Evidence Review Group's Report

Title: *Trastuzumab for the treatment of HER2 positive metastatic adenocarcinoma of the stomach or gastro-oesophageal junction.*

Produced by	CRD and CHE Technology Assessment Group
Authors	Gill Norman, Research Fellow, CRD Stephen Rice, Research Fellow, CRD Eldon Spackman, Research Fellow, CHE Lisa Stirk, Information Specialist, CRD Anthony Danso-Appiah, Research Fellow, CRD Dong Suh, Visiting Research Fellow, CHE Stephen Palmer, Senior Research Fellow CHE Alison Eastwood, Senior Research Fellow, CRD
Correspondence to	Dr Gill Norman Centre for Reviews and Dissemination University of York York YO10 5DD
Date completed	10/05/2010

Source of funding: This report was commissioned by the NIHR HTA Programme as project number HTA 08/219.

Declared competing interests of the authors
None

Acknowledgements
Daniel Swinson, Medical Oncologist, St James's Institute of Oncology,St James's University Hospital, Leeds

Matthew Seymour, Professor of Gastrointestinal Cancer Medicine, Cancer Research UK Clinical Centre, University of Leeds.

Ralph Crott, Senior Research Fellow, CRD, University of York, York, YO10 5DD

Gerry Richardson, Senior Research Fellow, CHE, University of York, York, YO10 5DD

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: *CRD and CHE Technology Assessment Group. Trastuzumab for the treatment of HER2 positive metastatic adenocarcinoma of the stomach or gastro-oesophageal junction.*

Contributions of authors

Gill Norman wrote the clinical effectiveness sections of the report, Eldon Spackman and Stephen Rice wrote the economic sections of the report and performed the economic modelling, Lisa Stirk wrote the sections of the report dealing with search strategies and provided information support, Anthony Danso-Appiah produced tables and commented on the report, Dong Suh reviewed and commented on the report, Stephen Palmer and Alison Eastwood managed the cost-effectiveness and clinical effectiveness parts of the project and reviewed and commented on the report.

Contents

List of Abbreviations

AE	Adverse events
aGC	Advanced gastric cancer
BNF	British National Formulary
CEAC	Cost-effectiveness acceptability curve
CSR	Clinical Study Report
CVAD	Central venous access device
ECF	Epirubicin, cisplatin, 5-FU
ECX	Epirubicin, cisplatin, capecitabine
EMEA	European Medicines Evaluation Agency
EOF	Epirubicin, oxaliplatin, 5-FU
EORTC	European Organisation for Research and Treatment of Cancer
EOX	Epirubicin, oxaliplatin, capecitabine
EQ-5D	EuroQol 5 Dimensions
ERG	Evidence Review Group
F	5-FU
FAS	Full analysis set (all randomised patients who received study medication at least once
GOJ	Gastroesophageal junction
HCF	Trastuzumab, cisplatin, 5-FU
HCX	Trastuzumab, cisplatin, capecitabine
H	Trastuzumab
HF	Trastuzumab, 5-FU

HUI	Health Utility Index
HX	Trastuzumab, capecitabine
ICER	Incremental cost-effectiveness ratio
IDMC	Independent data monitoring committee
ITT	Intention-to-treat
MS	Manufacturer’s submission-
mGC	Metastatic gastric cancer
OS	Overall survival
PFS	Progression free survival
PD	Progressive disease
PP	Per protocol
PSA	Probabilistic sensitivity analysis
PSSRU	Personal Social Services Research Unit
QALY	Quality-adjusted life year
QoL	Quality of life
RCT	Randomised controlled trial
SA	Sensitivity analysis
SPC	Summary of Product Characteristics
X	Capecitabine

1 SUMMARY

1.1 Scope of the submission

This report presents the ERG's assessment of the manufacturer's (Roche) submission (MS) to NICE on the use of trastuzumab (Herceptin®) for the treatment of HER2 positive advanced gastric cancer (aGC). The MS included a de-novo economic evaluation based on the eligibility of the application for end-of life status.

