+ MBC. The other trials[70-73, 76, 78, 79, 83-91]

included patients with mixed/unknown status and in many instances, patients who had advanced breast cancer. While the inclusion of patients with advanced breast cancer may arguably be of less concern, the importance of missing data on HR+/HER2+ status is two-fold. Firstly, patients with HR+/HER2+ MBC are the population of interest to this review and as Roche have acknowledged (MS, pg 18):[39] "All results should be treated with extreme caution when applied to the co-positive population as there is no evidence base capable of informing this analysis in the population specified by the decision problem." Secondly, both the EGF30008[51] and eLEcTRA trial[54] suggest that the effects of LET in patients with HR+/HER2+ MBC tumours are significantly compromised when compared to those with HR+/HER2- MBC. Thus the indirect comparisons analyses may be over-stating the benefit of AIs and if so, there is a need to adjust for the results based on HER2+ status. However, given the proportion of such patients is unknown, such adjustments are currently impossible. It is important to note, as Roche has also stated (MS, pg 18):[39] "...understanding of HER2 was not fully developed at the period when most of the evidence base identified was formed as many of the trials conducted were not stratified for HER2 positivity and it is clearly plausible that an imbalance in this strong indicator of extremely poor prognosis could have biased the estimates of relative efficacy generated."

Roche[39] also acknowledges a number of additional limitations to their indirect comparisons analyses, namely "the low number of trials by pairwise comparison, the heterogeneity in the length of follow-up observed in the selected studies and the different methods used to adjust for cross-over in the individual studies." (MS, pg 17) A final limitation is the fact that not all trials included patients

receiving first-line treatment. In fact, two trials[79, 83] were second-line, including the trial by Rose et al[83] that was a key trial for suggesting a 'class-effect' for AIs. The AG believes that pooling trials with different lines of treatment is inappropriate and misleading, thus these results should be interpreted with caution.

In summary, given the limitations described above, the AG believes that conducting indirect comparisons analyses with the limited data available is inappropriate. Therefore any findings generated from these analyses should be treated with caution.

Table 14: Comparison of indirect comparisons approaches undertaken by the manufacturers

	GlaxoSmithKline	Roche
Population included	Post-menopausal women with HR+ MBC, who have not received prior therapy for advanced or MBC, i.e. patients for whom treatment with endocrine therapy was considered appropriate	Post-menopausal women with HR+ MBC
Outcomes analysed	PFS/TTP, OS TTP has been used where possible and when TTP was not reported, PFS has been used assuming this was similar to TTP. Used Cox results not Log rank results	PFS, OS PFS=TTP if TTP was explicitly defined as the time from randomisation to disease progression or death from any cause (if the reason for death was not reported, it was assumed that the death was from any cause) Where HRs were unavailable, summary statistics were used based on Parmar et al [92]
Included studies	5 studies were included in both the PFS and OS analyses	7 studies were included in PFS analysis, 11 studies in the PFS/TTP analysis and 12 studies in the OS analysis
Synthesis methodology	No direct meta-analysis Series of the adjusted indirect comparison using the methods and principles as described by Bucher et al1996 [75]	For indirect comparisons, analyses were performed using Bayesian network meta-analyses (also known as mixed treatment comparisons), as described by Sutton and Higgins 2008 [77]
Assessment of homogeneity and similarity between included studies	No assessment was reported although the manufacturer stated in the methods that they anticipated systematic differences between studies (i.e. heterogeneity). Thus a random-effects model was used for the calculation of RR. Heterogeneity was intended to be assessed by measuring the degree of inconsistency in the studies' results (I^2). However, neither the I^2 statistic nor measures of relative risk were calculated; HRs were calculated and utilised in the analysis instead	For indirect comparisons, the manufacturer discussed with clinical experts and assessed statistically (from the posterior median variance of the random effects) the suitability of including particular trials in the analyses. A series of sensitivity analysis were performed to explore the nature of heterogeneity
Manufacturers' quality assessment	The manufacturer discussed the limitations of their indirect comparison. These included failure to fulfil basic assumptions of homogeneity, similarity and consistency for the indirect comparison	The manufacturer utilised clinical experts to assess the suitability of trials to be included in their analyses. The manufacturer discussed the limitations of their direct and network meta-analyses and sensitivity analyses were performed
AG comment	The studies included in the indirect comparisons analysis included trials in which the HR+/HER2+ status was unknown; only two trials included analyses of this specific population – EGF30008[51] and TAnDEM[52]	The manufacturer also utilised clinical experts to assess the suitability of trials to be included in their analyses. However there were only three trials in which the HR+/HER2+ status of patients analyses was known– EGF30008[51], TAnDEM[52] and eLEcTRA[54]
OS=overall survival; PFS=progression-free survival; RR=relative risk; TTP=time to progression		

4.2.3 Summary

The findings from the three main trials examining the efficacy of LAP+LET (EGF30008[51]), TRA+ANA (TAnDEM[52]) and TRA+LET (eLEcTRA[54]) all suggest that LAP+AI or TRA+AI result in improved outcomes when compared to AIs (LET or ANA). In the EGF30008[51] and TAnDEM[52] trials, while these differences were not significant for OS, statistically significantly different outcomes were reported for PFS and TTP and large differences were reported between TRA+LET and LET patients in eLEcTRA;[54] this latter trial lacked statistical power to adequately test for significant differences. In addition, both ORR and CBR appeared to be improved for patients taking LAP+AI or TRA+AI although the only statistically significant differences were found for TRA+ANA compared to ANA in TAnDEM.[52] No new safety concerns were identified from the trials although both AEs and SAEs were most common in the LAP+LET and TRA+AI groups than in AIs alone. For LAP+LET, the most significant AE was diarrhoea experienced by around a third of all patients. The impact this may have had on patient QoL is difficult to estimate as only the EGF30008 trial[51] attempted to measure QoL and to date the findings have only been presented as a conference abstract.[61] However, it would appear there are no significant differences between patients in either treatment group. Indeed, the majority of cases of AEs (including diarrhoea) were of grade 1 or 2 severity. Nevertheless, diarrhoea did result in around 1% of all patients who received LAP+LET discontinuing their treatment as a result; all other patients were managed by dose reduction, dose interruption or supportive intervention without treatment dose adjustments. For TRA+ANA patients, the most frequently reported AEs were fatigue, diarrhea and vomiting experienced by around a fifth of all patients, of which the majority were grade 1 or 2 severity. Fatigue was also a problem for around a quarter of patients who received TRA+LET but infections, gastrointestinal disorders and musculoskeletal and connective tissue disorders were even more common; over half of TRA+LET patients experienced these latter two AEs. Around a third of LET patients also reported gastrointestinal disorders and musculoskeletal and connective tissue disorders.

However, extreme caution must be exercised in comparing the aforementioned findings across trials. Thus, for example, it would be wrong to assume, based on the findings just presented, that TRA+ANA is superior to other treatments, for a number of reasons:

- Such a comparison would be considered to be too simplistic and naive as it breaks the randomisation procedure and would not account for differences in baseline characteristics in treatment groups across the trials. To compare the outcomes more accurately a direct comparison or indirect comparisons analyses would need to be considered.
- The ORR hides the fact that none of the patients in the TAnDEM trial[52] examining TRA+ANA were CRs, unlike in EGF30008[51] where 5% of LAP+LET and 4% of LET patients were CRs. It was not known if any patients taking TRA+LET or LET were CRs in eLEcTRA[54] because this trial did not report ORR by CR and PR.
- Trials did not always report data in the same way, so it is unclear, for example, if a greater proportion of patients in the eLEcTRA[54] trial experienced gastrointestinal disorders than in

other trials because such a category of AE did not exist (instead there were data on diarrohea and vomitting, etc). A similar problem was encountered in trying to compare baseline characteristics across trials.

- It is difficult to compare the results of eLEcTRA[54] with the other two trials because this trial was halted prematurely due to slow recruitment.
- Most crucially, it was apparent from the exclusion criteria that there were differences in the patients included in the EGF30008[51] and TAnDEM[52] trials. This appeared to be supported by data reported by the two trials which suggested large differences in median OS in the AI arms (LET and ANA respectively). Notwithstanding the dangers of crude comparisons across trials just highlighted, if patients were similar in terms of their baseline characteristics, given that the evidence to date suggests that there is no difference between LET and ANA in terms of efficacy, albeit in early HR+ breast cancer,[34, 93] then differences in OS of over 5 months would be unexpected as is the case here.

Alternative explanations for differences in OS are that there are real differences between LET and ANA or that differences between the AI groups occurred as a result of differences in second-line treatment received following progression. In relation to the first alternative, there appears to be a broad consensus within clinical practice that there is little to choose between LET and ANA, certainly in terms of efficacy, and there is also evidence of a 'class-effect' (i.e. LET=ANA) albeit from studies of early HR+ breast cancer.[34, 93] In relation to the second possible explanation, it is unclear in EGF30008[51]whether patients received second-line treatment whereas it is stated in TAnDEM[52] that patients did receive second-line treatment once they had progressed, either in the form of TRA+ANA (for ANA patients) or chemotherapy (both treatment groups). Thus, the AG compared median OS between the two AI groups, both in these patients who did not cross-over from the ANA group and ANA patients as a whole and still there were large differences in median OS (of between nearly 5 and 10 months), suggesting real differences in the patient populations of EGF30008[51] and TAnDEM.[52]

The fact that patients were able to cross-over in TAnDEM[52] has added an extra complication in interpreting and comparing the findings, namely, how much of the benefit in OS is attributable to the first-line treatment and how much of the benefit is attributable to subsequent treatment following disease progression? Post-hoc attempts have been made by both Kaufman et al[52] and Roche[39] to address this issue. Kaufman et al[52] compared the median OS between those receiving TRA+ANA with those who initially received ANA but did not cross-over to receive TRA, and those in the AI group who crossed over to receive TRA+ANA with those who did not. Significant differences, when the Wilcoxon test was employed, were reported in favour of TRA+ANA and those in the ANA group who crossed over. The AG believes this was an inappropriate method as it is prone to selection bias. A different method was employed by Roche,[39] namely they employed the RPSFT method which allowed for a comparisons between the TRA+ANA and ANA groups. This reported OS gains to be greater than when no adjustment was made. However, the main justification for employing the RPSFT approach appears to be that it has been used in previous submissions to NICE.[65, 66] The AG notes that other, possibly more appropriate, methods exist and believes that different randomisation-

based methods should ideally be used to compute and compare a range of OS estimates to assess sensitivity of treatment effects. Therefore, the AG believes the findings from the RPSFT approach should be treated with caution.

Because the manufacturers also believed that such a comparison across trials would be too crude and simplistic, both manufacturers conducted adjusted indirect comparisons analyses,[22, 39] with Roche employing a network meta-analyses based on a Bayesian approach[77] in which a number of different analyses and sensitivity analyses were performed and in which it was assumed, and tested, that there was a 'class-effect' for AIs. The findings from both the manufacturers' approaches appeared to support the trial findings suggesting that LAP+LET and TRA+AI were better than AIs alone in terms of PFS and/or TTP, but not OS. In addition, their analyses suggested that there was little difference between LAP+LET, TRA+ANA and TRA+AI. However, the AG believes these indirect comparisons must also be treated with caution for a number of important reasons. First and foremost, as discussed above, the AG does not believe the patient populations are sufficiently similar in the EGF30008[51] and TAnDEM[52] trials. Hence, these trials should not be compared with each other at all. If differences between trials can be explained by differences in second-line treatment subsequently received, then these would be sufficient grounds for not including either trial in an indirect comparisons analysis. On the other hand, if there are differences in efficacy between LET and ANA, then there may be grounds to conduct an indirect comparisons analysis if the other trials are sufficiently similar. However, both indirect comparisons analyses had one other major limitation, recognised by the manufacturers themselves,[22, 39] namely that the basic requirement for indirect comparisons with regard to exchangeability of relative treatment effect between trials in the two MS could not be assumed. Crucially, it was unknown if patients with HR+/HER2+ MBC were included and if so, how many such patients. Both the EGF30008[51] and eLEcTRA trial[54] suggest that the effects of LET in patients with HR+/HER2+ MBC tumours are significantly compromised when compared to those with HR+/HER2- MBC. As has been acknowledged by Roche:[39] "it is clearly plausible that an imbalance in this strong indicator [HER2+] of extremely poor prognosis could have biased the estimates of relative efficacy generated" (MS, pg 18). Other areas of heterogeneity include the proportion of patients with advanced breast cancer, length of follow-up and proportion of patients receiving patients receiving first-line treatment.

Thus, overall, the AG believes comparisons across trials cannot be made and that only the individual findings from each trial should be considered.

5 ASSESSMENT OF COST EFFECTIVENESS

5.1 Systematic review of existing cost-effectiveness evidence

In this section, firstly, a critical appraisal of the available economic evidence describing (i) LAP+LET and (ii) TRA+ANA is described. Secondly, the AG's critique of the two economic evaluations submitted by the manufacturers is presented.

5.1.1 Review of published cost-effectiveness studies

Full details of the search strategy conducted by the AG and the methods used for selecting evidence are presented in section 5. The AG concluded that none of the 107 economics studies identified from the electronic searches were eligible for inclusion in the literature review as they did not include any of the relevant interventions (LAP+AI or TRA+AI). The authors of the GlaxoSmithKline MS noted that "no economic evaluations of lapatinib plus an AI were identified" (MS, pg 100). The authors of the Roche MS stated that, although they summarised the characteristics and results of five studies,[94-98] "four were of poor relevance to the decision problem as they were not in the population of relevance and were not set in the UK" (MS, pg 209). The only study that Roche deemed to be relevant to the review was the poster by Hastings et al,[98] which is discussed below.

The AG notes that the poster presented at ASCO (June 2010) by Hastings et al[98] is relevant to the technologies under assessment. It is noted that the authors of this poster are employees of GlaxoSmithKline yet the poster was only discussed in the MS submitted by Roche.

Summary and critique of Hastings poster

The Hastings[98] poster describes an indirect comparison of the cost effectiveness of LAP+LET vs TRA+ANA in post-menopausal women with HR+/HER2+ MBC who have not received prior treatment. The perspective of the economic analysis is the UK NHS. The evidence network used to estimate treatment effectiveness appears to be the same as that described in the MS submitted by GlaxoSmithKline and includes both EGF30008[51] and TAnDEM;[52] the utility values for PFS and PPS health states are also the same. Base-case results are shown in Table 15 and are different to the estimates provided in the MS by GlaxoSmithKline. Hastings et al[98] conclude that LAP+LET is cheaper and more clinically effective compared to TRA+ANA and is therefore dominant. The AG is of the opinion that the results of the indirect analysis performed by Hastings et al[98] are unreliable as the studies which make up the evidence network are inappropriate (for more details see section 5). In addition, the AG notes that without access to more detailed information on costs, it is difficult to comment on the reliability of the cost-effectiveness results in this study.

Table 15 Base-case results from Hastings poster

Measure	LAP+LET	TRA+ANA	Difference
Total QALYs (discounted)	2.626	2.330	0.296
Total costs (discounted)	£60,614	£64,003	£-3.389
Cost per QALY			Dominant
QALY=quality adjusted life year			

5.1.2 Conclusions of the review of existing cost-effectiveness evidence

There is no relevant, currently available, published cost-effectiveness evidence to describe the use of LAP+LET or TRA+ANA in women who are HR+/HER2+ with MBC.

5.2 Overview and critique of GlaxoSmithKline economic evaluation

5.2.1 Overview of submitted economic evaluation and economic model

The purpose of the manufacturer's model is to assess the cost effectiveness of first-line treatment with LAP+LET in HR+/HER2+ patients with MBC. In the MS, the combination of LAP+LET is compared with the following interventions: LET monotherapy, TRA+ANA and ANA monotherapy. A decision analytic model was developed by the manufacturer to estimate PFS, OS, lifetime costs of treatment of MBC and QALYs. The model schema is presented in Figure 5. The model features three health states: alive and no progression, alive with progression and dead. The manufacturer estimates costs from the perspective of the NHS and Personal Social Services (PSS) and health outcomes in terms of life years, progression-free life years (PFLYs), post-progression life years (PPLY) and quality adjusted life years (QALYs); variables are estimated daily for 10 years. The economic evaluation has a time horizon of 10 years and both cost and benefits are discounted at 3.5% per annum. The manufacturer's reference case adequately reflects the NICE reference case[99] (Table 16).

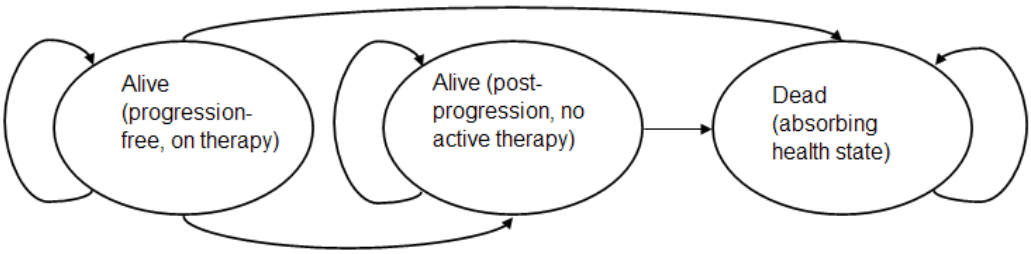

Figure 5 Structure of the model submitted by GlaxoSmithKline

Table 16 Reference case checklist for GlaxoSmithKline economic evaluation

NICE reference case requirements	Reference case[99]	Does the de novo economic evaluation match the reference case?
Defining the decision problem	The scope developed by the Institute	Yes
Comparators	Therapies routinely used in the NHS, including technologies currently regarded as best practice	Yes. Intervention is LAP+LET. Best practice: aromatase inhibitors. New intervention also under appraisal: TRA+ANA
Perspective on costs	NHS and PSS	NHS and PSS
Perspective on outcomes	All health effects on individuals	All health effects on individuals
Type of economic evaluation	Cost-effectiveness analysis	Cost-utility analysis
Synthesis of evidence on outcomes	Based on a systematic review	Based on a systematic review and an adjusted indirect comparison exercise
Measure of health benefits	QALYs	QALYs
Source of data for measurement of QoL	Reported directly by patients and/or carers	(i) Utilities are reported directly by patients from EGF30008 trial for PFS health states (ii) published utilities are used for PPS health states
Source of preference data for valuation of changes in QoL	Representative sample of general public	Algorithm used to map FACT-G values into EQ-5D values; EQ-5D valuations are based on values from representative values of general public
Discount rate	An annual rate of 3.5% on both costs and QALYs	An annual rate of 3.5% on both costs and QALYs
Equity weighting	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit
EQ-5D= EuroQol five dimensions; FACT= Functional Assessment of Cancer Therapy; PSS=Personal Social Services; QALY=quality adjusted life year; QoL=quality of life		

5.2.2 Summary and critique: clinical effectiveness data

Direct clinical evidence: LAP+LET vs LET

The key clinical data (PFS, OS and AEs) used in the manufacturer's economic model comparing LAP+LET vs LET are taken directly from the EGF30008 trial.[51] The AG's description and critique of this EGF30008 trial[51] is presented in section 5 of this report. However, it is worth repeating that (i) HR+/HER2+ patients are a sub-group of the EGF30008 trial[51] and (ii) the trial population does not include patients who have extensive symptomatic visceral or rapidly progressing or life threatening disease.

Progression-free survival and OS estimates for patients receiving LET were estimated by fitting Weibull survival functions to patient-level failure-time data for HR+/HER2+ patients in the EGF30008 trial.[51] Progression-free survival and OS estimates for patients receiving LAP+LET were obtained by applying the HRs for LAP+LET vs LET to the PFS and OS curves for LET for HR+/HER2+ patients in EGF30008 trial.[51] In general, the AG's preferred approach to projective

modelling is to assess PFS and post-progression survival (PPS) separately and then combine them in order to get a more reliable estimate of OS, rather than simply modelling OS as a single entity.

Indirect clinical evidence: LAP+LET vs TRA+ANA; LAP+LET vs ANA

An indirect comparison analysis was carried out by the manufacturer to compare LAP+LET with other drugs. In order to perform the indirect comparison and derive (PFS, OS and AEs) parameter estimates for TRA+ANA and ANA monotherapy, data from TAnDEM[52] and other published AI trials (with or without an anti-HER2 therapy and mixed HER2 populations) were used. A summary and critique of the evidence network employed by the manufacturer is presented in section 5 of this report.

As noted earlier, the AG is not confident that the results of the indirect comparison analysis conducted by the manufacturer are reliable. Firstly, TAnDEM[52] is included in the network and this trial does not exclude patients who have extensive symptomatic visceral or rapidly progressing or life threatening disease which means that the patient populations of the studies in the network are different; in particular, the patients in TAnDEM[52] and EGF30008[51] are not comparable. Secondly, the AG highlights that GlaxoSmithKline uses TAnDEM[52] data published in 2009 that does not use the RPFST method to adjust (for cross-over) OS estimates for patients receiving ANA in TAnDEM.[52] Finally, a further criticism is that in the indirect analysis described in the MS, it is likely that the AI trials include patients who are both HER2+ and HER2-; the AG is of the opinion that inclusion of HER2- patients is inappropriate given that the decision problem is focussed on treating women who are HER2+ and it is becoming apparent[51, 54] that an AI as monotherapy is less effective in women who are HER2+ than in women who are HER2-.

In summary, the AG considers that the submitted model results for the comparison of LAP+LET vs LET is the only comparison that is wholly valid to inform decision making in this area due to the limited comparative clinical effectiveness data available.

5.2.3 Summary and critique: costs and resource use

The manufacturer estimates the following costs for each treatment strategy: acquisition and administration of medications, patient monitoring, treatment of AEs, other costs during PFS and PPS and total costs. The key cost parameters used in the model are summarised in Table 17.

Table 17 Key cost parameters used in the model

Item	Most (£)	Source
Lapatinib (250mg) 70 pack	804.30/pack; 11.49 per tablet	BNF 59[29]
Letrozole (2.5mg) 28 pack	66.50/pack; 2.38 per tablet	BNF 59[29]
Trastuzumab (150mg) vial	407.40	BNF 59[29]
Anastrazole (1mg) 28 pack	68.56/pack; 2.45 per tablet	BNF 59[29]
Dispensing costs	8.50	PSSRU (15mins of Community Pharmacist time, £34/hr)[100]
ECHO/MUGA monitoring costs	46.50/month	National Schedule of Reference Costs 08-09[101] (50%:50%, testing every 3 months)
Total pre-progression cost	562.00/month of PFS	Remak[17]/PSSRU[100]
Total post-progression cost	803.92/month of PPS	Remak[17]/PSSRU[100]
Non-severe adverse event (e.g. chills, constipation, cough, epistaxis, hot flush, nasopharingitis)	99	Probability for hospitalization for G3+ AEs was based on data from EGF30008 trial. Visit and hospitalisation costs for G3+ based on National Schedule of Reference Costs 08-09[101]
Alopecia	158	
Dyspnoea	722	
Headache	255	
Nausea	420	
Vomiting	1398	
BNF= British National Formulary; ECHO= echocardiogram; MUGA= multi gated acquisition scan; PFS=progression-free survival; PSSRU= Personal Social Services Research Unit		

The economic model makes use of pre/post progression cost data from the Remak and Brazil study;[17] however the manufacturer does not comment on the relevance and/or generalisability of the assumptions employed in this study to co-positive patients with MBC in England and Wales.

In summary, the AG notes that, in the MS, the methods used by the manufacturer to identify, measure and value cost items are not fully described. The AG notes that further information is provided in the manufacturer's accompanying cost-effectiveness report. Table 18 identifies key costs which could have been discussed in further detail by the manufacturer.

Table 18 Examples of limited costing methods described in the MS by GlaxoSmithKline

Assumption made	Limitation
Economic evaluation uses a 4mg/kg loading dose of TRA followed by subsequent doses of 6mg/kg in 3-weekly scenario	SPC states that a 8mg/kg loading dose be used then followed by subsequent doses of 6mg/kg in a 3-weekly scenario. In the model a 10mg/kg loading dose is costed (4mg/kg + 6mg/kg)
MS describes costs of PFS and PPS using cost categories described in Remak[17] paper	Remak[17] paper also describes cost categories related to end of life treatment which are not included in the economic evaluation
MS appears to assume that patients did not receive 2nd line chemotherapy or post-progression treatments as further treatments are not discussed	Inappropriate assumption – (i) not a valid assumption for clinical practice in England and Wales as often patients go on to receive additional treatments (ii) high proportion of patients received 2nd line chemotherapy treatment in TAnDEM[52] study which gives first line treatment in a similar setting to patients with MBC
MS assumes a 14 day wastage of oral tablets	GlaxoSmithKline model uses drug costs on a per tablet basis. As drugs can only be bought in packs (and any unused drugs cannot be shared) this is inappropriate. The full (rather than half) pharmacy dispensing cost should also be included in the cost associated with wastage
GlaxoSmithKline estimates drug costs per tablet which leads to inaccuracies: Daily cost per tablet (ANA) = £2.45 Daily cost per tablet (LET) = £2.38 28-day cost (ANA) = £68.60 28-day cost (LET) = £66.64	Pack prices from BNF 59[29]: 28-day cost (ANA) = £68.56 28-day cost (LET) = £66.50
GlaxoSmithKline assumes that delivery of trastuzumab is always an outpatient procedure. Delivery of simple parenteral chemotherapy (SB12Z Deliver parenteral chemotherapy at first attendance [£272] and SB15Z Deliver subsequent elements of a Chemotherapy cycle [£272])	Depending on patient condition and local circumstance, delivery could be on a daycase basis. A weighted average of outpatient and daycase costs would be more meaningful
Monitoring costs in the economic model are £46.50 per month for both LAP+LET and TRA+ANA patients	In the base case, GlaxoSmithKline assumes that cardiac monitoring occurs every 3 months and that both MUGA and ECHO scans are used in equal proportions. AG clinical advisors have stated that MUGA scans are used less frequently (30%) than ECHOs (70%)
Adverse events	No real information is presented in MS to explain methods used to cost concurrent events; all AEs appear to be costed as individual episodes which is unrepresentative of clinical practice
AG=Assessment Group; ECHO= echocardiogram; MUGA= multi gated acquisition scan; PFS=progression-free survival; PPS=post-progression survival; SPC-summary of product characteristics	

5.2.4 Summary and critique: utilities

Utility values for PFS without AEs were estimated using data from EGF30008[51] on the Functional Assessment of Cancer Therapy General Scale (FACT-G)[102] plus Breast Cancer subscale (FACT-B)[22] and an algorithm was used to map from the FACT-G to patient preference-based utilities.[103, 104] The pre-progression utility value used in the model was 0.86.

In the EGF0008 trial[51] FACT assessments were routinely completed by patients only until withdrawal of study medications (i.e. typically at disease progression). This means that post-progression utility values for patients are largely unavailable and the manufacturer states that the generalisability of the values that are available is uncertain. In order to identify a utility decrement for PD to apply to patients with PD, the manufacturer used the results of a study[105] by Lloyd et al of societal preferences

for different stages of MBC in the UK. The absolute reduction in utility compared with no progression used in the model was 0.23; this means that PP utility value can be no higher than 0.62.

Disutility values from grade 3+ AEs were obtained from published and unpublished sources[105-107] and where no data were available, assumptions were made. The utility decrements employed in the economic model include: nausea (0.1); vomiting (0.1); diarrhoea (0.1); alopecia (0.11); asthenia/fatigue/lethargy (0.12); skin and nail disorders (0.15).

The AG notes that the manufacturer does not sufficiently describe the results of the FACT assessments from EGF30008,[51] nor does the manufacturer adequately describe (or test the sensitivity of) the mapping exercise undertaken. Therefore it is difficult to comment on the usefulness of the PFS utility values.

For PPS, the AG agrees with the manufacturer that the paper by Lloyd et al[105] describing UK based societal preferences is relevant to health care decision-making in the UK. However as (i) the health states described in the Lloyd et al[105] paper were derived from literature reviews, exploratory interviews with physicians and an oncology focus group made up of specialist nurses and (ii) the health states were gender neutral and there was no mention of "cancer" in the health state descriptions, the AG is also very aware that health state descriptions and the valuations of the general public may not fully reflect the experiences of patients with cancer nor the true preferences of the general public.

5.2.5 Summary and critique: results

The manufacturer presents detailed summaries of costs and outcomes (PFLY, PPLY, LY, QALYs) for the following regimens: LAP+LET, LET monotherapy, TRA+ANA and ANA monotherapy. Base-case results for the pair-wise comparisons are reported in Table 19. The results show that LAP+LET is not cost effective compared to any of the AIs. LAP+LET appears to be cost effective compared to TRA+ANA; however, as there is much uncertainty about the reliability of the indirect comparison results, these results are not considered by the AG to be meaningful.

Table 19 Base-case results: pair-wise comparisons

					Incremental comparisons: LAP+LET vs		
	LAP+LET	**LET**	**TRA+ANA**	**ANA**	**LET***	**TRA+ANA**	**ANA**
Progression-free life years	1.181	0.738	1.042	0.592	**0.444**	0.139	0.589
Post-progression life years	2.218	2.079	2.004	2.065	**0.138**	0.214	0.153
Life years	3.399	2.817	3.045	2.657	**0.582**	0.354	0.742
QALYs	**2.389**	**1.923**	**2.137**	**1.788**	**0.467**	**0.252**	**0.601**
Acquisition costs	£30,219	£688	£23,818	£576	**£29,531**	£6,401	£29,643
Administration costs	£260	£83	£4,236	£66	**£177**	-£3,976	£194
Monitoring costs	£659	£0	£581	£0	**£659**	£78	£659
Treatment-specific adverse events costs	£113	£71	£109	£67	**£42**	£4	£46
Other progression-free costs	£7,966	£4,975	£7,026	£3,991	**£2,991**	£940	£3,975
Other post-progression costs	£21,396	£20,060	£19,330	£19,919	**£1,336**	£2,066	£1,447
Total costs	£60,614	£25,878	£55,101	£24,620	**£34,737**	£5,513	£35,995
Cost per LYG	-	-	-	-	**£59,684**	£15,590	£48,478
Cost per PFLYG	-	-	-	-	**£78,317**	£39,532	£61,074
Cost per QALY gained	-	-	-	-	**£74,448**	£21,836	£59,895

LYG=life year gained; PFLYG=progression-free life year gained; QALY= quality adjusted life year

*AG only considers the results of the LAP+LET vs LET comparison to be valid [highlighted in bold]

5.2.6 Summary and critique: sensitivity analysis and probabilistic sensitivity analysis

The AG notes that 51 scenarios were examined by the manufacturer using sensitivity analysis. The results of the sensitivity analysis are presented in Table 41 of the MS. For the comparison with LET monotherapy, the incremental cost per QALY is in the range of £41,877 to LAP+LET being dominated by LET monotherapy. The cost per QALY gained versus ANA monotherapy ranges from £38,170 to £378,674. For the comparison with TRA+ANA, the range is LAP+LET dominating the comparator to a cost per QALY estimate of £45,106. The approach to sensitivity analysis adopted by the manufacturer makes it is difficult for the AG to acquire any real insight the true drivers affecting the size of the ICERs.

A summary of the PSA is presented in Figure 6 and the cost-effectiveness acceptability curve (CEAC) is shown in Figure 7. The CEAC shows that at a cost-effectiveness threshold of £30,000 per QALY gained, the probability of LAP+LET being cost effective is very low (<25%) compared with any AI and low (approximately 50%) compared with TRA+ANA.

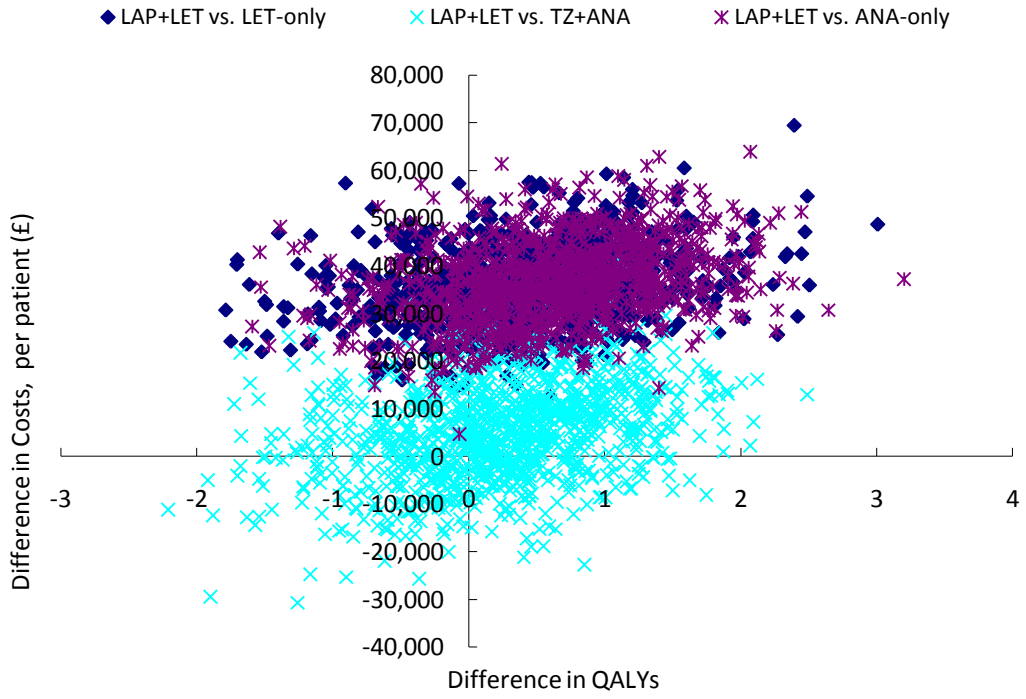

Figure 6 Incremental cost-effectiveness ratios for LAP+LET versus comparators

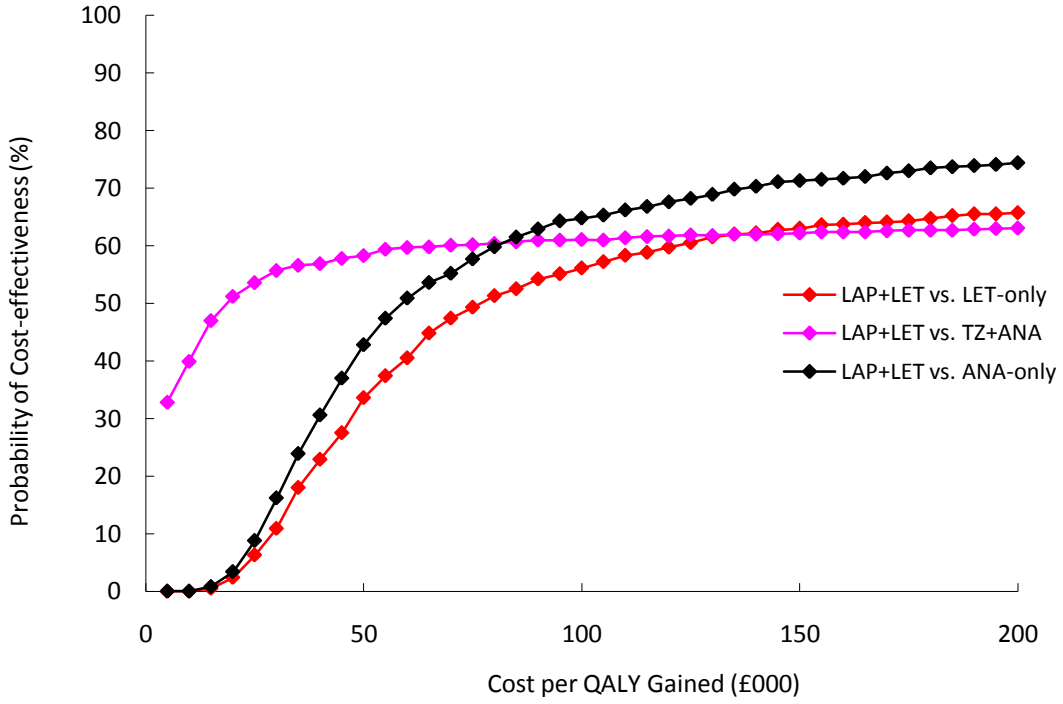

Figure 7 CEAC for LAP+LET vs LET, TRA+ANA and ANA

5.2.7 Summary and critique: End of life treatment criteria

The AG notes that the manufacturer has not requested that LAP+LET be considered by NICE as an end of life treatment.

5.2.8 Conclusions of the Assessment Group

From the information presented in the MS, the AG agrees with the manufacturer that LAP+LET is not cost effective when compared with LET.

The AG also considers that the methods used in the indirect analysis undertaken by the manufacturer are unreliable and concludes that the ICERs derived from the remaining comparisons (LAP+LET vs TRA+ANA; LAP+LET vs ANA) are not meaningful.

GlaxoSmithKline did not make a case for LAP to be considered as an end of life treatment.

5.3 Overview and critique of Roche economic evaluation

5.3.1 Overview of submitted economic evaluation and economic model

The purpose of the manufacturer's model is to assess the cost effectiveness of first-line treatment with TRA+ANA in HR+/HER2+ patients with MBC. In the MS, the combination of TRA+ANA is compared with the following interventions: ANA monotherapy, LAP+LET and LET monotherapy. An AUC model was designed to calculate the present value of the health outcomes and NHS/PSS costs attributable to each possible treatment option calculated. The model schema is presented in Figure 8. The model features three health states (PFS, PD and death) and has a cycle length of one month. The manufacturer estimates costs from the perspective of the NHS/PSS and health outcomes in terms of LYG and QALYs. The economic evaluation has a time horizon of 15 years and both costs and benefits are discounted at 3.5% per annum (implemented monthly). The manufacturer's economic evaluation adequately reflects the NICE reference case[99] (Table 20).

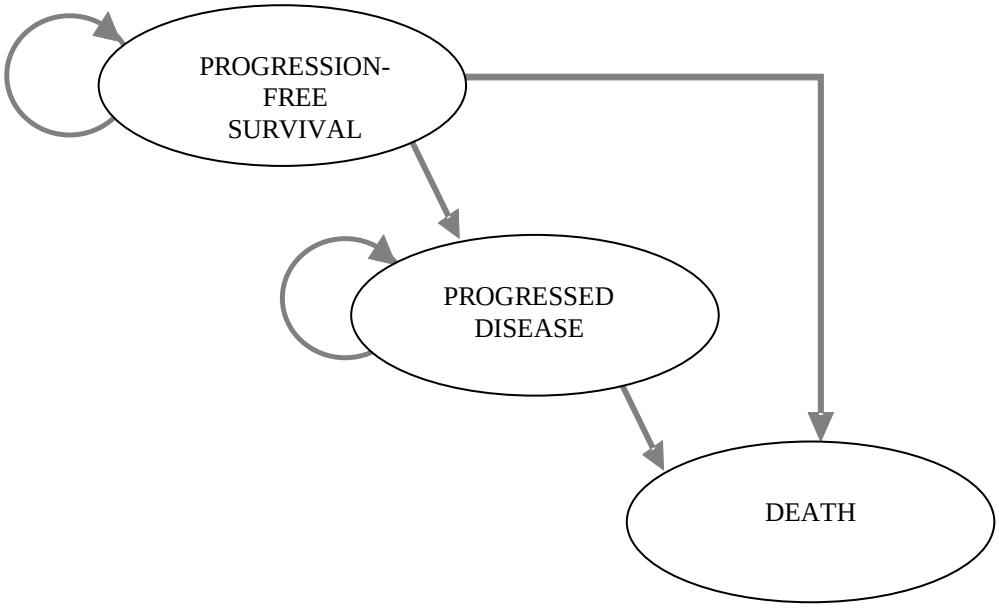

Figure 8 Structure of the model submitted by Roche

Table 20 Reference case checklist for Roche economic evaluation

NICE reference case requirements	Reference case[99]	Does the de novo economic evaluation match the reference case?
Defining the decision problem	The scope developed by the Institute	Yes
Comparators	Therapies routinely used in the NHS, including technologies currently regarded as best practice	Yes. Best practice: aromatase inhibitors. New intervention also under appraisal: LAP+LET
Perspective on costs	NHS and PSS	NHS and PSS
Perspective on outcomes	All health effects on individuals	All health effects on individuals
Type of economic evaluation	Cost-effectiveness analysis	Cost-utility analysis
Synthesis of evidence on outcomes	Based on a systematic review	Based on a systematic review and indirect analysis exercise
Measure of health benefits	QALYs	QALYs
Source of data for measurement of QoL	Reported directly by patients and/or carers	Health state descriptions derived from reviews of the literature and lay and professional focus groups
Source of preference data for valuation of changes in QoL	Representative sample of general public	Representative sample of general population
Discount rate	An annual rate of 3.5% on both costs and QALYs	An annual rate of 3.5% on both costs and QALYs (implemented monthly)
Equity weighting	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit
PSS= Personal Social Services; QALY= quality adjusted life year; QoL=quality of life		

5.3.2 Summary and critique: clinical effectiveness data

Direct clinical evidence

The key clinical data (PFS and OS) used in the MS are taken directly from the TAnDEM[52] trial and utilise some previously unpublished data as the model inputs were taken from an April 2008 data cut (the published paper from the trial uses an older data cut). As the PFS curves from TAnDEM[52] were practically complete, the Kaplan-Meier PFS curves for the two regimens were used directly to model the majority of disease progression of patients within the economic model (uncertainty in the Kaplan-Meier PFS curves was addressed in the sensitivity analysis). In general, the AG's preferred approach to projective modelling would be to assess PFS and PPS separately and then combine them in order to get a more reliable estimate of OS, rather than simply modelling OS as a single entity.