1.2 Summary of submitted clinical effectiveness evidence

The MS focused on the direct evidence from the ToGA trial.[1] This was a phase III randomised controlled trial (RCT) which compared a doublet regimen of cisplatin plus a fluoropyrimidine (capecitabine or 5-fluorouracil (5-FU) (CX/F)) alone or in combination with trastuzumab (HCX/F) in patients with HER2 positive advanced adenocarcinoma of the stomach or gastroeospohageal junction (GOJ). The choice of fluoropyrimidine was at the discretion of the investigator; 87% of patients received capecitabine and 13% 5-FU. This is the only trial of direct relevance to the licensed population of HER2 positive metastatic gastric cancer (mGC). The licensed population is therefore slightly more restrictive than the population defined in the NICE scope which is aGC patients. The primary outcome was overall survival (OS). All patients given at least one dose of treatment constituted the full analysis set (FAS). A subgroup of patients from this trial with metastatic tumours which were IHC2+/FISH+ or IHC3+ constituted the population for which EMEA licensed trastuzumab. Outcome data are summarised in Table 1.

Analyses of the FAS, and of the EMEA approved subgroup of patients who were highly HER2 positive (IHC2+/FISH+ or IHC3+), were reported for the ToGA trial. The hazard ratio (HR) for OS in the EMEA subgroup (74% of the FAS population) was 0.65 (95% CI: 0.51 to 0.83) which corresponded to median survival of 16 months for the HCX/F group versus 11.8 months for the CX/F group. **(confidential information removed.)**

Table 1. Summary of outcome data reported in the MS for the EMEA subgroup (IHC2+/FISH + or IHC3+; 74% of the FAS population) of the ToGA trial

Outcome	HCX/F	CX/F	Statistical results
OS (median)	16.0 months	11.8 months	HR 0.65 (95% CI: 0.51 to 0.83)
PFS (median)	7.6 months	5.5 months	HR 0.64; (95% CI: 0.51 to 0.79)
Response Rate (%)*	47.3	34.5	OR 1.70 (95% CI 1.22 to 2.38)
QoL	• Graphical presentation of EORTC data only: ERG unable to form conclusions as to differences between the groups.		
Adverse events*	• Statistically significantly more grade 1 and 2 adverse events (multiple categories) in HCX/F group. • Statistically significantly more asymptomatic LVEF reductions in HCX/F group did not translate into increased symptomatic cardiac events. • No statistically significant differences in Grade 3 or 4 events.		

*Data for FAS population LVEF = Left ventricular ejection fraction

For the EMEA population, progression free survival (PFS) was also improved in the HCX/F group compared to the CX/F group (HR 0.64; 95% CI: 0.51 to 0.79) which corresponded to a 2.1 month difference in time to progression or death. In the FAS population there was a statistically significantly higher overall response rate in the HCX/F group (47.3% versus 34.5%; odds ratio 1.70, 95% CI: 1.22 to 2.38, p = 0.002).

No statistical test results were reported for QoL measures; it was reported that both groups showed improvements in QoL, (assessed using the EORTC-QLQ C30 and the gastric module ST022)[2] over the course of treatment, which endured for the duration of PFS.[3] It was therefore contended that improved QoL is longer in the HCX/F arm than in the CX/F arm, corresponding to the increased duration of PFS. Graphical presentations of the FAS data were supplied in support of this statement, while equivalent presentations of the EMEA data were supplied following a request from the ERG. Given the lack of statistical analysis, which accorded with the ToGA protocol, it was not possible to assess whether there was a statistically significant difference in QoL changes between the groups.

Adverse events data for the FAS population showed evidence of increased grade1 and 2 adverse events in the HCX/F arm compared to the CX/F arm, but no excess of grade 3 or 4 events was found. The significant differences between the arms were distributed across the event categories, and were attributed in the MS to longer duration of chemotherapy in the HCX/F arm. While there were statistically significant increases in asymptomatic reduction of LVEF in the HCX/F arm, there was no corresponding increase in symptomatic cardiac adverse events of any type.

The comparison between HCX/F and CX/F does not represent a comparison with the current UK standard chemotherapy for optimally fit patients, which is triplet therapy comprising an anthracycline (epirubicin (E)), a platinum agent (cisplatin or oxaliplatin(C or O)) and a fluoropyrimidine (capecitabine or 5-FU (X or F)). Accordingly the manufacturer assessed the possibility of performing a network meta-analysis to assess the relative efficacy of HCX/F and the four possible triplet regimens investigated in the REAL-2 trial.[4] Of necessity, this involved trials conducted in populations with mixed HER2 status.