In the TAnDEM[52] trial estimates of OS were affected by (i) high rates of cross-over of patients from ANA to TRA+ANA and (ii) second-line chemotherapy imbalance by trial group. Both of these factors have been considered by the AG in section 5. The AG is aware that the OS estimates for TRA+ANA used in the base case are the adjusted values; use of the RPSFT approach reduced the OS HR of TRA+ANA vs ANA. To date, the manufacturer has been unable to account for the second-line therapy imbalance using quantitative methods. As the Kaplan-Meier OS curves were not complete, parametric fitting of the curves was carried out to allow extrapolation beyond the follow-up period.

The manufacturer concluded that the exponential distribution most accurately portrayed the OS curves of the two regimens for the time period beyond the availability of the Kaplan-Meier OS curves.

The AG considers that the manufacturer's approach to adjusting for cross-over in the TAnDEM[52] trial is limited and requires further exploration/justification before confidence can be placed in the OS results generated.

Indirect clinical evidence

As the manufacturer's systematic review found no network capable of linking all of the regimens in the population of interest, the manufacturer assumed that LET and ANA hold an 'AI class effect' in terms of PFS and OS and that the PFS and OS curves observed for ANA patients in TAnDEM[52] would therefore hold for LET patients. In order to integrate the combination therapies into the evidence network the manufacturer also assumed that HER2 status is independent of the relative treatment effect of the AI therapies. The AG's critique of the manufacturer's evidence approach to indirect analysis is fully discussed in section 5.

In summary, the AG is not confident that the results of the indirect comparisons analyses conducted by the manufacturer are reliable. Firstly, the EGF30008[51] trial is included in the network and this trial excludes patients who have extensive symptomatic visceral or rapidly progressing or life threatening disease which means that the patient populations in EGF30008[51] and TAnDEM[52] are not comparable. Secondly, the AG agrees with the manufacturer that "given...mixing of heterogeneous populations, the results produced by the [indirect] analysis should be treated with caution" (MS, pg 216). Specifically, the AG is of the opinion that inclusion of evidence based derived from HER2- patients is inappropriate given that the decision problem is focussed on treating women who are HER2+ and it is becoming apparent[51, 54] that an AI as monotherapy is less effective in women who are HER2+ than in women who are HER2-.

In summary, the AG considers that the submitted model results for the comparison of TRA+ANA vs ANA is the only comparison that is wholly valid to inform decision making in this area.

5.3.3 Summary and critique: resource use and costs

The manufacturer presents a detailed and comprehensive description of resource use and costs used in the economic model. The cost categories are presented as follows: monthly drug costs, treatment duration, administration and monitoring costs, pharmacy preparation, response assessment, cardiac monitoring, PFS background supportive care, AEs, progressed disease costs and 'end of life' costs. The key cost parameters used in the economic model are summarised in Table 21. In terms of costs, the key difference between the treatment regimens is the acquisition costs of the drugs (LAP and TRA are much more expensive than the AIs) followed by administration costs (IV TRA is much more

expensive to administer than the other oral drugs) and finally, the costs of 'PD, BSC and second-line treatment'(post PFS costs are higher for the patients taking LAP and TRA).

Table 21 Key parameters used by Roche

Item	Cost (£)	Source
Lapatinib* (250mg)	11.49 per tablet purchased; 12.32 per tablet taken; 2249.53/month	BNF 59[29]
Letrozole* (2.5mg)	2.38 per tablet purchased; 2.55 per tablet taken; 77.50/month	BNF 59[29]
Trastuzumab** (150mg vial)	1956.33/month (vial sharing used in model); 2230.16/month (full wastage)	BNF 59[29]
Anastrazole* (1mg)	2.45 per tablet purchased; 2.62 per tablet taken; 79.90/month	BNF 59[29]
ECHO/MUGA monitoring costs	46.31 /month	Ward et al 2006;[108]NHS Reference Costs 08-09[101] (70%:30%, testing every 4 months)
Total subsequent monthly cost (administration, cardiac monitoring, pharmacy prep)	92.99 (ANA/LET) 273.58 (LAP+LET) 297.87 (TRA+ANA)	NHS Reference Costs 08-09;[101] PSSRU 2009[109] and Clinician Advisory Board
Progressive disease costs (2nd line treatment with exemestane monotherapy)	92.88	Roche Advisory Board; BNF 59[29]
PFS BSC costs	192.83/month	NICE CG81; [24] PSSRU 2009[109]
Post-progression BSC costs	542/month	NICE CG81; [24] PSSRU 2009[109]
End of life costs	3,418/last 14 days of life	NICE CG81[24]
Examples of AE costs:		
Back pain	194	NHS Reference Costs 08-09[101] (PS05A)
Cardiac failure	370	NHS Reference Costs 08-09[101] (EB05Z)
Chest pain	400	NHS Reference Costs 08-09[101] (PA22Z)
Hypertension	560	NHS Reference Costs 08-09[101] (EB041)
Vomiting	553	NHS Reference Costs 08-09[101] (PA28B)
*6.72% of dispensed oral tablets are wasted (based on TAnDEM[52] data) **80% of patients receiving trastuzumab do so in vial sharing centres		
Utility values		
Cooper[110] /NICE[24] (base case)	PFS=0.73; PD=0.45; stable disease=0.65	
Hastings et al[98] (sensitivity analysis)	PFS=0.86; PD=0.62	
BNF= British National Formulary; PSSRU= Personal Social Services Research Unit		

The AG notes that the economic evaluation in the MS uses a three-weekly schedule for TRA; although this is not the weekly schedule used in TAnDEM,[52] the AG agrees that the three-weekly schedule is typically used in clinical practice in England and Wales.

Table 22 summarises the key costs that the AG believes the manufacturer could have considered in more detail in the MS.

Table 22 Examples of limited costing methods employed in MS by Roche

Assumption made	Limitation
Wastage (tablets)	Roche uses an average pill count per month to compare with a notional number of whole packs - this is not correct and underestimates wastage
Subsequent administration of trastuzumab	Roche uses estimate from old interim local source instead of correct NHS reference cost
Roche estimates drug costs per tablet purchased/taken which leads to inaccuracies	Only pack prices are available from the BNF 59[29] and drug usage would be more accurately costed accordingly
Roche assumes that delivery of trastuzumab is always a daycase procedure and uses daycase costs	Depending on patient condition and local circumstance, delivery could be on an outpatient basis. A weighted average of outpatient and daycase costs would be more meaningful
Cardiac monitoring costs used in the economic model	In the base case, Roche assumes that cardiac monitoring occurs every 4 months and that MUGA and ECHO scans are used in unequal proportions (30%:70%). AG clinical advisors have stated that cardiac monitoring occurs every 3 months
MUGA cost=£316.64 based on uplifted cost from Ward et al 2006	NHS Reference cost (08-09[101]) RA37Z OP=£203.05 is more appropriate
Adverse events are limited and poorly described and sourced. AE costs described by Roche appear to be underestimates	Examples: Back pain uses “paramedic attendance” cost and ignores IP/OP/DC episode costs. In MS hypercalcaemia cost is given hypertension cost (model); hypercalcaemia cost is not estimated in model
AG=Assessment Group; BNF= British National Formulary; DC=day case; ECHO= echocardiogram; IP=inpatient; MUGA= multi gated acquisition scan; OP=outpatient;PFS=progression-free survival; PSSRU= Personal Social Services Research Unit	

5.3.4 Summary and critique: utilities and adverse events

The TAnDEM[52] trial conducted by Roche did not collect data using a generic health-utility instrument. In order to estimate utility values for co-positive patients, the manufacturer underook a focussed review of the literature, identified 20 relevant studies (1996 to 2009) and presented utility values from six of these published studies (MS, pg 247). However, as no studies were identified as being relevant specifically to co-positive patients, the manufacturer made a decison to use the assumptions of Winstanley and Murray in the recent breast cancer publication[24] and to apply utilities as identified by Cooper et al;[110] this decision was made to ensure alignment with the most recent relevant piece of research[24] commissioned by NICE in breast cancer. In the sensitivity analysis, the manuacturer made use of the utility values cited by Hastings et al[98] in the indirect comparison of LAP+LET vs TRA+ANA. The values used by Hastings et al[98] were derived via mapping FACT-G[102] data collected in EGF30008[51] to EQ-5D. Both sets are values are shown in Table 21. Disutility from AEs is not a feature of the economic model developed by Roche.

The Cooper al[110] paper pools utilities from many different sources (all derived from oncology nurses using the standard gamble technique). In contrast, the AG notes that the paper by Lloyd et al,[105] identified by the manufacturer, asks 100 members of the general public to rank health states using the standard gamble technique to determine utility values. As the study by Lloyd et al[105] is a large preference study designed to obtain UK-based societal preferences for distinct stages of MBC, the AG considers the paper by Lloyd et al,[105] with caveats previously mentioned, to be the most useful evidence available that could help to inform the decision problem.

In the Roche model, only grade 3 or grade 4 AEs are considered. In the MS it is assumed that the AEs recorded for TRA+ANA are the same for LAP+LET and that the AEs recorded for ANA can be applied to LET. For comparison of TRA and LAP this seems unlikely as episodes of diarrhoea are reported more often for LAP patients. The AG also notes that the costs of several of the AEs listed in the MS (e.g. anaemia, cardiac failure, hypercalcaemia, hypertension) are not the cost inputs used in the economic model (‘AE cost data’). In summary, the AG is of the opinion that the AE costs used in the economic model are underestimated and require revision to make them reliable.

5.3.5 Results: summary and critique

The manufacturer presents detailed summaries of estimated costs and outcomes (time in PFS, time in PD, life years, QALYs in PFS, QALYs in PD and total QALYs) for the following regimens: TRA+ANA, LAP+LET, ANA monotherapy and LET monotherapy. As there are four regimens of interest, the manufacturer has chosen to represent the results of the economic evaluation in terms of the efficiency frontier. The efficiency frontier links the regimens that are not dominated. Figure 9 shows that, using this approach, LAP+LET does not lie on the efficiency frontier and that the key comparison is between TRA+ANA vs LET. When TRA+ANA is compared with LET, the ICER is estimated at approximately £54,336 per QALY gained. When TRA+ANA is compared with LAP+LET, it appears to be cost effective.

Table 23 Incremental cost effectiveness ratios (cost per QALY gained)

	TRA+ANA	LAP+LET	ANA	LET
Total costs	£54,748.92	£51,882.53	£23,340.88	£23,327.52
Total QALYs	1.87	1.71	1.29	1.29
ICER: TRA+ANA vs...				
LAP+LET	£17,914/QALY gained			
ANA	£54,151/QALY gained			
LET	£54,174/QALY gained			
ICER=incremental cost-effectiveness ratio				

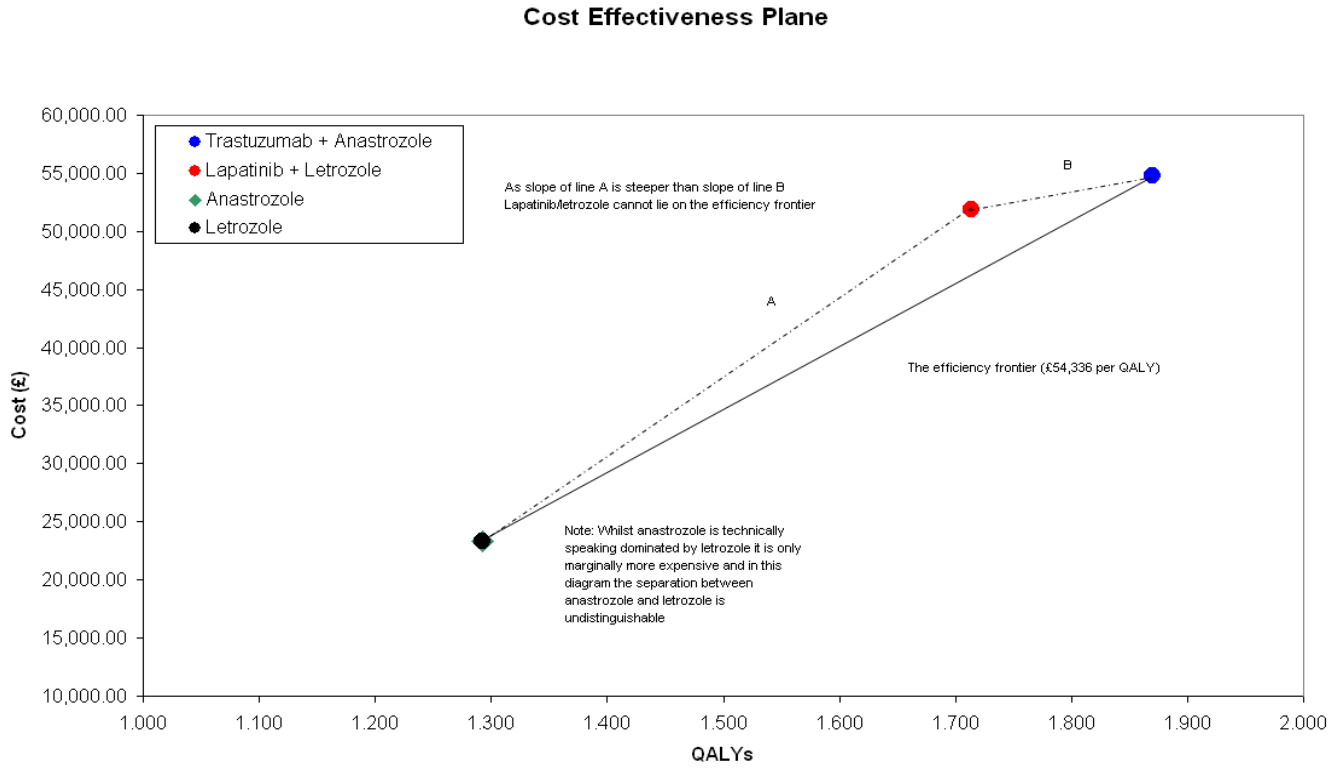

Figure 9 Cost-effectiveness plane

5.3.6 Summary and critique: sensitivity analysis and probabilistic sensitivity analysis

Sensitivity analysis

Twenty-three different parameters were modified in the univariate sensitivity analysis. The results are presented only for the comparison of TRA+ANA vs LET. The base case ICER was most sensitive to variation in PFS utility values (£50,099 to £59,355) and the rate used to discount health outcomes (£48,664 to £58,400).

The manufacturer also described three multivariate/scenario analyses. The implementation of the indirect comparisons analysis PFS and OS HRs into the model (using TAnDEM[52] ANA curves as a baseline with showing that LET is slightly preferred to ANA) leads to a change in the efficiency frontier and the results are as follows: ANA represents a cost-effective option up to a threshold of £3,594; LET is the most cost-effective option from £3,594 to £57,773; above £57,773 TRA+ANA represents the most cost-effective treatment.

When pessimistic (PFS=0.65; PD=0.35) and optimistic (PFS=0.8; PD=0.55) utility values are used in the multivariate analysis, the ICER ranges from £48,715 to £62,239. The manufacturer estimates that the base case ICER would fall to £44,497 if the utility values used in the Hastings[98] paper are employed (these values do not fall within the +/-10% of PFS and PD values used in the univariate sensitivity analysis).

Finally, the manufacturer attempts to account for the confounding influence of the imbalance in second-line chemotherapy in the TAnDEM[52] trial and demonstrates that the base case ICER could fall to around £49,426.

Probabilistic sensitivity analysis

The PSA carried out by the manufacturer is summarised in the MS (pg 278). The CEAC (MS, pg 279) shows that at a threshold of £30,000 per QALY gained, the combination therapies are **never** cost effective. At a threshold of £55,000 per QALY gained, TRA+ANA was shown to be cost effective in approximately 35% of simulations (i.e. had a low probability of cost effectiveness).

5.3.7 End of life treatment criteria: summary and critique

This section provides an overview and critique of the manufacturer's case for TRA+ANA as an end of life maintenance treatment for patients MBC. The NICE end of life treatment criteria has three key points:

- The treatment is indicated for patients with a short life expectancy, normally less than 24 months and;
- There is sufficient evidence to indicate that the treatment offers an extension to life, normally of at least an additional three months, compared to current NHS treatment, and;
- The treatment is licensed or otherwise indicated for small patient populations.

Patient life expectancy of less than 24 months

The published literature[111]on prognosis after a diagnosis of MBC confirms that the disease is incurable and patient life expectancy is short. In a previous scope[112] issued by NICE (Lapatinib for the treatment of previously treated women with advanced, metastatic or recurrent breast cancer) it was stated that "...The average life expectancy after diagnosis of metastatic breast cancer is 18-24 months. This is reduced by up to 50% for patients with tumours over-expressing HER2."

The manufacturer cites data from the comparator (ANA) group of the TAnDEM[52] trial to support the argument that patients with MBC who are co-positive have a very poor prognosis. Median OS is shown to range between 17.2 months (excluding all patients who crossed over) and be no greater than 32.1 months (excluding all patients with liver metastases); other OS estimates are also generated depending on the methodology used to undertake survival analysis (Table 24). The AG notes that although data from the EGF30008[51] trial show that patients with MBC who are co-positive have a median OS of 33.3 months, this trial excludes patients who have extensive symptomatic visceral or rapidly progressing or life threatening disease. The AG acknowledges that use of second-line therapies may also influence estimates of OS.

Table 24 Overall survival (AI monotherapy)

Trial	Median OS: ANA Unadjusted	Median OS: ANA Centrally confirmed status	Median OS: ANA RPFST adjusted	Median OS: ANA Excluding all patients who crossed over	Median OS: LET
TAnDEM[52]	23.9 months	28.6 months	21.98 months	17.2 months	n/a
EGF30008[51]	n/a	n/a	n/a	n/a	33.3 months
n/a=not applicable					

Life extension of at least three months

The manufacturer attempts to demonstrate that there is sufficient evidence from the TAnDEM[52] trial to indicate that TRA+ANA offers an extension to life of at least an additional three months, compared to current NHS treatment. In TAnDEM[52] unadjusted median OS gained and RPFST adjusted median OS gained were estimated to be 4.6 and 6.54 months respectively.

The AG is of the opinion that TRA+ANA compared to ANA yields a life extension of at least 3 months for patients who are co-positive and who have had no prior treatment for MBC.

Licensed for a small patient population

The size of the patient population eligible for treatment with TRA+AI in England and Wales i.e. women with HR+/HER2+ MBC is estimated to be around 50 patients by both GlaxoSmithKline and Roche. However, trastuzumab has indications in MBC, metastatic gastric cancer and early breast cancer. The manufacturer reports that, in England and Wales, across all the indications 7,158 patients are eligible to receive TRA each year (2,333 from MBC, 506 from metastatic gastric cancer and 4,319 from early breast cancer). It was difficult for the AG to verify these population figures as the references cited were not included in the references package as part of the MS; the data were from pharmaceutical company reports which the AG could not access easily. The AG notes that there is currently an ongoing STA[113] of TRA for the treatment of HER2+ metastatic adenocarcinoma of the stomach or gastro-oesophageal junction (second Appraisal Committee meeting was held in August 2010). During this STA[113] the Appraisal Committee considered the size of the eligible patient population and was not satisfied that the population for which TRA is licensed met the criterion of a small patient population. The population in the ongoing STA[113] was estimated to be 7,144 people who have HER2+metastatic gastric cancer, HER2+early and locally advanced breast cancer or HER2+metastatic breast cancer, i.e. very similar to the size of population estimated in this STA.

5.3.8 Conclusions of the Assessment Group

From the information presented in the MS, the AG agrees with the manufacturer that TRA+ANA is not cost effective when compared with ANA. The AG also considers that the methods used in the indirect analysis undertaken by the manufacturer are unreliable and concludes that the ICERs derived from the remaining comparisons (TRA+ANA vs LET; TRA+ANA vs LAP+LET) are not meaningful.

The manufacturer submitted a case for TRA+ANA to be considered as an end of life treatment for women with HR+/HER2+ MBC. The AG agrees that TRA+ANA meets the criteria as a treatment for patients with a short life expectancy and that it extends life by an additional 3 months when compared

to current NHS treatment. However, the AG makes no comment on whether or not the criterion of a small patient population is met.

5.4 Independent economic assessment

Each of the novel treatment regimens considered in this AG report rely upon clinical evidence derived from a single small RCT (EGF30008[51] or TAnDEM[52]). Moreover the comparator treatments differ between these trials, albeit both drugs were drawn from the same class of compounds. These disparities alone suggest the need for caution when generalising these results of the RCTs. However, an even greater difficulty arises if the two study populations do not appear to match.

As discussed in earlier sections, there is reason to believe that in some important respects the protocol criteria governing the selection of subjects for these two trials were sufficiently dissimilar as to be likely to generate non-equivalent patient populations. In particular, the requirement in the EGF30008[51] trial to exclude "Subjects with extensive symptomatic visceral disease including hepatic involvement and pulmonary lymphangitic spread of tumor, or the disease was considered by the investigator to be rapidly progressing or life threatening" is not matched by a similar exclusion in the TAnDEM[52] trial protocol.

As a result, it is reasonable to expect that patients in EGF30008[51] may have been somewhat fitter and with better prognoses than those recruited into TAnDEM.[52] However, direct comparison of patient characteristics in the trials is restricted by differences in how measures were defined and/or reported in the two CSRs. For example, ****. Tables setting out the number and location of metastatic lesions at baseline in the two trials are available. However they cannot be compared with full confidence since they are defined somewhat differently. There is strong evidence of a significant difference in the mean age of the populations (**** years in EGF30008,[51] and **** years in TAnDEM,[52] $p<0.0001$). There is also evidence of a greater incidence of soft tissue metastases in TAnDEM[52] patients (43.5% vs 30.14%, $p=0.004$). Metastases at other sites are broadly comparable, with the exception of bone metastases (14.2% in EGF30008[51] vs 56.5% in TAnDEM[52]) though this is likely to be an artefact of differing reporting methods. Overall the frequency of metastatic sites affected per patient (1.77 in EGF30008[51] vs 2.40 in TAnDEM[52]) also suggests more severe advanced disease in TAnDEM[52] patients, but this too could be a reporting effect.

Coupled with the serious problems identified in earlier sections relating to the indirect comparison of treatment effects, these uncertainties led the AG to conclude that whilst two separate assessments of cost effectiveness, each based on one of the principal RCTs, could be undertaken with some confidence, the evidence base is too insecure to allow a meaningful comparison of the two innovative compounds against each other. In this section, two separate cost-effectiveness analyses are reported,

using a common modelling framework and common parameter values, but employing effectiveness data drawn only from a single RCT (either TAnDEM[52] or EGF30008[51]).

5.4.1 Methods: common model features and parameters

Model design

A common model structure has been adopted for both *de novo* cost-effectiveness analyses (Figure 10), and wherever possible has been implemented using the same parameter values. The *de novo* model employs outcome data derived from the relevant clinical trial in the form of Kaplan-Meier estimated survival values augmented by projected survival estimates calibrated against the observed data. The preferred approach uses PFS and PPS estimates directly as the basis for calculating expected OS in each group of the RCT. Progression-free survival and PPS values then furnish the information required to calculate all components of health service costs, and also to estimate the expected future patient utility. These survival estimates are calculated separately for each date on which a resource is expected to be used (e.g. when prescriptions are dispensed, or when a hospital visit or test takes place) avoiding the need for a general model cycle or for mid-cycle corrections.

Both of the manufacturers' models use PFS and OS as the primary sources for survival information, and derive time in PPS as the difference between OS and PFS. The AG finds this approach generally liable to generate substantial bias in OS estimates when projective parametric modelling is used. This is because recorded OS data are a result of combining patient experience in two distinct phases in which hazard rates would be expected to exhibit quite different dynamics (in PFS the patient is likely to have reduced event risks whilst the active drug continues to be effective, but in PPS event risks are more likely to revert to higher levels of uncontrolled disease progression). In most cases standard parametric statistical models cannot accurately represent an outcome measure (such as OS) which is a compound of two very different processes, and modelled OS projected over several decades can result in very large cumulative errors. By contrast, in advanced disease the risk profile of patients entering PPS is usually quite stable and allows projective modelling with greater confidence (i.e. narrow confidence intervals), and limits the risk of some of the more extreme estimates of long-term survival which can occur when modelling OS directly. At a pragmatic level, deriving PPS as the difference between OS and PFS can sometimes lead to modelling anomalies with negative estimates of PPS during projection, an error which cannot occur when PFS and PPS estimates are summed.

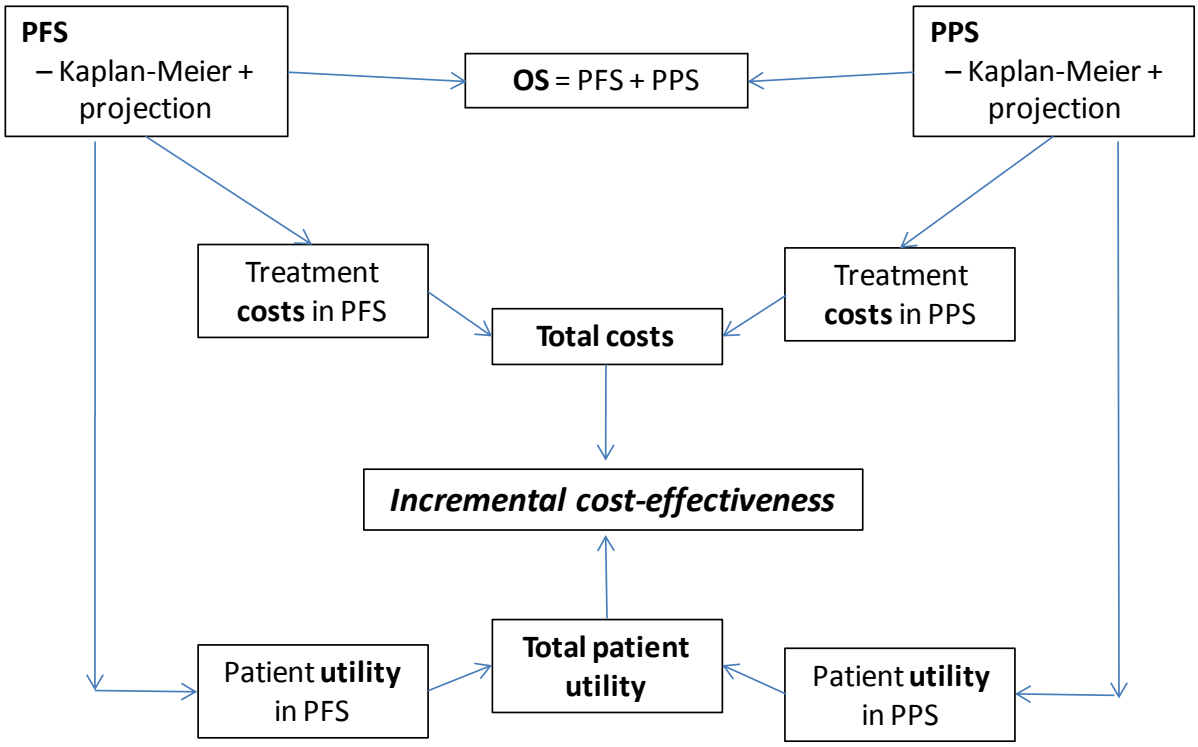

Figure 10 Schematic of AG model structure for each treatment option

Undiscounted and discounted (3.5% p.a. for costs and outcomes) deterministic results were generated for each year of remaining life up to 30 years. A PSA was carried out for all model variables for which sampling uncertainty could be estimated, and a range of univariate sensitivity analyses were performed for other variables and assumptions.

Cost parameter values

Model costing variables common to both models are listed in Table 25 with the parameter values used in the base case analyses and the data sources employed.

Table 25 Costing parameter values and sources common to both AG economic models

Cost item	Value	Source
Pharmacy dispensing costs	£9.00 from hospital pharmacy £6.90 from community pharmacy	Roche MS, based on hourly cost of pharmacist time[109] and 12 minutes per script dispensed
Frequency of cardiac monitoring	Every 3 months	Clinical advisor opinion
ECHO: MUGA proportion of cardiac scans	70: 30	Clinical advisor opinion
Unit cost per ECHO scan	£74.37	NHS Reference Costs 2008-09;[101]Direct access diagnostics code DA02
Unit cost per MUGA scan	£203.05	NHS Reference Costs 2008-09[101] OP Nuclear medicine code RA37Z
Frequency of OP follow-up (including CT scan) in PFS	Every 3 months	Clinical advisor opinion
Unit cost per OP follow-up visit	£98.51	NHS Reference Costs 2008-09:[101] Consultant led follow-up attendance, non-admitted, face to face, code 800 (clinical oncology)
Unit cost per CT scan	£138.27	NHS Reference Costs 2008-09:[101]code RA12Z - OP CT Scan (2 areas with contrast)
Annual cost of Best Supportive Care in PFS	£1,831.54	NICE Guideline[24]updated for inflation[109]
Annual cost of Best Supportive Care in PPS	£5,597.82	NICE Guideline[24]updated for inflation[109]
Terminal care costs (last 2 weeks of life)	£1,788.55	Remak & Brazil paper[17] updated for inflation[109]
Unit cost of exemestane (per 30 tablet pack)	£88.80	BNF 59, March 2010[29]
Wastage per patient (half pack)	£44.40	BNF 59, March 2010 [29]
Proportion of PPS patients receiving exemestane	50%	Modelling assumption
Discounting rate (costs)	3.5% p.a.	NICE Methods Guide[99]
Discounting rate (outcomes)	3.5% p.a.	NICE Methods Guide[99]
BNF= British National Formulary ; CT=computed tomography; ECHO= echocardiogram; MUGA= multi gated acquisition scan OP= outpatient; PFS=progression-free survival; PPS=post-progression survival		

Patient utility valuation

The AG considered a number of sources for utility values for breast cancer patients referenced by the manufacturers and conducted an exploratory (but not systematic) search which failed to identify any useful additional material. Overall there appears to be a particular dearth of relevant utility studies appropriate to this appraisal and conforming to the NICE recommended[99] approach (UK data capturing population preferences using EQ-5D or either time-trade off or standard gamble methodology). Several standard gamble studies have been reported, but normally use a very small sample of health professionals to assess quite general health states, not particularly focused on advanced disease. The AG concluded that the best available option was the study reported by Lloyd et al[105] in 2006, which considered health states and a limited set of treatment-related AEs specific to MBC and developed a mixed model using data collected from a sample of 100 UK residents broadly similar in age and sex to the general population. The values presented have face validity for both

absolute values and inter-state differences. The particular benefit is that they furnish an integrated system of utility estimates (rather than adopting values from disparate sources as is often the case).

The Lloyd et al[105] model includes age and treatment response as model variables. However, it is important to note that the relevant age is not that of the patient, but of the study participant. To ensure consistency with the UK EQ-5D standard value scheme, the AG adopted the average age of respondents to the original multi-vessel disease study[114] of just over 47 years in arriving at utility parameter values. For patients who are pre-progression it was necessary to calculate a weighted average of the model values for stable disease and treatment response, based on the reported response rate (CR+PR) in each group of each trial; those are shown below in sections 5.4.2 and 5.4.3. A common health state utility value was obtained for post-progression patients of 0.496 (se 0.160) for use in both models.

5.4.2 Specific model features and parameters: LAP+LET vs LET

Expected PFS

The manufacturer of LAP provided full details of survival analyses (PFS, PPS and OS) requested by the ERG relating to data from the EGF30008 trial.

Confidential information withdrawn

Examination of the Kaplan-Meier estimates of PFS over time (**Error! Reference source not found.**) shows an early advantage for the combination therapy (LAP+LET) compared to LET alone, but also indicates that this benefit steadily eroded over time until the two treatments were indistinguishable beyond about 16 months. The AG decided that the most reliable estimate of the mean expected PFS would be obtained by using the difference between the Kaplan-Meier AUC estimates up to the time of convergence (505 days). Thereafter a single exponential model of PFS was applied to both the intervention and the comparator in the AG model, calibrated on pooled Kaplan-Meier data for the period >500 days.

This approach yielded estimates of PFS up to 505 days:

198.5 (se 17.6) PFS days for LET only

266.2 (se 16.1) PFS days for LAP+LET

i.e. a **gain of 67.6** (se 16.9) PFS days attributable to use of LAP.

Expected PPS

Examination of the Kaplan-Meier analysis of PPS in the EGF30008[51] trial indicated that following disease progression patients in both groups of the trial were at the same risk of death which appears to be constant over time. Therefore, a single exponential model was calibrated from the pooled trial data for use in the AG model (**Error! Reference source not found.**) yielding an estimated mean survival for patients in PPS of 764.8 (se 5.0) days.

Confidential information withdrawn

Expected overall survival

Since no cross-over following disease progression was permitted in the EGF30008[51] trial, and there is no evidence of significant imbalance in post-progression therapies no further adjustments to PFS and PPS estimates are necessary.

The best estimate of OS is obtained by summing PFS and PPS, after adjusting PPS to exclude patients dying at or before disease progression:

928.0 (se 18.2) OS days for LET only

982.8 (se 16.8) OS days for LAP+LET

i.e. a **gain of 54.8** (se 24.8) OS days attributable to use of LAP

Patient utility values

Health state utility values for patients in PFS obtained from the Lloyd et al[105] model differ slightly between the EGF30008[51] trial groups due to differential treatment response rates:

0.7623 (se 0.1141) in the LET group

0.7663 (se 0.1136) in the LAP+LET group.

Adverse events

Examining the incidence of the six grade 3/4 adverse events featured in the Lloyd et al[105] model (fatigue, diarrhoea/vomiting, stomatitis, febrile neutropenia, hand-foot syndrome and alopecia) showed generally low incidence in all categories with the exception of diarrhoea/vomiting which was six times more common in the combination group than the LET group. However, the absolute difference in estimated utility per study patient is very small (less than 0.01) with a wide confidence interval, so the AG decided to examine the influence of the disutility of this AE through sensitivity analysis rather than through setting a value in the base case. A similar approach was taken to the differential cost per patient of AEs.

Cost of lapatinib: acquisition

The acquisition cost of LAP is £804.30 per pack of 70 tablets. The standard dose requires patients to take six tablets per day. In the AG model it is assumed that LAP is prescribed to non-progressed patients every 28 days, with sufficient packs to complete treatment for the next four weeks taking account of any unused tablets from previous prescriptions. This requires two or three packs to be prescribed at each visit. Wastage is automatically included in this calculation as the dispensed tablets are unused at the time of progression (on average 14 days supply), and no mid-cycle correction is necessary. It is assumed that prescriptions will be dispensed by a hospital pharmacist.

Cost of letrozole: acquisition

The acquisition cost of LET is £66.50 per pack of 28 tablets. It is assumed that one pack is dispensed every 28 days to all patients remaining in PFS on that day. This implies that wastage is limited to an average 14 days of treatment per patient. It is assumed that prescriptions will be dispensed by a community pharmacist, except for the first prescription which is provided in the hospital.

Actual and expected delivery of treatment

Information provided by the manufacturer of LAP indicates that adjustments made to the dose intensity of treatments in the EGF30008[51] trial were based on similar pill counts to those discussed in the next section in relation to ANA. The AG has no reason to consider these data are any more reliable than those for ANA which were assessed at the individual patient level. The AG therefore decided not to make adjustments to calculations based on 100% compliance with the treatment protocol. This ensures a consistent approach in both appraisals.

5.4.3 Specific model features and parameters: TRA+ANA vs ANA

Expected PFS

The manufacturer of TRA provided full details of the survival analyses (PFS, PPS and OS) requested by the ERG relating to data from the TAnDEM[52] trial. Examination of the cumulative hazard plots for PFS suggested *****. It was found that a two parameter Weibull model offered an acceptable representation of the long-term trend in both groups as shown in **Error! Reference source not found.**. The AG decided that the most reliable estimate of the mean expected PFS would be obtained by using the Kaplan-Meier area-under-curve (AUC) estimate up to the last recorded event in each group, and then adding the area under the projected long-term Weibull model curve at later times.

This approach yielded estimates of:

189.6 (se 21.4) PFS days for ANA only

514.8 (se 64.1) PFS days for TRA+ANA

i.e. a **gain of 325.1** (se 67.6) PFS days attributable to use of TRA

Confidential information withdrawn

Expected PPS

Examination of the cumulative hazard plots for PPS indicated that disease progression was not associated with any variation in risk away from a continuous long-term trend. However, analysis confirmed that simple exponential models (i.e. linear trends in cumulative hazard) were not adequate to describe the observed PPS data. Two parameter Weibull models were fitted to data from both trial groups and offered an acceptable representation as shown in **Error! Reference source not found.**. The AG decided that the most reliable estimate of the mean expected PFS would be obtained by using the Kaplan-Meier AUC estimate up to the last recorded event in each group, and then adding the area under the projected long-term Weibull model curve at later times.

Confidential information withdrawn

Using this approach generated estimates of:

869.6 (se 46.3) PPS days for ANA only

649.6 (se 63.1) PPS days for TRA+ANA

i.e. a **loss of 220.0** (se 78.3) PPS days attributable to use of TRA.

Expected overall survival

Combining estimates of mean PFS and mean PPS in each group, and adjusting for the minority of patients who die at or before progression (5.8% in the ANA group and 9.3% in the TRA+ANA group), combined estimates for OS were obtained:

1009.0 (se 50.5) OS days for ANA only

1101.3 (se 85.6) OS days for TRA+ANA

i.e. a **gain of 92.2** (se 99.4) OS days attributable to use of TRA.

The manufacturer of TRA has drawn attention to two factors in the TAnDEM[52] trial which are considered likely to distort the estimation of PPS in the comparator group:

- a large number of patients in the comparator (ANA) group chose to 'cross-over' to TRA+ANA therapy following disease progression, and are likely to gain additional benefit in terms of extended PPS;

- a greater proportion of patients in the comparator group received second-line chemotherapy, potentially also extending PPS.

In the MS, the results of applying a statistical technique to attempt to counter the first of these confounding factors were presented, but no attempt was made to make any further adjustment to overcome the second-line chemotherapy imbalance.

As discussed in section 5 the suggested statistical adjustment is not universally accepted as the most suitable method to employ, and may rely on restrictive assumptions that are not valid for use in the TAnDEM[52] trial. The AG asked for additional survival analyses to be undertaken in order to explore the sensitivity of OS estimates to other approaches to correcting PPS for cross-over and imbalance in second-line chemotherapy; how the results of the survival analyses were used by the AG are discussed below.

Cross-over: Separate Kaplan-Meier analyses of patients in the comparator group of the TAnDEM[52] trial, split by whether they did or did not receive cross-over TRA following progression, demonstrated a clear advantage for cross-over patients. However, the data suggest that, after about 6 months have elapsed, this advantage diminishes, and disappears altogether after about three years. The AG found that these complex trends in PPS could be well described by fitting bi-phase exponential models (**Error! Reference source not found.**), from which it is possible to estimate the mean survival gain attributable to cross-over for patients in the post-progression phase.

The net benefit of cross-over (i.e. the area between the two modelled PPS lines) is estimated as 150.5 days. However, this advantage only accrues to those patients who do not die at or before progression (91% of the total), so that the mean PPS adjustment which may be subtracted in the calculation of OS in the control group is 137.5 (se 11.7) days.