The manufacturer concluded, correctly in the view of the ERG, that it was impossible to construct a meaningful network meta-analysis of efficacy in comparable population groups. Accordingly they presented a narrative synthesis of evidence from relevant trials and a meta-analysis.[4-8] This discounted the evidence from the meta-analysis showing superiority of ECF over CF and contended that the absence of individual trial evidence showing a statistically significant benefit of ECX/F over CX/F represented evidence that there was no such benefit. The ERG did not consider this argument to be convincing. The primary justification for questioning the pooled estimate of the Wagner meta-analysis[8] was the inclusion of a trial which used a comparator of mitomycin plus CF rather than CF alone;[9] thus the pooled estimate is likely to be conservative with respect to ECF. Given that ECX is at least as effective as ECF,[4] and may be more effective,[10] ECX may also be more effective than CX.

1.3 Summary of submitted cost effectiveness evidence

No previously published cost-effectiveness studies of trastuzumab in HER2 positive patients with metastatic gastric cancer were identified by the

manufacturer. Therefore the manufacturer's de novo economic evaluation formed the basis of the submitted economic evidence. The evaluation included two separate trastuzumab regimens in combination with either cisplatin and capecitabine (HCX) or cisplatin and 5-FU (HCF). The trastuzumab regimens were compared with three other triplet regimens containing epirubicin in combination with either: cisplatin and capecitabine (ECX); oxaliplatin and capecitabine (EOX) or cisplatin and 5-FU (ECF). The comparator regimens included by the manufacturer were reported to be based on the final NICE scope and routine NHS practice. The patient population reflected the EMEA licensed subgroup of patients from the ToGA trial which was conducted in patients with mGC whose tumours have HER2 overexpression (defined as IHC2+/FISH+ or IHC3+).

The manufacturer's cost-effectiveness analysis was based on a simple three state cohort model (progression free, disease progression and death). The model was used to estimate PFS and OS for each of the alternative regimens. Quality of life was quantified by applying utility weights to the separate model states in order to estimate quality-adjusted life years (QALYs). Costs were assessed from an NHS perspective and incorporated the acquisition and monitoring costs of the alternative regimens, HER2 testing, adverse events and other supportive care costs associated with the management of progression-free and progressive disease. An 8-year time horizon was employed and was considered to represent a lifetime analysis. Both one-way and probabilistic sensitivity analyses (PSA) were undertaken.

In the absence of direct evidence comparing all five regimens, the manufacturer combined the results from the ToGA trial with a series of assumptions, based on their narrative synthesis of evidence from relevant trials, a meta-analysis and expert opinion in order to determine the treatment effectiveness of the alternative regimens. The EMEA subgroup results from the ToGA trial were used to estimate parameters for two of the regimens (HCX and ECX). Data from the HCX/F arm of the ToGA trial was assumed to be equivalent to the HCX regimen in the model given that the majority of patients (87%) randomised to trastuzumab also received capecitabine. The

CX/CF arm of the ToGA study was assumed by the manufacturer to be equivalent to the ECX regimen in the model and was justified by: (i) the absence of individual trial evidence showing a statistically significant benefit of ECX/F over CX/F; (ii) the higher dose of CX/F used in ToGA compared to routine practice and (iii) the proportion of patients in the CX/F arm who received capecitabine. The effectiveness of these two regimens applied in the model was then based on a statistical extrapolation of the PFS and OS patient-level data from the ToGA trial in order to estimate mean survival times.

The equivalent survival times for PFS and OS for each of the other comparators (HCF, EOX and ECF) were informed by linking these statistical extrapolations to a series of additional assumptions regarding the relative effectiveness of the other alternative regimens. To make this link the manufacturer made the following assumptions: (i) that PFS for HCF would be equivalent to HCX, but that OS would be less favourable for HCF due to the lower effectiveness of 5-FU compared to capecitabine; (ii) that PFS for EOX and ECF were equivalent to ECX (which, itself, was assumed to be equivalent to the CX/CF arm of the ToGA trial); (iii) that OS for EOX was equivalent to ECX, and (iv) that OS for ECF would be less favourable than ECX (again due to the assumption that 5-FU is less effective than capecitabine). From these assumptions the manufacturer derived PFS and OS estimates for all 5 regimens. These estimates were then applied to the proportion of patients assumed to test HER2 positive based on IHC2+/FISH+ or IHC3+ (66% of mGC patients).

The manufacturer's results showed that the combination of HCX resulted in a mean gain of 4.8 months of life compared with ECX/EOX and that HCF resulted in a mean gain of 4.3 months of life compared with ECF. The results also showed approximately half of this extension in life resulted from an extension of PFS. Adjusting for the quality of life associated with the separate model states showed that HCX resulted in a mean gain of 0.25 QALYs compared to ECX/EOX, 0.31 QALYs compared to ECF and 0.07 QALYs compared to HCF.