Confidential information withdrawn

Second-line chemotherapy: Four Kaplan-Meier analyses of TAnDEM[52] patients, stratified by treatment group and post-progression use of second-line chemotherapy, were compared. No clear distinctions were apparent, though in both trial groups those receiving second-line therapy seemed to have a modest advantage. Generally, the hazard time profiles did not markedly differ from linearity indicating that an exponential parametric model would be appropriate. The AG chose to compare all patients receiving second-line chemotherapy to all patients who did not, recognising that this is necessarily only an exploratory analysis lacking a full standardisation. However, in view of the small numbers of patients in each stratum, more detailed analysis would most likely be unproductive. Exponential survival parameters were estimated suggesting a hazard ratio of 0.83 in favour of chemotherapy, and a gain in PPS of 145.2 (se 31.1) days. This figure must be adjusted for three factors:

- the difference in use of second-line chemotherapy between the trial groups is 24% (32% - 8%);

- examination of the trial data indicated that the majority (82%) of chemotherapy patients also benefited from cross-over TRA, and for these patients the effect of chemotherapy is already included in the cross-over adjustment discussed above;

- the absolute difference in PPS only applies to patients who did not die at or before disease progression.

The net effect of these adjustments is an estimated additional second-line chemotherapy gain in PPS in the comparator group of 5.9 (se 1.2) days.

When the adjustments for cross-over and chemotherapy are included, the following final estimates for OS were obtained:

861.2 (se 52.1) OS days for ANA only

1101.3 (se 85.6) OS days for TRA+ANA

i.e. a **gain of 240.1** (se 100.2) OS days attributable to use of TRA.

Patient utility values

Health state utility values for patients in PFS obtained from the Lloyd et al[105] model, differ slightly been the TAnDEM[52] trial groups due to differential treatment response rates:

0.7639 (se 0.1139) in the ANA group

0.7687 (se 0.1133) in the TRA+ANA group.

Adverse events

Examining the incidence of the six grade 3/4 AEs featured in the Lloyd et al[105] model (fatigue, diarrhoea/vomiting, stomatitis, febrile neutropenia, hand-foot syndrome and alopecia) showed very low incidence in all categories and no significant differences on which to base any estimate of differential disutility from AEs for this comparison; the AG decided to examine the potential importance of this issue via sensitivity analysis. A similar approach was taken to the differential cost per patient of AEs.

Cost of TRA: acquisition and administration

The cost of TRA treatment was estimated using the distribution of body weight recorded at baseline in the TAnDEM[52] trial. These data indicated that a lognormal distribution was appropriate, and parameters were estimated by the method of moments (i.e. a weighted average of the individual doses and vials of TRA which would be required to treat the population of patients without vial sharing was estimated). This calculation automatically incorporated drug wastage. For the initial loading dose (8mg/kg) the cost per dose was estimated as £1,657.86 and for a regular dose (6mg/kg) the cost per dose is £1,292.88. These costs were applied to all patients remaining in PFS at the beginning of each 3-week period.

The costs of administering TRA are derived from the NHS Reference Costs 2008-09,[101] using average costs for Day cases and Outpatient weighted by national activity levels. For the loading dose, this uses HRG code SB14Z, and for the regular dose, code SB12Z as specified in clinical coders guidance.[115] The unit cost per treatment is £284.66 (loading dose) and £198.63 (regular doses).

Cost of ANA: acquisition

The acquisition cost of ANA is £68.56 per pack of 28 tablets. It is assumed that one pack is dispensed every 28 days to all patients remaining in PFS on that day. This implies that wastage is limited to an average 14 days of treatment per patient. It is assumed that prescriptions will be dispensed by a community pharmacist, except for the first prescription which is provided in the hospital.

Actual and expected delivery of treatment

In their submitted model the manufacturer of TRA adjusted the quantity of each treatment by a multiplier to represent the ratio of treatment actually received by patients and that expected during their time in PFS. This seems to be a compound of patient compliance, missed doses, dose adjustments and the lack of precision in the estimation of treatment volumes in their model. For TRA infusions this factor has a minor effect (x 0.987), but the effect is more pronounced for ANA. The AG considered this issue carefully on the basis of the detailed individual treatment records included in the TAnDEM[52] CSR. In the case of TRA infusions there are almost no occasions when scheduled infusions were not administered on time and at the prescribed dose. A few instances of a missed appointment (e.g. at Christmas holiday) were generally followed by a double dose administered at the next scheduled visit. It appears therefore that the trial data do not support the notion that there is any serious systematic discrepancy between planned and administered delivery of TRA. There may be some merit in a minor adjustment in the submitted model to take account of the approximation involved in estimating PFS at monthly intervals, but this problem does not arise in the AG's model which calculates PFS daily.

For ANA the estimation of an adjustment factor appears to have been based on estimated compliance data using pill counts undertaken during the trial. The individual patient data track the issue of tablet packs at each patient visit, and the number of tablets returned unused at the following visit. These data reveal that the method of calculating compliance is fundamentally unsound, since it takes no account of the occasional failure of patients to return unused tablets during the dosing period, and a systematic failure to return unused tablets at the end of treatment. As a consequence individual compliance figures ranging between 32% and 300% were estimated. Closer examination of individual patients' drug issues and returns shows a generally exemplary adherence to schedule in all patients.

The AG is satisfied that there is no evidential basis for making any adjustments to the calculated expected use of either treatment on the grounds of deviation from treatment protocol, or to correct for approximations arising from the model structure since the AG model is designed to avoid such problems.

5.4.4 Results

The results obtained from modelling the costs and outcomes of each of the trial-based appraisals are shown separately in this section. No attempt has been made to make any comparisons between the groups of the two trials, as the populations are not considered to be directly comparable, and reliable indirect comparisons of treatment effects could not be undertaken (see section 5). Therefore the only questions which may be addressed legitimately are:

- can LAP+LET be considered a cost-effective alternative to LET alone?

- can TRA+ANA be considered a cost-effective alternative to ANA alone?

Base case result: LAP+LET vs LET

The base case cost-effectiveness results based on the AG model are shown in Table 26. A small expected mean health gain per patient (less than 2 months life extension, and under 0.12 additional QALYs) is generated by an additional cost of more than £25,000 per patient most of which is incurred in the first 5 years. The cost-effectiveness ratio is stable over long time periods and exceeds £220,000 per QALY gained.

Univariate sensitivity analysis

Results from a sensitivity analysis covering the main model variables are shown in Table 27. The ICER is most sensitive to the health state utility parameter values, and to the cost of LAP, but is insensitive to most of the other variables. In all cases the ICER remains above £137,000 per QALY indicating that uncertainty in any single parameter value is unlikely to alter the cost effectiveness of LAP+LET relative to conventional thresholds.

Table 26 Cost-effectiveness results for base case analysis of LAP+LET vs LET (discounted) using AG model

Treatment	Cost per patient						Outcomes per patient		ICER
Time horizon (years)	Drugs	Monitoring	Adverse events	BSC	Terminal care	Total costs	Life years	QALYs	£ / QALY gained
LET									
10	£718	£757	-	£12,266	£1,622	£15,362	2.526	1.444	
20	**£718**	**£757**	**-**	**£12,407**	**£1,643**	**£15,524**	**2.549**	**1.455**	
30	£718	£757	-	£12,408	£1,643	£15,525	2.549	1.455	
LAP+LET									
10	£25,082	£1,397	£98	£12,374	£1,622	£40,573	2.670	1.558	
20	**£25,082**	**£1,397**	**£98**	**£12,513**	**£1,643**	**£40,733**	**2.693**	**1.570**	
30	£25,082	£1,397	£98	£12,514	£1,643	£40,734	2.693	1.570	
Incremental									
10	£24,364	£640	£98	£108	£0	£25,211	0.145	0.114	£220,252
20	**£24,365**	**£640**	**£98**	**£106**	**£0**	**£25,209**	**0.144**	**0.114**	**£220,626**
30	£24,365	£640	£98	£106	£0	£25,209	0.144	0.114	£220,628

Table 27 Univariate sensitivity analysis of the cost-effectiveness results of LAP+LET vs LET to variations in main variables in AG model (base case with 20 year horizon)

	Variation in value		ICER for:	
Model variable	**low**	**high**	**low**	**high**
Base case	-	-	**£220,628**	
Discount rate – costs	0%	6%	£223,696	£218,618
Discount rate – outcomes	0%	6%	£222,875	£219,318
Dispensing costs: community	£5	£10	£220,416	£221,005
Dispensing costs: hospital	£7	£11	£220,610	£220,646
Frequency of cardiac monitoring (p.a.)	3	6	£219,875	£222,464
ECHO as % of scans	50%	100%	£221,528	£219,279
Frequency of PFS follow-up & CT scan (p.a.)	2	6	£219,674	£221,307
Proportion of progressed patients on exemestane	0%	100%	£220,950	£220,306
Net extra cost of AEs in LAP+LET group	£0	£1,000	£219,766	£228,382
Net extra disutility of AEs in LAP+LET group	0	-0.01	£206,767	£224,907
Utility in PFS: LET only	-10%	+10%	£149,469	£421,474
Utility in PFS: LAP+LET	-10%	+10%	£547,822	£138,190
Utility in PFS: both groups	-10%	+10%	£250,889	£196,886
Utility in PPS	-10%	+10%	£217,583	£223,760
Acquisition cost of LAP	-10%	+10%	£199,558	£241,698
Cost of cardiac scan	-10%	+10%	£220,233	£221,023
BSC annual costs	-10%	+10%	£220,503	£220,753
Terminal care costs	-10%	+10%	£220,628	£220,628

Probabilistic sensitivity analysis

Probabilistic sensitivity was explored running 1000 random iterations for all variables subject to measurable parameter uncertainty, using the base case scenario over a 20 year horizon. The PSA results are compared with the corresponding deterministic results in Table 28.

Table 28 Comparison of deterministic and probabilistic cost-effectiveness results for LAP+LET vs LET (base case with 20 year horizon)

	Incremental cost	**Incremental QALYs**	**ICER**
Deterministic	£25,209	0.114	£220,628
PSA	£24,878	0.009	£2,895,994

The scatterplot of iteration results in the cost-effectiveness plane (Figure 11) indicates that all iterations lie substantially outside the region normally considered cost effective. Figure 12 confirms that there is no measurable probability of the combination therapy being cost effective at a willingness-to-pay threshold of £40,000 per QALY gained, and does not reach 50% probability until nearly £3,000,000 per QALY gained due to the serious uncertainty concerning whether the combination treatment delivers any real benefit to patients in the long-term.

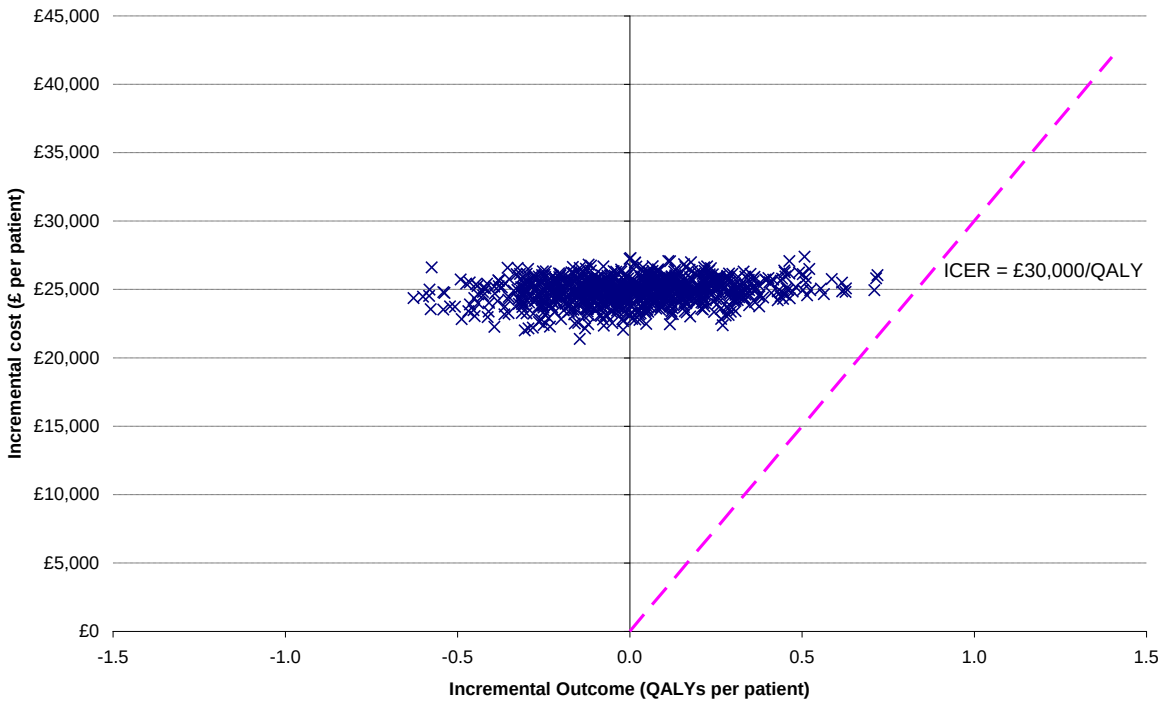

Figure 11 PSA of LAP+LET vs LET only: scatterplot of 1000 probabilistic iterations

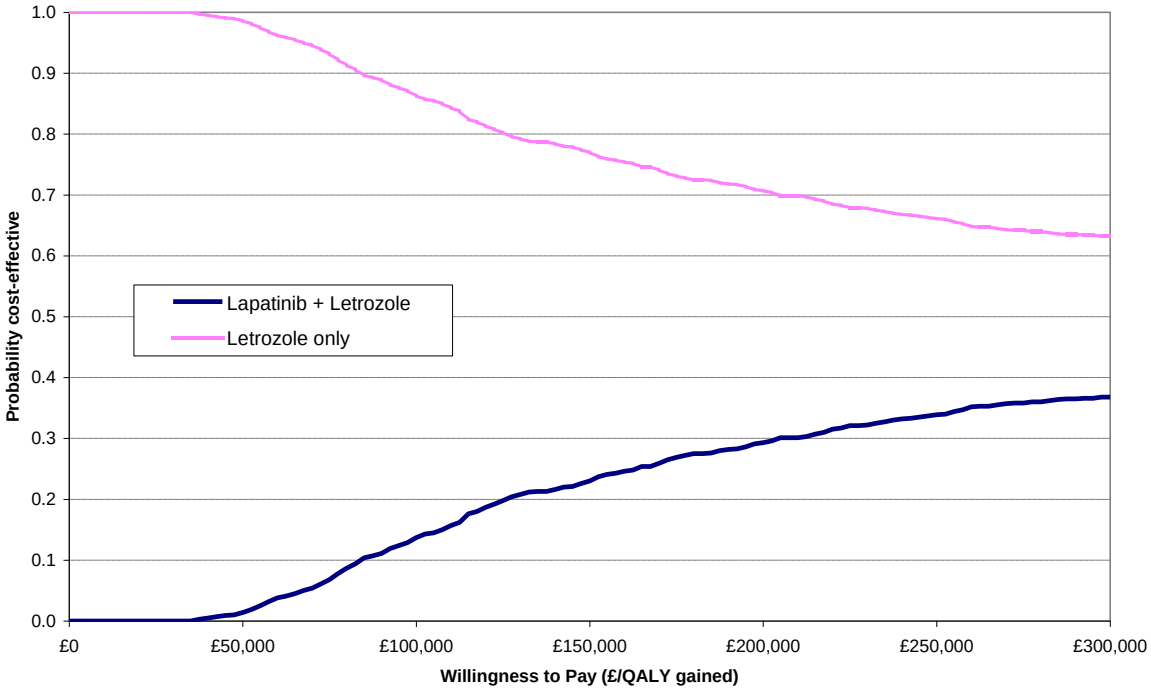

Figure 12 PSA of LAP+LET vs LET only: cost-effectiveness acceptability curve

Base case results: TRA+ANA vs ANA

The base case cost-effectiveness results based on the AG model are shown in Table 29. A modest expected mean health gain per patient (less than 6 months life extension, and under 0.5 additional QALYs) is generated by a substantial additional cost of more than £35,000 per patient most of which is incurred in the first 5 years. The cost-effectiveness ratio is stable over long time periods and exceeds £80,000 per QALY gained.

Univariate sensitivity analysis

Results from a sensitivity analysis covering the main model variables are shown in Table 30. The ICER is most sensitive to the health state utility parameter values, and to the cost of TRA, and discounting rates, but very insensitive to most of the other variables. In all cases the ICER remains above £65,000 per QALY indicating that uncertainty in any single parameter value is unlikely to alter the cost effectiveness of TRA+ANA relative to conventional thresholds.

Table 29 Cost-effectiveness results for base case analysis of TRA+ANA vs ANA (discounted) using AG model

Treatment	Cost per patient						Outcomes per patient		ICER
Time horizon (years)	Drugs	Monitoring	Adverse events	BSC	Terminal care	Total costs	Life years	QALYs	£ / QALY gained
ANA									
10	£549	£602	-	£11,101	£1,632	£13,884	2.204	1.235	
20	**£549**	**£602**	**-**	**£11,194**	**£1,647**	**£13,992**	**2.220**	**1.243**	
30	£549	£602	-	£11,194	£1,648	£13,993	2.220	1.243	
TRA+ANA									
10	£35,197	£1,843	£90	£10,664	£1,695	£49,488	2.652	1.660	
20	**£36,251**	**£1,898**	**£92**	**£10,741**	**£1,696**	**£50,679**	**2.692**	**1.690**	
30	£36,370	£1,905	£93	£10,749	£1,696	£50,813	2.696	1.694	
Incremental									
10	£34,648	£1,241	£90	-£437	£63	£35,604	0.448	0.425	£83,776
20	**£35,702**	**£1,297**	**£92**	**-£453**	**£49**	**£36,687**	**0.472**	**0.448**	**£81,956**
30	£35,821	£1,303	£93	-£445	£49	£36,820	0.476	0.451	£81,644

Table 30 Univariate sensitivity analysis of the cost-effectiveness results of TRA+ANA vs ANA to variations in main variables in AG model (base case with 20 year horizon)

	Variation in value		ICER for:	
Model variable	**low**	**high**	**low**	**high**
Base case	-	-	**£81,956**	
Discount rate – costs	0%	6%	£87,366	£78,830
Discount rate – outcomes	0%	6%	£75,988	£85,735
Dispensing costs: community	£5	£10	£81,923	£82,016
Dispensing costs: hospital	£7	£11	£81,854	£82,059
Frequency of cardiac monitoring (p.a.)	3	6	£81,634	£82,611
ECHO as % of scans	50%	100%	£82,279	£81,473
Frequency of PFS follow-up & CT scan (p.a.)	2	6	£81,258	£82,707
Proportion of progressed patients on exemestane	0%	100%	£82,346	£81,566
Net extra cost of AEs in TRA+ANA group	£0	£1,000	£81,750	£83,816
Net extra disutility of AEs in TRA+ANA group	0	-0.01	£81,956	£83,685
Utility in PFS: ANA only	-10%	10%	£75,208	£90,035
Utility in PFS: TRA+ANA	-10%	10%	£105,508	£67,000
Utility in PFS: both groups	-10%	10%	£94,583	£72,304
Utility in PPS	-10%	10%	£79,300	£84,797
Administration of TRA costs	-10%	10%	£80,917	£82,995
Acquisition cost of TRA costs	-10%	10%	£75,238	£88,675
Cost of cardiac scan	-10%	10%	£81,815	£82,098
BSC annual costs	-10%	10%	£82,019	£81,894
Terminal care costs	-10%	10%	£81,945	£81,967

Probabilistic sensitivity analysis

Probabilistic sensitivity was explored running 1000 random iterations for all variables subject to measurable parameter uncertainty, using the base case scenario over a 20 year horizon. The PSA results are compared with the corresponding deterministic results in Table 31.

Table 31 Comparison of deterministic and probabilistic cost-effectiveness results for TRA+ANA vs ANA (base case with 20 year horizon)

	Incremental cost	**Incremental QALYs**	**ICER**
Deterministic	£36,687	0.472	£81,956
PSA	£32,277	0.363	£88,933

The scatterplot of iteration results in the cost-effectiveness plane (Figure 13) indicates a strong positive correlation between incremental cost and incremental benefit. Figure 14 confirms that there is no measurable probability of the combination therapy being cost effective at a willingness-to-pay threshold of £50,000 per QALY gained, and only a 4.1% probability at £60,000 per QALY gained.

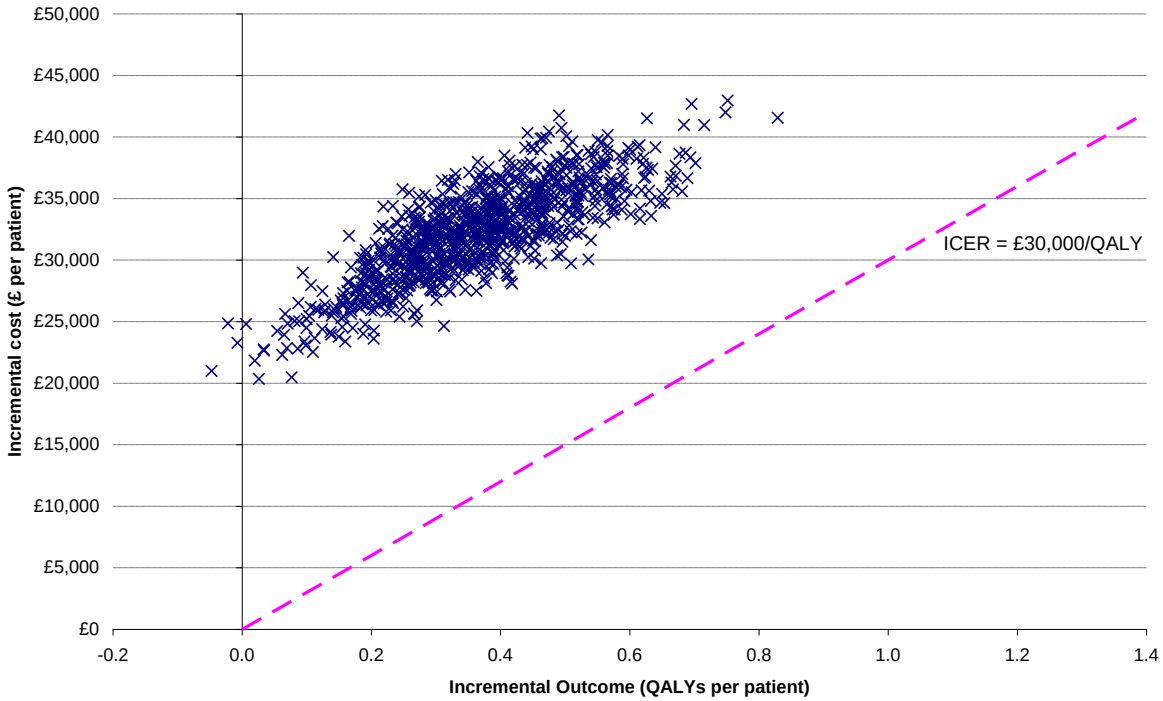

Figure 13 PSA of TRA+ANA vs ANA only: scatterplot of 1000 probabilistic iterations

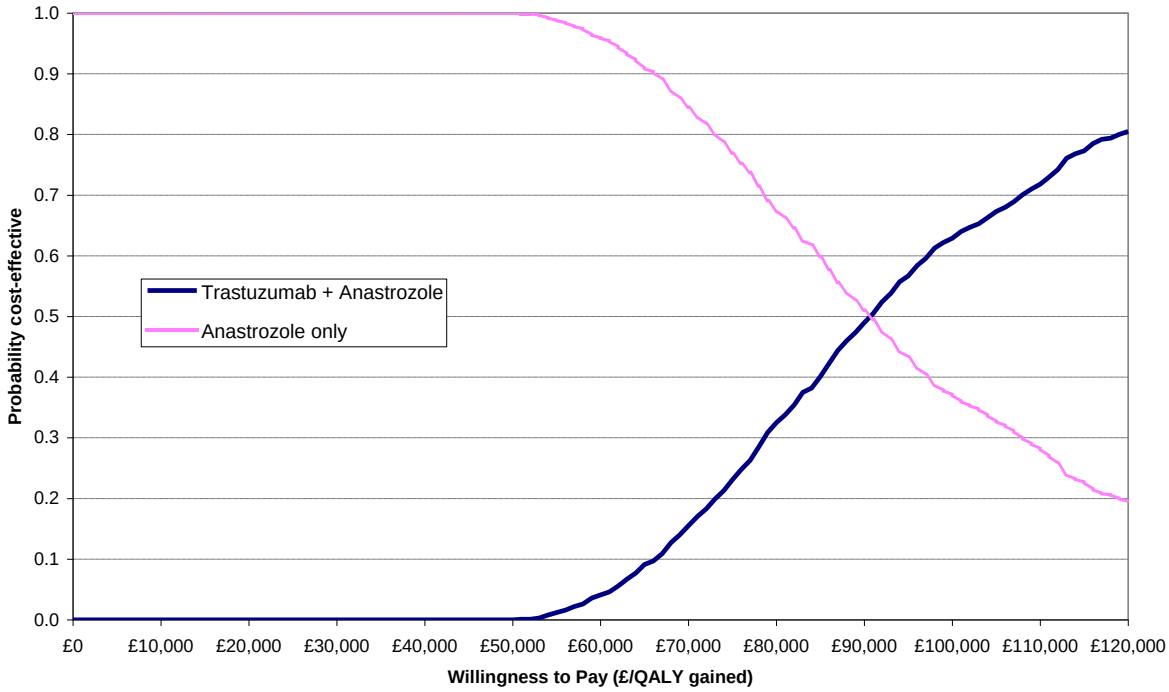

Figure 14 PSA of TRA+ANA vs ANA only: cost-effectiveness acceptability curve

5.5 Summary of cost-effectiveness evidence

5.5.1 Cost-effectiveness review

In summary, the AG did not identify any relevant papers for inclusion in the cost effectiveness review of LAP+AI or TRA+AI in patients who are HR+/HER2+ with MBC. The manufacturer of TRA identified a poster[98] which was presented at the ASCO 2010 conference; the study described compared LAP+LET vs TRA+ANA using an indirect comparisons analysis. The AG is of the opinion that the results of the indirect analysis performed by Hastings et al[98] are unreliable as the studies which make up the evidence network are inappropriate. In addition, the AG notes that without access to more detailed information on costs, it is difficult to comment on the reliability of the cost-effectiveness results in this study.

5.5.2 Submitted economic evaluations by manufacturers

The two economic evaluations submitted by the manufacturers appear to meet the NICE reference case criteria.[99] However, the AG is critical of the approaches used by the manufacturers to estimate OS in each of their models; the AG is of the opinion that projective modelling in this group of patients can lead to substantial bias in OS estimates. In addition, the AG also identified several costing inaccuracies and inconsistencies in both of the economic evaluations submitted.

For the direct comparisons, GlaxoSmithKline demonstrated that LAP+LET is not cost effective compared with LET and Roche demonstrated that TRA+ANA is not cost effective compared with ANA.

Both of the manufacturers undertook indirect comparisons analyses in order to be able to compare LAP+LET vs TRA+ANA. GlaxoSmithKline demonstrated that LAP+LET is cost effective compared with TRA+ANA. Roche demonstrated that TRA+ANA is cost effective compared with LAP+LET. The AG concludes that the indirect comparisons analyses conducted by the manufacturers are unreliable and that only the ICERs estimated from the direct comparisons are valid.

Roche makes the case for TRA+ANA to be considered as an end of life treatment for women with HR+/HER2+ MBC. The AG does not have sufficient information to verify whether all three NICE criteria for consideration of end of life treatments are met.

5.5.3 AG's cost-effectiveness results and sensitivity analysis

The AG reports the results of two separate *de novo* cost-effectiveness analyses using a common framework and common parameter values but employing effectiveness data drawn only from a single RCT (either EGF30008[51] or TAnDEM[53]). The AG model employs outcome data derived from the relevant clinical trial in the form of Kaplan-Meier estimated survival values augmented by projected survival estimates calibrated against the observed data. The AG uses PFS and PPS estimates directly as the basis for calculating expected OS in each group of the RCT.

As the AG is of the opinion that the evidence base is too unstable to allow meaningful comparison of LAP+AI vs TRA+ANA, the only questions that may be addressed legitimately are:

- Can LAP+LET be considered a cost-effective treatment compared with LET alone?
- Can TRA+ANA be considered a cost-effective treatment compared with ANA alone?

Base case result: LAP+LET vs LET

The AG concludes that in HR+/HER2+ women with MBC, LAP+LET compared with LET is not cost effective. Using a time horizon of 20 years, the AG estimates an ICER which exceeds £220,000 per QALY gained for the comparison of LAP+LET vs LET; the incremental total costs and QALYs per patient treated are estimated as £25,209 and 0.114 respectively.

Base case result: TRA+ANA vs ANA

The AG concludes that in HR+/HER2+ women with MBC, TRA+ANA compared with ANA is not cost effective. Using a time horizon of 20 years, the AG estimates an ICER which exceeds £80,000 per QALY gained for the comparison of TRA+ANA vs ANA; the incremental total costs and QALYs per patient treated are estimated as £36,687 and 0.448 respectively.

LAP+AI vs TRA+AI

The AG emphasises again that the currently available clinical evidence base is too unstable to allow meaningful comparison of LAP+AI vs TRA+AI.

Sensitivity analyses undertaken by the AG

For the comparison of LAP+LET vs LET the univariate sensitivity analysis shows that the ICER is most sensitive to the choice of health state utility parameter values, the cost of LAP and is insensitive to most of the other variables. In all cases, the ICER remains above £137,000 per QALY gained. The PSA shows that there is no measureable probability of LAP+LET being cost effective at a willingness-to-pay threshold of £40,000 per QALY gained; to achieve a 50% probability of LAP+LET being cost effective, the willingness-to-pay threshold needs to increase to around £3,000,000 per QALY gained.

For the comparison of TRA+ANA vs ANA, the univariate sensitivity analysis shows that the ICER is most sensitive to the choice of health state utility parameter values, the cost of TRA and discounting rates only. In all cases, the ICER exceeds £65,000 per QALY gained. The PSA shows that there is no measureable probability of TRA+ANA being cost effective compared to ANA at a willingness-to-pay threshold of £50,000.

6 DISCUSSION

The size of the relevant study population of interest to this appraisal is small. The manufacturers of both LAP and TRA are in agreement that the eligible population of women with HR+/HER2+MBC in England and Wales is approximately 50 patients per year.

Only three RCTs[51, 52, 54] have been identified which present head-to-head comparisons of the interventions of interest to this appraisal. It was not possible to compare the data across the trials because of differences in the patient populations. Nevertheless, all three[51, 52, 54] suggest that either LAP or TRA in combination with an AI improves efficacy, in terms of PFS and/or TTP, over an AI alone. These findings are only statistically significantly different in two trials[51, 52] as the eLEcTRA[54] trial lacks statistical power due to being halted early due to slow recruitment. The trials[51, 52] which measured OS did not report any statistically significant differences between treatment groups. Adverse events were more common in the groups in which either LAP or TRA was given in combination with an AI but on the whole, were of grade 1 or 2 severity. However, around 1% of patients taking LAP+LET had to discontinue their treatment as a result of AEs related to diarrhoea. No new safety concerns were reported in any of the trials.

The comparison of LAP+AI vs TRA+AI is also of interest to this appraisal and, as there are no head-to-head trials of these interventions, the manufacturers used indirect comparisons analyses using mainly clinical data from EGF30008[51] and TAnDEM[52] to assess this comparison. The AG believes that the results of any indirect comparisons analyses of LAP+LET vs TRA+ANA using data from EGF30008[51] and TAnDEM[52] are unreliable due to heterogeneous patient populations. The AG considers that there are apparent differences in the study populations of these two key trials which prohibit comparison of patients and therefore results; these differences may be explained by the fact that patients were excluded from the EGF30008 trial[51] if their disease was rapidly progressing or life threatening.

In addition, to complete the evidence network in the indirect comparisons analyses presented in the submitted MS, the manufacturers had to use trials with mixed HER2- and HER2+ populations. The AG is of the opinion that use of clinical effectiveness evidence from a mixed population adds to the uncertainty regarding the results of the indirect analyses conducted by the manufacturers. To illustrate, in EGF30008[51] which included both HR+/HER2+ and HR+/HER2- populations, the clinical effectiveness of LET appears to be compromised in patients who are HR+/HER2+ compared to patients who are HR+/HER2-; this was also apparent for patients in the LAP+LET arm of the trial but to a lesser extent. The significance of this finding is unclear but from a purely clinical viewpoint, could suggest that LET alone is relatively ineffective in patients with HR+/HER2+ MBC; more

evidence is required from other studies which are able to compare HR+/HER2+ and HR+/HER2- populations.

In summary, the AG is of the opinion that it is not useful to compare findings across the two trials[51, 52] because of heterogeneous patient populations. In addition, reliance on clinical evidence from a mixed population adds to the uncertainty of the validity of the results for a HR+/HER2+ population. The AG therefore considers the results of the indirect comparisons analysis presented by the manufacturers to be unreliable.

A final issue that needs to be considered relates to the generalisability of these trials to the actual population of interest in the UK, namely post-menopausal women with HR+/HER2+ MBC, who have not previously received treatment for MBC and for whom treatment with an AI is suitable. None of the patients in EGF30008[51] or TAnDEM[52] have received prior treatment with TRA; this is not surprising as, at the time the trials were recruiting, the use of TRA for patients with early or advanced breast cancer was relatively rare. This contrasts very much with what happens in clinical practice in the NHS today. Now, when a patient is diagnosed with early HER2+ breast cancer, TRA is the standard treatment of choice and in reality it is likely that only *de novo* patients with HR+/HER2+ MBC will be eligible for TRA+AI as per the wording of the EMA licence (i.e. TRA-naive). Patients who have been treated with TRA previously are eligible for treatment with LAP+AI; however, whether the clinical effectiveness of LAP+AI is the same for patients who are and who are not TRA-naive is uncertain.

From a health economics perspective, the AG has confirmed by its independent analyses the assertion made by both manufacturers that LAP+LET and TRA+ANA are not cost effective compared with AIs alone for women with HR+/HER2+MBC. The ICERs estimated by the AG for LAP+LET vs LET and TRA+ANA vs ANA are higher than those estimated by the manufacturer.

The AG is of the opinion that the protocol criteria governing the selection of patients for EGF30008[51] and TAnDEM[52] are sufficiently dissimilar to be likely to generate non-equivalent patient populations. This means that the results of any indirect comparisons analyses that include both of these trials in the evidence network are unreliable. Consequently, the AG did not address the cost effectiveness of LAP+LET vs TRA+ANA as there were insufficient comparative clinical data available to allow estimation of meaningful ICERs.

7 CONCLUSIONS

Clinical effectiveness evidence from two good quality RCTs[51, 52] demonstrates that LAP+LET or TRA+ANA improves median PFS and/or TTP compared with AI monotherapy in patients who are HR+/HER2+ MBC. To date, the trials[51, 52] do not show a statistically significant benefit in terms of OS for patients taking LAP+LET vs AI monotherapy or TRA+ANA vs AI monotherapy. The results of the economic evaluations conducted by the manufacturers and confirmed by the AG demonstrate that LAP+LET is not cost effective compared with AI monotherapy, nor is TRA+ANA cost effective compared with AI monotherapy.

Due to differences in the patient populations of EGF30008[51] and TAnDEM[52] the AG believes the results of the indirect comparisons analyses conducted by the manufacturers are inappropriate and for the same reason believes that it would be unsound to compare LAP+LET vs TRA+ANA in an economic evaluation.

7.1 Suggested research priorities

As EGF30008[51] reports, there were large differences in PFS for HER2+ and HER2- patients receiving both LAP+LET and, in particular, LET. Further research may be warranted comparing the clinical effectiveness of AIs as monotherapy in patients with HER2+ and HER2- breast cancer.

Given most patients who present for HR+/HER2+ MBC are likely to have been previously treated for early breast cancer and given this is almost certain to have included TRA (unlike at the time the pivotal trials in this appraisal were conducted), further research may be required into treating MBC in the HR+/HER2+ population who are not TRA (or LAP) naive.

As increasingly, trials allow patients to cross-over following disease progression, attempts should be made to consider how to adjust for cross-over at the trial design stage, rather than as a post-hoc analysis.

8 REFERENCES

1. Germano S, O'Driscoll L. Breast cancer: Understanding sensitivity and resistance to chemotherapy and targeted therapies to aid in personalised medicine. Current Cancer Drug Targets. 2009; 9(3):398-418.

2. Cancer Research UK. Breast cancer - risk factors. 2010 [updated 30/06/2010]; Available from: http://info.cancerresearchuk.org/cancerstats/types/breast/riskfactors/ (Accessed: 06/09/2010).

3. Ward S, Simpson E, Davis S, Hind D, Rees A, Wilkinson A. Taxanes for the adjuvant treatment of early breast cancer: systematic review and economic evaluation. *Health Technol Assess*. 2007; 11(40).

4. Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BA. Polygenic susceptibility to breast cancer and implications for prevention. Nat Genet. 2002; 31(1):33-6.

5. Hannah L CT, MacBeth F, editor. Practical Clinical Oncology. First ed: Cambridge University Press; 2008.

6. Endogenous Hormones Breast Cancer Collaborative Group. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. J Natl Cancer Inst. 2002; 94:606 - 16.

7. Cassidy J, Bissett D, Spence R, Payne M, editors. Oxford Handbook of Oncology. Second ed: Oxford University Press; 2006.

8. American Joint Committee on Cancer. AJCC cancer staging manual. 7th ed. Chicago: Springer; 2009.

9. International Union Against Cancer. TNM classification of malignant tumours. 7th ed. New York: Wiley-Liss; 2009.

10. Cancer Research UK. Predicted improvements in breast cancer survival Cancer Research UK; 2010 [updated 05/07/2010]; Available from: http://www.cancerhelp.org.uk/about-cancer/cancer-questions/predicted-improvements-in-breast-cancer-survival (Accessed: 06/09/2010).

11. Andreopoulou E, Hortobagyi GN. Prognostic factors in metastatic breast cancer: Successes and challenges toward individualized therapy. J Clin Oncol. 2008; 26(22):3660-2.

12. Cold S, Jensen NV, Brincker H, Rose C. The influence of chemotherapy on survival after recurrence in breast-cancer - a population-based study of patients treated in the 1950s, 1960s and 1970s. Eur J Cancer. 1993; 29A(8):1146-52.

13. National Institute for Health and Clinical Excellence (NICE). Guidance on the use of trastuzumab for the treatment of advanced breast cancer. NICE Technology Appraisal Guidance. No.34. 2002; Available from: http://www.nice.org.uk/nicemedia/pdf/TA107guidance.pdf (Accessed: 26/01/2010).

14. Cancer Research UK. Breast cancer - UK incidence statistics. 2010 [updated 08/07/2010]; Available from: http://info.cancerresearchuk.org/cancerstats/types/breast/incidence/ (Accessed: 06/09/2010).

15. Quinn M, Cooper N, Rowan S. Cancer Atlas of the United Kingdom and Ireland 1991-2000. 2005.

16. Woods LR, B. Coleman, MP. Origins of socio-economic inequalities in cancer survival: a review Ann Oncol 2006; 17.

17. Remak E, Brazil L. Cost of managing women presenting with stage IV breast cancer in the United Kingdom. Br J Cancer. 2004; 91(1):77-83.

18. Stebbing J, Slater S, Slevin M. Breast cancer (metastatic). 2007; Available from: http://clinicalevidence.bmj.com/ceweb/conditions/woh/0811/0811_background.jsp (Accessed: 26/01/2010).

19. Cancer Research UK. Trastuzumab (Herceptin). Cancer Research UK; 2009 [updated 5 July 2010]; Available from: http://www.cancerhelp.org.uk/about-cancer/treatment/cancer-drugs/trastuzumab. (Accessed: 06/09/2010)

20. National Institute for Health and Clinical Excellence (NICE). Final scope for the proposed appraisal of lapatinib and trastuzumab in combination with an aromatase inhibitor for the first line treatment of hormone receptor positive metastatic breast cancer which over expresses HER2. January 2010.

21. MBC Advocacy Working Group. Metastatic breast cancer patients: addressing their unmet needs. Community Oncology. 2008; 5(12):645-9.

22. GlaxoSmithKline. Manufacturer Submission To The National Institute for Health and Clinical Excellence By GlaxoSmithKline UK. Multiple technology appraisal of lapatinib (Tyverb®) for the treatment of HR+ HER2+ Metastatic Breast Cancer. Date: 7th June 2010.

23. Turner J, Kelly B, Swanson C, Allison R, Wetzig N. Psychosocial impact of newly diagnosed advanced breast cancer. Psychooncology. 2005; 14(5):396-407.

24. National Institute for Health and Clinical Excellence (NICE). Advanced breast cancer: diagnosis and treatment. NICE clinical guideline 81. 2009; Available from: http://www.nice.org.uk/Guidance/CG81 (Accessed: 26/01/2010).

25. Quinlan E, Thomas-Maclean R, Hack T, Kwan W, Miedema B, Tatemichi S, *et al*. The impact of breast cancer among Canadian women: Disability and productivity. Work. 2009; 34(3):285-96.

26. Yoo GJ, Levine EG, Aviv C, Ewing C, Au A. Older women, breast cancer, and social support. Support Care Cancer. 2009:1-10.

27. Montazeri A, Vahdaninia M, Harirchi I, Ebrahimi M, Khaleghi F, Jarvandi S. Quality of life in patients with breast cancer before and after diagnosis: An eighteen months follow-up study. BMC Cancer. 2008; 8.

28. Kedrin D, Wyckoff J, Boimel PJ, Coniglio SJ, Hynes NE, Arteaga CL, *et al*. ERBB1 and ERBB2 Have Distinct Functions in Tumor Cell Invasion and Intravasation. Clin Cancer Res. 2009; 15(11):3733-9.

29. British National Formulary. BNF 59. 2010; Available from: http://www.bnf.org/bnf/ (Accessed: 06/07/2010).

30. Cancer Research UK. NICE rejects lapatinib for advanced breast cancer. 2010.

31. European Medicines Agency (EMEA). Committee for Medicinal Products for Human use summary of positive opinion for Tyverb. 2008; Available from: http://www.ema.europa.eu/pdfs/human/opinion/Tyverb_38267507en.pdf (Accessed: 26/01/2010).

32. News P. Further indication for Tyverb approved. Pharma News. 2010.

33. Jones AL, Barlow M, Barrett-Lee PJ, Canney PA, Gilmour IM, Robb SD, *et al.* Management of cardiac health in trastuzumab-treated patients with breast cancer: updated United Kingdom National Cancer Research Institute recommendations for monitoring. Br J Cancer. 2009; 100(5):684-92.

34. National Institute for Health and Clinical Excellence (NICE). Hormonal therapies for the adjuvant treatment of early oestrogen-receptor-positive breast cancer. Technology Appraisal TA112. 2006; Available from: http://www.nice.org.uk/TA112 (Accessed: 26/01/2010).

35. Lavelle K, Todd C, Moran A, Howell A, Bundred N, M C. Non-standard management of breast cancer increases with age in the UK: a population based cohort of women greater than or equal to65 years. Br J Cancer. 2007; 96.

36. Wyld L, Garg DK, Kumar ID, Brown H, Reed MWR. Stage and treatment variation with age in postmenopausal women with breast cancer: compliance with guidelines. Br J Cancer. 2004; 90(8):1486-91.

37. Ladva S. Continued use of trastuzumab following disease progression in metastatic breast cancer. London New Drugs Group. 2010.

38. European Medicines Agency (EMEA). Committee for Medicinal Products for Human use post-authorisation summary of positive opinion for Herceptin. 2009; Available from: http://www.ema.europa.eu/pdfs/human/opinion/Herceptin_82246709en.pdf (Accessed: 26/01/2010).

39. Roche. Trastuzumab in combination with aromatase inhibitors for the 1st line treatment of metastatic hormone receptor positive breast cancer which over-expressed HER2. Roche Submission to the National Institute for Health and Clinical Excellence, 7th June 2010

40. Centre for Reviews and Dissemination (CRD). CRD's guidance for undertaking reviews in healthcare: Systematic Reviews (3rd Edition). York: CRD, University of York 2008.

41. Egger M, Smith GD, Altman DG. Systematic reviews in health care – Meta-analysis in context: BMJ books; 2001.

42. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med. 2004; 23(20):3105-24.

43. Langer B. Phase II/III randomized study of Anastrozole with or without Trastuzumab (Herceptin) in postmenopausal women with hormone-receptor positive HER2-overexpressing metastatic breast cancer. Physician Data Query (PDQ). 2001: Available from: http://www.mrw.interscience.wiley.com/cochrane/clcentral/articles/021/CN-00362021/frame.html (Accessed: 30/06/2010).

44. Piccart-Gebhart MJ, Coleman RE. Phase III Randomized Study of Trastuzumab (Herceptin) in Women With HER2-Positive Primary Breast Cancer. Physician Data Query (PDQ). 2002: Available from: http://www.mrw.interscience.wiley.com/cochrane/clcentral/articles/007/CN-00404007/frame.html (Accessed: 30/06/2010).

45. Stein S. Phase III randomized study of letrozole with or without GW572016 in postmenopausal women with stage IIIB or IV breast cancer. Physician Data Query (PDQ). 2004: Available from: http://www.mrw.interscience.wiley.com/cochrane/clcentral/articles/289/CN-00548289/frame.html (Accessed: 30/06/2010).

46. Novartis. Study of the Efficacy and Safety of Letrozole Combined With Trastuzumab in Patients With Metastatic Breast Cancer. Physician Data Query (PDQ). 2006: Available from: http://www.mrw.interscience.wiley.com/cochrane/clcentral/articles/302/CN-00614302/frame.html (Accessed: 30/06/2010).

47. Morris P, Modi S. Trastuzumab-refractory HER2-positive metastatic breast cancer - Use of tanespimycin, a novel Hsp90 inhibitor. American Journal of Hematology/ Oncology. 2008; 7(6).

48. Maung K, O'Shaughnessy JA. Inhibition of ErbB1 and ErbB2 by lapatinib ditosylate, a dual kinase inhibitor: Promising activity in pretreated advanced breast cancer. Clinical Breast Cancer. 2004; 4(6):398-400.

49. Ranganathan A, Moore Z, O'Shaughnessy JA. Trastuzumab prolongs progression-free survival in patients with hormone-sensitive and HER2-overexpressing metastatic breast cancer. Clinical Breast Cancer. 2007; 7(6):450-1.

50. Johnston S, Arteaga C. Lapatinib (Tykerb) plus letrozole (Femara) or letrozole alone for postmenopausal hormone receptor-positive metastatic breast cancer. P and T. 2009; 34(2):102.

51. Johnston S, Pippen Jr J, Pivot X, Lichinitser M, Sadeghi S, Dieras V, *et al.* Lapatinib combined with letrozole versus letrozole and placebo as first-line therapy for postmenopausal hormone receptor - Positive metastatic breast cancer. J Clin Oncol. 2009; 27(33):5538-46.

52. Kaufman B, Mackey JR, Clemens MR, Bapsy PP, Vaid A, Wardley A, *et al.* Trastuzumab plus anastrozole versus anastrozole alone for the treatment of postmenopausal women with human epidermal growth factor receptor 2-positive, hormone receptor-positive metastatic breast cancer: results from the randomized phase III TAnDEM study. J Clin Oncol. 2009; 27(33):5529-37.

53. Schwarzberg LS, Franco SX, Florance A, O'Rourke L, Maltzman J, Johnston SRD. Lapatinib plus letrozole as first-line therapy for HER-2+ hormone receptor-positive metastatic breast cancer. Oncologist. 2010; 15(2):122-9.

54. Huober J, Fasching P, Paepke S, Kubista E, Barsoum M, Wallwiener D, *et al.* Letrozole in Combination with Trastuzumab Is Superior to Letrozole Monotherapy as First Line Treatment in Patients with Hormone-Receptor-Positive, HER2- Positive Metastatic Breast Cancer (MBC) - Results of the eLEcTRA Trial. Cancer Res. 2009; 69(24_MeetingAbstracts):4094-.

55. Gomez HL, Doval DC, Chavez MA, Ang PC-S, Aziz Z, Nag S, *et al.* Efficacy and Safety of Lapatinib As First-Line Therapy for ErbB2-Amplified Locally Advanced or Metastatic Breast Cancer. J Clin Oncol. 2008; 26(18):2999-3005.

56. Cameron D, Casey M, Press M, Lindquist D, Pienkowski T, Romieu C, *et al.* A phase III randomized comparison of lapatinib plus capecitabine versus capecitabine alone in women with advanced breast cancer that has progressed on trastuzumab: updated efficacy and biomarker analyses. Breast Cancer Res Treat. 2008; 112(3):533-43.

57. Food and Drug Administration. Guidance for industry: Clinical trial endpoints for the approval of cancer drugs and biologics. 2007; Available from: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071590.pdf (Accessed: 22/07/2010).

58. Committee for Medicinal Products for Human Use. Guideline on the evaluation of anticancer medicinal products in man. 2005; Available from: http://www.tga.gov.au/docs/pdf/euguide/ewp/020595enrev3.pdf (Accessed: 22/07/2010).

59. Fleming TR, Rothmann MD, Lu HL. Issues in using progression-free survival when evaluating oncology products. J Clin Oncol. 2009; 27(17):2874-80.

60. Saad ED, Katz A. Progression-free survival and time to progression as primary end points in advanced breast cancer: often used, sometimes loosely defined. Ann Oncol. 2009; 20(3):460-4.

61. Sherif B, Sherrill B, Amonkar M, Wu Y, J. Maltzman J, O'Rourke L, *et al.* Lapatinib plus letrozole compared with letrozole alone as first-line therapy in hormone receptor positive HER2+ metastatic breast cancer (MBC): A quality-of-life (QOL) analysis. J Clin Oncol ASCO Annual Meeting Proceedings (Post-Meeting Edition). 2009; 27(15s):Abstr 1039.

62. Robins JM, Tsiatis AA. Correcting for non-compliance in randomized trials using rank preserving structural failure time models. Communications in Statistics Theory 1991; 20(8):2609-31.

63. Mark SD, Robins JM. A method for the analysis of randomized trials with compliance information: an application to the Multiple Risk Factor Intervention Trial. Control Clin Trials. 1993; 14(2):79-97.

64. Morden JP, Lambert PC, Latimer N, Abrams KR, AJ W. Assessing methods for dealing with treatment switching in randomised controlled trials. University of Leicester; 2010; Available from: http://www.le.ac.uk/hs/kra1/2010_BS01.pdf (Accessed: 17/08/2010).

65. National Institute for Health and Clinical Excellence. Final appriasal determination: Sunitinib for the treatment of gastrointestinal stromal tumours. 2009; Available from: http://www.nice.org.uk/nicemedia/live/11817/43174/43174.pdf (Accessed: 17/08/2010).

66. National Institute for Health and Clinical Excellence. Final appraisal determination: Everolimus for the second-line treatment of advanced renal cell carcinoma. 2010; Available from: http://guidance.nice.org.uk/TA/WaveCRS1/48 (Accessed: 17/08/2010).

67. Robins J, Finkelstein D. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with Inverse Probability of Censoring Weighted (IPCW) log-rank tests. Biometrics. 2000; 56 (3):779-88.

68. Branson M, Whitehead J. Estimating a treatment effect in survival studies in which patients switch treatment. Stat Med. 2002; 21(17):2449-63.

69. Walker A, White I, Babiker A. Parametric randomization-based methods for correcting for treatment changes in the assessment of the causal effect of treatment. Stat Med. 2004; 23(4):571-90.

70. Bonneterre J, Thürlimann B, Robertson JF, Krzakowski M, Mauriac L, Koralewski P, *et al.* Anastrozole versus tamoxifen as first-line therapy for advanced breast cancer in 668

postmenopausal women: results of the Tamoxifen or Arimidex Randomized Group Efficacy and Tolerability study. J Clin Oncol. 2000; (22):3748-57.

71. Nabholtz JM, Buzdar A, Pollak M, Harwin W, Burton G, Mangalik A, *et al.* Anastrozole is superior to tamoxifen as first-line therapy for advanced breast cancer in postmenopausal women: Results of a North American multicenter randomized trial. J Clin Oncol. 2000; (22):3758-67.

72. Buzdar AU, Jones SE, Vogel CL, Wolter J, Plourde P, Webster A. A phase III trial comparing anastrozole (1 and 10 milligrams), a potent and selective aromatase inhibitor, with megestrol acetate in postmenopausal women with advanced breast carcinoma. Arimidex Study Group. Cancer. 1997; (4):730-9.

73. Jonat W, Howell A, Blomqvist C, Eiermann W, Winblad G, Tyrrell C, *et al.* A randomised trial comparing two doses of the new selective aromatase inhibitor anastrozole (Arimidex) with megestrol acetate in postmenopausal patients with advanced breast cancer. Eur J Cancer. 1996; 32(3):404-12.

74. Puhan M, Bachmann L, Kleijnen J, ter Riet G, Kessels A. Inhaled drugs to reduce exacerbations in patients with chronic obstructive pulmonary disease: a network meta-analysis. BMC Medicine. 2009; 7(1):2.

75. Bucher H, Guyatt G, Griffith L, Walter S. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol. 1997; 50:683-91.

76. Mouridsen HT. Letrozole in advanced breast cancer: The PO25 trial. Breast Cancer Res Treat. 2007; 105(1 SUPPL.):19-29.

77. Sutton AJ, Higgins JPT. Recent developments in meta-analysis. John Wiley & Sons, Ltd.; 2008. p. 625-50.

78. Buzdar A, Douma J, Davidson N, Elledge R, Morgan M, Smith R, *et al.* Phase III, multicenter, double-blind, randomized study of letrozole, an aromatase inhibitor, for advanced breast cancer versus megestrol acetate. J Clin Oncol. 2001; (14):3357-66.

79. Dombernowsky P, Smith I, Falkson G, Leonard R, Panasci L, Bellmunt J, *et al.* Letrozole, a new oral aromatase inhibitor for advanced breast cancer: double-blind randomized trial showing a dose effect and improved efficacy and tolerability compared with megestrol acetate. J Clin Oncol. 1998; (2):453-61.

80. Kaufmann M, Bajetta E, Dirix LY, Fein LE, Jones SE, Cervek J, *et al.* Exemestane improves survival compared with megoestrol acetate in postmenopausal patients with advanced breast cancer who have failed on tamoxifen. results Of a double-blind randomised phase III trial. Eur J Cancer. 2000:S86-7.

81. Mouridsen H, Sun Y, Gershanovich M, Perez-Carrion R, Becquart D, Chaudri-Ross HA, *et al.* Superiority of letrozole to tamoxifen in the first-line treatment of advanced breast cancer: evidence from metastatic subgroups and a test of functional ability. Oncologist. 2004; (5):489-96.

82. Paridaens R, Dirix L, Beex L, Nooij M, Cufer T, Lohrisch C, *et al.* Promising results with exemestane in the first-line treatment of metastatic breast cancer: a randomized phase II EORTC trial with a tamoxifen control. Clinical breast cancer. 2000:S19-21.

83. Rose C, Vtoraya O, Pluzanska A, Davidson N, Gershanovich M, Thomas R, *et al.* An open randomised trial of second-line endocrine therapy in advanced breast cancer. comparison of the aromatase inhibitors letrozole and anastrozole. Eur J Cancer. 2003; (16):2318-27.

84. Campos SM, Guastalla JP, Subar M, Abreu P, Winer EP, Cameron DA. A comparative study of exemestane versus anastrozole in patients with postmenopausal breast cancer with visceral metastases. Clinical breast cancer. 2009; (1):39-44.

85. Howell A, Robertson JF, Quaresma Albano J, Aschermannova A, Mauriac L, Kleeberg UR, *et al.* Fulvestrant, formerly ICI 182,780, is as effective as anastrozole in postmenopausal women with advanced breast cancer progressing after prior endocrine treatment. J Clin Oncol. 2002; (16):3396-403.

86. Osborne CK, Pippen J, Jones SE, Parker LM, Ellis M, Come S, *et al.* Double-blind, randomized trial comparing the efficacy and tolerability of fulvestrant versus anastrozole in postmenopausal women with advanced breast cancer progressing on prior endocrine therapy: results of a North American trial. J Clin Oncol. 2002; (16):3386-95.

87. Robertson JF, Llombart-Cussac A, Rolski J, Feltl D, Dewar J, Macpherson E, *et al.* Activity of fulvestrant 500 mg versus anastrozole 1 mg as first-line treatment for advanced breast cancer: results from the FIRST study. J Clin Oncol. 2009; (27):4530-5.

88. Paridaens RJ, Dirix LY, Beex LV, Nooij M, Cameron DA, Cufer T, *et al.* Phase III study comparing exemestane with tamoxifen as first-line hormonal treatment of metastatic breast cancer in postmenopausal women: the European Organisation for Research and Treatment of Cancer Breast Cancer Cooperative Group. J Clin Oncol. 2008; (30):4883-90.

89. Chernozemsky I, Kalinov K, Tzekov H, Racheva M, Hristova S, Tomova A, *et al.* Randomized phase III trial of exemestane or tamoxifen in first-line hormonal treatment of postmenopausal women with metastatic breast cancer. Breast Cancer Res Treat. 2007; 106(Suppl 1):S119.

90. Mauriac L, Romieu G, Bines J. Activity of fulvestrant versus exemestane in advanced breast cancer patients with or without visceral metastases: data from the EFECT trial. Breast Cancer Res Treat. 2009; (1):69-75.

91. Kaufmann M, Bajetta E, Dirix LY, Fein LE, Jones SE, Zilembo N, *et al.* Exemestane is superior to megestrol acetate after tamoxifen failure in postmenopausal women with advanced breast cancer: Results of a phase III randomized double-blind trial. J Clin Oncol. 2000; (7):1399-411.

92. Parmar M, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. Statistics in Medicine. 1998; (17):2815-34.

93. Hind D, Ward S, De Nigris E, Simpson E, Carroll C, Wyld L. Hormonal therapies for early breast cancer: systematic review and economic evaluation. Health Technol Assess. 2007; 11(26):1-134.

94. Dranitsaris G, Leung P, Mather J, Oza A. Cost utility analysis of second line hormonal therapy in advanced breast cancer: a comparison of two aromatase inhibitors to megestrol acetate. Anticancer Drugs. 2000; 11:591-601.

95. Dranitsaris G, Verma S, Trudeau M. Cost utility analysis of first-line hormonal therapy in advanced breast cancer. Am J Clin Oncol. 2003; 26(3):289-96.

96. Marchetti M, Caruggi M, Colombo G. Cost utility and budget impact of third-generation aromatase inhibitors for advanced breast cancer: A literature based model analysis of cost in the Italian National Health Service. Clin Ther. 2004; 26(9).

97. Verma S, Rodchi A. Economic evaluation of antiaromatase agents in the second line treatment of metastatic breast cancer. Support Care Cancer. 2003; 11:728-34.

98. Hastings V, Amonkar M, Lykopoulos K, Diaz J, Johnston S. Indirect comparison of the cost-effectiveness of letrozole plus lapatinib (LET+LAP) versus anastrozole plus trastuzumab (ANA+TZ) as first line treatment for postmenopausal women with HER2+ and HR+ metastatic breast cancer (MBC) from the U.K. National Health Service (NHS) perspective ASCO 2010; Abstract 1035 (poster presentation).

99. National Institute for Health and Clinical Excellence. Guide to the methods of technology appraisal. London: National Institute for Health and Clinical Excellence; 2008; Available from: http://www.nice.org.uk/pdf/TAP_Methods.pdf (Accessed: 26/01/2010).

100. Curtis L. Unit Costs of Health and Social Care 2008. Personal Social Services Research Unit, University of Kent; 2008; Available from: http://www.pssru.ac.uk/uc/uc2008contents.htm (Accessed: 25/08/2010).

101. Department of Health. NHS Reference Costs 2008-2009. 2010: Available from: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_111591 (Accessed: 25/08/2010).

102. Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, *et al*. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. J Clin Oncol. 1993; 11(3):570-9.

103. Dobrez D, Cella D, Pickard S, Lai J-S, Nickolov A. Estimation of patient preference-based utility weights from the Functional Assessment of Cancer Therapy—General. Value in Health. 2007; 10(4):266-72.

104. Delea T, Sofrygin O, Amonkar M. Patient preference based utility weights from the Functional Assessment of Cancer Therapy - General in women with hormone receptor positive metastatic breast cancer receiving lapatinib plus letrozole or placebo plus letrozole. ISPOR 2010

105. Lloyd A, Nafees B, Narewska J, Dewilde S, Watkins J. Health state utilities for metastatic breast cancer. Br J Cancer. 2006; 95:683-90.

106. Oxford Outcomes. Utility measurement study for patients with metastatic breast cancer treated with tyverb. 2009.

107. Doyle S, Lloyd A, Walker M. Health state utility scores in advanced non-small cell lung cancer. Lung Cancer. 2009; 62(3):374-80.

108. Ward S, Pilgrim H, Hind D. Trastuzumab for the treatment of primary breast cancer in HER2 positive women:A Single Technology Appraisal. ERG Report (ScHARR). 2006.

109. Curtis L. Unit Costs of Health and Social Care 2009. Personal Social Services Research Unit, University of Kent; 2010; Available from: http://www.pssru.ac.uk/uc/uc2009contents.htm (Accessed: 25/08/2010).

110. Cooper N, Abrams K, Sutton A, Turner D, Lambert P. A Bayesian approach to Markov modelling in cost effectiveness analyses: application to taxane use in advanced brest cancer. J Roy Statistical Society. 2003; 166:389-405.

111. Regional Drugs and Therapeutics Centre. The use of bevacizumab in the manaagment of metastatic breast cancer 2007: Available from: http://www.nyrdtc.nhs.uk/docs/eva/Bevacizumab%20MBC%20RDTC%20FINAL.pdf (Accessed: 25/08/2010).

112. National Institute for Health and Clinical Excellence. Lapatinib for the treatment of previously treated women with advanced, metastatic or recurrent breast cancer (final scope). 2007: Available from: http://guidance.nice.org.uk/nicemedia/live/11731/35243/35243.pdf (Accessed: 25/08/2010).

113. CRD and CHE Technology Assessment Group. Trastuzumab for the treatment of HER2 positive metastatic adenocarcinoma of the stomach or gastro-oesophageal junction: A Single Technology Appraisal. 2006; Available from: http://www.nice.org.uk/nicemedia/live/12152/49646/49646.pdf (Accessed: 25/08/2010).

114. Dolan P, Gudex C, Kind P, Williams A. The time trade-off method: results from a general population survey. Health Econ. 1996; 5:141-54.

115. Swain L. OPCS-4.5 Chemotherapy Regimens List & Guidance for Clinical Coders Version 2.0 dated 10th March 2010. 2010 Available from: www.connectingforhealth.nhs.uk/clinicalcoding/codingstandards/opcs4/downloads/ (Accessed: 25/08/2010).

9 APPENDICES

Appendix 1: Literature search strategies

MEDLINE 1950 to April Week 4 2010

Searches		Results
1	(lapatinib or tykerb or tyverb or lapatinib ditosylate).af.	456
2	(trastuzumab or herceptin).af.	340
3	(letrozole or femara or anastrozole or arimidex or exemestane or aromasin).af.	2106
4	exp Aromatase Inhibitors/	4804
5	aromatase inhibitor$.tw.	3518
6	1 or 2	3627
7	3 or 5	4323
8	6 and 7	121
9	exp Breast Neoplasms/	172296
10	(breast$ adj5 (neoplasm$ or cancer$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or dcis or ductal or infiltrat$ or intraductal$ or lobular or medullary)).mp.	203286
11	9 or 10	203391
12	8 and 11	115

EMBASE 1980 to 2010 Week 18

Searches		Results
1	(lapatinib or tykerb or tyverb or lapatinib ditosylate).af.	2435
2	(trastuzumab or herceptin).af.	10741
3	(letrozole or femara or anastrozole or arimidex or exemestane or aromasin).af.	5962
4	Aromatase Inhibitors.mp. or exp aromatase inhibitor/	11914
5	1 or 2	11728
6	3 or 4	11959
7	5 and 6	1472
8	(breast$ adj5 (neoplasm$ or cancer$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or sarcoma$ or dcis or ductal or infiltrat$ or intraductal$ or lobular or medullary)).mp.	195921
9	exp breast cancer/	167175
10	8 or 9	197019
11	7 and 10	1378
12	limit 11 to (human and english language)	1195

The Cochrane Library Issue 4, 2010

Searches		Results
1	(lapatinib or tykerb or tyverb or lapatinib ditosylate or trastuzumab or herceptin or letrozole or femara or anastrozole or arimidex or exemestane or aromasin or aromatase inhibitor*)	1196
2	MeSH descriptor **Breast Neoplasms** explode all trees	6865
3	(breast cancer* or breast neoplasm* or breast tumor* or breast tumour* or breast carcinoma*)	14097
4	(#2 OR #3)	14097
5	(#1 AND #4)	932

Search Results by each database in the Cochrane Library	Results
Cochrane Database of Systematic Reviews	24
Database of Abstracts of Reviews of Effects (DARE)	26
Cochrane Central Register of Controlled Trials (CENTRAL)	757
Cochrane Methodology Register (CMR)	11
Health Technology Assessment Database	34
NHS Economic Evaluation Database (NHS EED)	79

The search strategy for the Cochrane Library is broader than MEDLINE or EMBASE, combining all the drug-related free text words with breast cancer (both using MeSH descriptor Breast Neoplasms and free text words) to identify relevant reviews and particularly economic evaluations in the area.

All databases

Total number of results from all databases: 2228

After electronic removal of duplicates: 2202

After manual removal of duplicates: 2069

Appendix 2: Table of excluded studies with rationale

Excluded studies from clinical review

The following citations were excluded by the AG at screening stage 2:

Study	Reason for exclusion
Langer 2001[43]	Data Physicians Query apparently relating to TAnDEM [52]
Maung and OShaughnessy 2004[48]	LAP monotherapy, not LAP+AI
Morris and Modi 2008[47]	TRA+ tanespimycin, not TRA+AI
Novartis 2006 [46]	Data Physicians Query apparently relating to eLEcTRA[54]
Piccart-Gebhart and Coleman 2002[44]	Data Physicians Query apparently relating to TAnDEM [52]
Ranganathan et al 2007 [49]	News article reporting on conference presentation for TAnDEM [52]
Stein 2004 [45]	Data Physicians Query apparently relating to EGF30008[51]

Studies included in the indirect comparisons analyses performed by the manufacturers

The following studies were included in the GlaxoSmithKline[22] and/or Roche[39] submissions but excluded by the AG:

Study	Submission(s) included in	Reason for exclusion
TARGET(outside North America)[70]	GlaxoSmithKline[22] and Roche[39]	ANA vs. TAM, not HER2+
PO25[76]	GlaxoSmithKline[22] and Roche[39]	LET vs TAM, not limited to first-line, not HER2+
TARGET (North America) [71]	GlaxoSmithKline[22] and Roche[39]	ANA vs. TAM, not HER2+
Rose et al 2003[83]	Roche[39]	ANA vs LET, second-line, not HER2+
Campos et al 2009[84]	Roche[39]	ANA vs EXE, not limited to first-line, not HER2+
020[85]	Roche[39]	ANA vs fulvestrant, not limited to first-line, not HER2+
021[86]	Roche[39]	ANA vs fulvestrant, not limited to first-line, not HER2+
FIRST[87]	Roche[39]	ANA vs fulvestrant, HER2 status not clear
Jonat et al 1996 [73]	Roche[39]	ANA vs megestrol acetate, not clear if first-line, not HER2+
Buzdar et al 1997[72]	Roche[39]	ANA vs megestrol acetate, not limited to first-line, not HER2+
Dombernowsky et al 1998[79]	Roche[39]	LET vs megestrol acetate, second-line, not HER2+
Buzdar et al 2001[78]	Roche[39]	LET vs megestrol acetate, not limited to first-line, not HER2+
EORTC[88]	Roche[39]	EXE vs TAM, not HER2+
Chernozemsky et al 2007[89]	Roche[39]	EXE vs TAM, not HER2+
EFFECT[90]	Roche[39]	EXE vs fulvestrant, not clear if first-line, not HER2+
Kaufmann et al 2000 [91]	Roche[39]	EXE vs megestrol acetate, not clear if first-line, not HER2+

Appendix 3: Participant characteristics of EGF30008, TAnDEM and eLEcTRA

Table 32 Participant characteristics: EGF30008

	HER2+				ITT			
Demographic or clinical characteristic of patients	**LET (n=108)**		**LAP+LET (n=11)**		**LET (n=644)**		**LAP+LET (n=642)**	
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Age, years[a]								
Median		59		60		63		62
Range		45-87		44-85		35-95		31-94
ECOG performance status[a]								
0	51	47	59	53	349	54	370	58
≥1	57	53	51	46	286	44	268	42
Hormone receptor status[a]								
ER/PgR positive	69	64	74	67	414	64	420	65
ER positive/PgR negative	20	19	19	17	90	14	91	14
Disease stage								
IIIB or IIIC	7	6	5	5	30	5	25	4
IV	101	94	106	95	613	95	616	96
Number of metastatic sites[a]								
Median		2		2		2		2
Range		1-7		1-7		0-7		0-7
Disease stage								
Bone only	18	17	16	14	85	13	94	15
Visceral or soft tissue	90	83	95	86	559	87	548	85
Liver	37	34	33	30	171	27	146	23
Lung	40	37	43	39	242	38	248	39
Lymph node	43	40	57	51	304	47	312	49
Soft tissue	31	29	35	32	218	34	212	33
Other	18	17	19	17	127	20	125	19
Previous therapy								
Endocrine[a]	62	57	60	54	317	49	313	49
Tamoxifen or toremifene only	60	56	59	53	302	47	300	47
Aromatase inhibitor only	1	<1	1	<1	3	<1	5	<1
Chemotherapy[a]	51	47	61	55	280	43	281	44
Anthracycline only	38	35	41	37	172	27	171	27
Anthracyclines and taxanes	9	8	9	8	41	6	42	7
Other	4	4	11	10	66	10	68	11
Biologic therapy (any)	1	<1	1	<1	1	<1	2	<1
Interval since prior adjuvant endocrine therapy[a]								
≥ 6 months or no prior therapy	67	62	73	66	487	76	501	78
< 6 months	41	38	38	34	157	24	141	22

ECOG=Eastern Cooperative Oncology Group; ER=oestrogen receptor; HER2=human epidermal growth factor receptor 2; ITT=intent to treat; PgR=progesterone receptor.

[a] Indicates prespecified baseline prognostic factors used in predefined Cox regression model. Additional factors included treatment, disease-free interval, and serum HER2 (extracellular domain) at baseline.

Data taken from Johnston et al 2009[51]

Table 33 Participant characteristics: TAnDEM

Demographic or clinical characteristic of patients	HER2+			
	ANA (n=104)		TRA+ANA (n=103)	
	N	%	N	%
Age, years				
Median	54		56	
Range	27-77		31-85	
Hormone receptor status[a]				
Primary and/or metastatic lesion ER+ and/or PgR+ (local)	104	100	103	100
Primary and/or metastatic ER+ and/or PgR+ (central)	73	70.2	77	74.8
Time from diagnosis of metastatic disease, months				
Median	1.2		1.6[b]	
Range	0.3-19.3		0.3-67.1	
Number of metastatic sites per patient				
Median	2		2	
Range	1-5		1-5	
Number of lesions per patient				
Median	4		4	
Range	1-13		1-14	
Site of therapy				
Lung	48	46.2	43	41.7
Liver	29	27.9	33	32.0
Bone	53	51.0	64	62.1
Soft tissue	44	42.3	46	44.7
Other	65	62.5	72	69.9
Previous therapy				
Hormonal	69	66.3	62	60.2
Tamoxifen for metastatic disease	3	2.9	5	4.9
Chemotherapy	62	59.6	55	53.4
Anthracycline	53	51.0	46	44.7
Bisphosphonate	27	26.0	28	27.2
LVEF, %				
Median	63		62	
Range	51-89		50-82	

ER=estrogen receptor-positive; LVEF=left ventricular ejection fraction; PgR+=progesterone receptor-positive

[a] Hormone receptor status determined locally as defined by institutional criteria

[b] n =101; two patients were not considered as having metastatic disease but, instead, were considered as having local recurrence

Data taken from Kaufman et al 2009[52]

Table 34 Participant characteristics: eLEcTRA

	LET (HER2+) (n=31)	TRA+LET (HER2+) (n=26)	LET (HER2-) (n=35)
	%	%	%
Age, years			
Median	61	61.5	70
Range	47-88	39-87	45-81
ECOG Performance status			
0	55	31	54
1	45	69	43
Not reported	0	0	3
Hormone receptor status			
ER+ and/or PgR+	97	100	100
ER+ and/or PgR unknown	3	0	0
Time from primary diagnosis to randomisation, months			
Median	30	3	2
Range	0-75	0-486	0-292
Site of metastases			
Locoregional	29	46	43
Lung	13	15	26
Liver	39	19	23
Bone	61	58	57
Bone only	23	23	17
Soft tissue	36	31	23
Other	3	8	0
Previous and second-line therapy			
Any adjuvant therapy	71	42	31
Adjuvant endocrine therapy	65	31	26
Second line trastuzumab (after assigned treatment)	52	31	N/A

ECOG=Eastern Cooperative Oncology Group; ER+=oestrogen receptor-positive; PgR+=progesterone receptor-positive
Data taken from Roche MS[39]

Appendix 4: Findings from the indirect comparisons analyses performed by the manufacturers

Table 35 Adjusted indirect comparisons analysis conducted by GlaxoSmithKline: median OS

	TRA+ANA	LET	ANA	TAM
LAP+LET	0.85 (0.47, 1.54)	0.77 (0.52, 1.14)	0.71 (0.45, 1.14)	0.74 (0.49, 1.12)
TRA+ANA		0.90 (0.60, 1.36)	0.84 (0.59, 1.19)	0.87 (0.59, 1.27)
LET			0.93 (0.76,1.15)	0.96 (0.84, 1.09)
ANA				1.03 (0.88, 1.22)

Bold = significant difference in terms of OS; Hazard ratio < 1 indicates greater likelihood of better response on treatment versus comparator

Table 36 Adjusted indirect comparisons analysis conducted by GlaxoSmithKline: median PFS/TTP

	TRA+ANA	LET	ANA	TAM
LAP+LET	0.89 (0.54,1.47)	**0.65 (0.47, 0.89)**	**0.53 (0.36, 0.80)**	**0.45 (0.32, 0.65)**
TRA+ANA		0.73 (0.50, 1.07)	**0.60 (0.45, 0.81)**	**0.51 (0.36, 0.71)**
LET			0.82 (0.65,1.04)	**0.70 (0.60, 0.82)**
ANA				0.85 (0.71, 1.01)

Bold = significant difference in terms of TTP/PFS; Hazard ratio < 1 indicates greater likelihood of better response on treatment versus comparator

Table 37 Network analysis conducted by Roche: median OS – base case (cross-over adjustment for TAnDEM applied)

	LAP+LET	LET	EXE	ANA	Megastrol acetate	TAM
TRA+ANA	1.01 [0.58;1.75]	0.75 [0.51;1.09]	0.73 [0.48;1.11]	0.73 [0.51;1.04]	**0.60 [0.41;0.89]**	0.74 [0.51;1.09]
LAP+LET		0.74 [0.50;1.10]	0.72 [0.46;1.14]	0.72 [0.48;1.10]	**0.60 [0.39;0.91]**	0.74 [0.49;1.12]
LET			0.98 [0.78;1.22]	0.98 [0.85;1.12]	**0.81 [0.70;0.94]**	1.00 [0.87;1.14]
EXE				1.00 [0.80;1.25]	0.83 [0.68;1.02]	1.02 [0.82;1.26]
ANA					**0.83 [0.71;0.97]**	1.02 [0.89;1.16]
Megestrol acetate						**1.23 [1.04;1.45]**
Bold = significant difference in terms of OS; Hazard ratio < 1 indicates greater likelihood of better response on treatment versus comparator						

Table 38 Network analysis conducted by Roche: median OS – final sensitivity analysis, AIs as a class (cross-over adjustment for TAnDEM applied)

	LAP+AI	AI
TRA+AI	0.78 [0.52;1.18]	0.55 [0.42;0.74]
LAP+AI		0.71 [0.53;0.95]
*Head-to-head comparison; **Bold** = significant difference in terms of OS; Hazard ratio < 1 indicates greater likelihood of better response on treatment versus comparator		

Table 39 Network analysis conducted by Roche: median OS (sensitivity analysis using original unadjusted findings for TAnDEM) – initial analysis

	LAP+LET	LET	EXE	ANA	Megastrol acetate	TAM
TRA+ANA	1.16 [0.67;2.01]	0.86 [0.59;1.25]	0.84 [0.55;1.28]	0.84 [0.59;1.19]	0.70 [0.47;1.02]	0.85 [0.58;1.25]
LAP+LET		0.74 [0.50;1.10]	0.72 [0.46;1.14]	0.72 [0.48;1.10]	**0.60 [0.39;0.91]**	0.74 [0.49;1.12]
LET			0.98 [0.78;1.22]	0.98 [0.85;1.12]	**0.81 [0.70;0.94]**	1.00 [0.87;1.14]
EXE				1.00 [0.80;1.25]	0.83 [0.68;1.02]	1.02 [0.82;1.26]
ANA					**0.83 [0.71;0.97]**	1.02 [0.89;1.16]
Megestrol acetate						**1.23 [1.04;1.45]**
Bold = significant difference in terms of OS; Hazard ratio < 1 indicates greater likelihood of better response on treatment versus comparator						

Table 40 Network analysis conducted by Roche: median OS (sensitivity analysis using original unadjusted findings for TAnDEM) – final sensitivity analysis, AIs as a class

	LAP+AI	AI
TRA+AI	1.13 [0.67;1.92]	0.84 [0.59;1.19]
LAP+AI		0.74 [0.50;1.10]
Bold = significant difference in terms of OS; Hazard ratio < 1 indicates greater likelihood of better response on treatment versus comparator		

Table 41 Network analysis conducted by Roche: median PFS– base case

	LAP+LET	LET	EXE	ANA	Megastrol acetate	TAM
TRA+ANA	0.87 [0.48;1.55]	0.62 [0.37;1.02]	**0.58 [0.39;0.85]**	**0.55 [0.41;0.74]**	**0.49 [0.32;0.76]**	**0.50 [0.36;0.69]**
LAP+LET		**0.71 [0.53;0.96]**	0.67 [0.38;1.18]	0.64 [0.38;1.06]	**0.57 [0.38;0.84]**	**0.58 [0.34;0.98]**
LET			0.94 [0.58;1.53]	0.90 [0.60;1.36]	0.80 [0.62;1.03]	0.82 [0.53;1.26]
EXE				0.96 [0.74;1.24]	0.85 [0.56;1.29]	0.87 [0.70;1.08]
ANA					0.89 [0.64;1.23]	0.91 [0.79;1.04]
Megestrol acetate						1.02 [0.72;1.45]

Bold = significant difference in terms of OS; Hazard ratio < 1 indicates greater likelihood of better response on treatment versus comparator

Table 42 Network analysis conducted by Roche: median PFS– final sensitivity analysis, AIs as a class

	LAP+AI	AI
TRA+AI	0.78 [0.52;1.18]	0.55 [0.42;0.74]
LAP+AI		0.71 [0.53;0.95]

Bold = significant difference in terms of OS; Hazard ratio < 1 indicates greater likelihood of better response on treatment versus comparator

Table 43 Summary of the network analysis conducted by Roche: median PFS/TTP – initial analysis

	TRA+LET	LAP+LET	LET	EXE	ANA	Megastrol acetate	TAM
TRA+ANA	0.97 [0.46;2.02]	0.91 [0.58;1.44]	**0.65 [0.46;0.92]**	**0.62 [0.43;0.88]**	**0.55 [0.41;0.74]**	**0.54 [0.38;0.77]**	**0.49 [0.36;0.68]**
TRA+LET		0.94 [0.46;1.94]	0.67 [0.35;1.29]	0.64 [0.32;1.25]	0.57 [0.29;1.13]	0.56 [0.29;1.09]	**0.51 [0.26;0.99]**
LAP+LET			**0.71 [0.53;0.96]**	**0.68 [0.48;0.96]**	**0.61 [0.43;0.86]**	**0.60 [0.43;0.83]**	**0.54 [0.39;0.75]**
LET				0.95 [0.80;1.13]	0.85 [0.72;1.02]	**0.84 [0.73;0.96]**	**0.76 [0.67;0.87]**
EXE					0.90 [0.74;1.10]	0.88 [0.76;1.02]	**0.80 [0.68;0.94]**
ANA						0.98 [0.82;1.18]	0.89 [0.78;1.02]
Megestrol acetate							0.91 [0.78;1.06]

Bold = significant difference in terms of OS; Hazard ratio < 1 indicates greater likelihood of better response on treatment versus comparator

Table 44 Summary of the network analysis conducted by Roche, AIs assumed to have class-effect: median PFS/TTP – final sensitivity analysis, AIs as a class

	LAP+AI	AI
TRA+AI	0.81 [0.54;1.20]	0.57 [0.44;0.75]
LAP+AI		0.71 [0.53;0.95]

Bold = significant difference in terms of OS; Hazard ratio < 1 indicates greater likelihood of better response on treatment versus comparator

NB. Roche performed and presented a number of other sensitivity analyses in their report which have not been reproduced here. In addition to median, mean values for all outcomes were also presented by Roche