The QALY estimates were then combined with the costs to calculate the incremental cost-effectiveness. The manufacturer compared the cost-effectiveness of all 5 regimens simultaneously and demonstrated that 3 of these regimens were ruled out either by dominance (EOX and ECF) or by extended dominance (HCF). Of the two remaining (non-dominated) regimens, ECX appeared both less costly and less effective than HCX. The incremental cost-effectiveness ratio (ICER) of HCX vs ECX was £53,010 per QALY. This ICER was subsequently altered by the manufacturer to £51,927 per QALY during the clarification stage, following minor corrections to their original costing estimates.

The manufacturer also presented the ICER results using three separate pair-wise comparisons (figures based on corrected estimates provided during the clarification stage): (i) HCX vs ECX = £51,927 per QALY; (ii) HCF vs ECF = £50,838 per QALY and (iii) HCX vs EOX = £40,711 per QALY. The presentation of a separate pair-wise comparison of the ICER for HCF vs ECF was justified on the basis that some patients may not be suitable for capecitabine making this specific comparison relevant. No justification was made for the separate presentation of the ICER of the HCX vs EOX comparison.

The results of the one-way sensitivity analyses appeared to demonstrate that the ICER estimates remained relatively robust. Across the majority of these analyses the ICER estimates only altered by a small amount. The maximum ICER reported in the one-way sensitivity analyses by the manufacturer for the comparison of HCX vs ECX was £56,175 per QALY. The results appeared most sensitive to the statistical extrapolation of OS, the inclusion of an additional treatment effect of trastuzumab on utility during PFS and the efficacy assumptions made on the relative effectiveness of particular comparators.

The results of the probabilistic sensitivity analysis (PSA) produced a 95% confidence interval around the ICER comparing HCX to ECX from £37,180 to £95,238 per QALY. As PSA simultaneously accounts for uncertainty in several parameters, much higher and lower ICER estimates are produced

than in one-way sensitivity analyses. However, extreme estimates only occur in a small percent of the simulated population. The PSA results were presented graphically using cost-effectiveness acceptability curves (CEACs). The associated probabilities of HCX being cost-effective at thresholds of £20,000 and £30,000 per QALY were not reported by the manufacturer. From the CEAC the probability that HCX is cost-effective at £30,000 is 0% and at £50,000 is 42%.

1.4 Commentary on the robustness of submitted evidence

1.4.1 Strengths

The ToGA trial is a well-conducted phase III RCT which directly compared trastuzumab in its licensed therapeutic combination with CX/F to CX/F alone; which comprises two of the key components of standard therapy in the UK.[1] The comparator regimen used is considered to be standard therapy in other non-UK settings. ToGA was appropriately randomised and protocol amendments took place on the advice of an independent data monitoring committee (IDMC).

The economic model structure was considered appropriate for the decision problem and the general approach employed by the manufacturer to estimate lifetime cost-effectiveness met the requirements of the NICE reference case approach. Both one-way and PSA were used to reflect uncertainty in the model inputs and assumptions and these were informative in exploring the robustness of the results and identifying potential key drivers of cost-effectiveness. The ERG also acknowledges that the manufacturer provided detailed additional information in response to the clarification points which were central to key aspects of the ERG's review.

1.4.2 Weaknesses

There are two principal weaknesses of the MS. The first of these is that ToGA represents the only evidence of the efficacy of any treatment in the licensed population. The HER2 positive mGC population has not been identified within

previous trials and therefore the efficacy of standard triplet regimens (or indeed any therapy) in this particular group is not known. The comparator used in ToGA (CX/F) is not the standard UK treatment and, where employed in frailer patients, is used at lower doses. Indirect evidence must therefore be relied on to assess the efficacy of HCX/F compared with current standard treatment for fit patients (ECX or EOX). This requires the assumption that the HER2 positive population is equivalent to a mixed HER2 positive population, containing an unknown proportion of HER2 positive patients. It is known that the rate of HER2 positivity varies with histological subtype; whether the histology seen in the ToGA trial is representative of the UK population is not clear since ToGA was primarily conducted in non-European settings.

Accepting the assumption of population comparability, the MS attempted to create a network meta-analysis to compare HCX/F with ECX and thence with EOX. The manufacturer concluded that this was not possible. The original MS did not provide details of studies excluded from the systematic review process, so it was not possible to determine whether all relevant studies were identified. Information on excluded studies was requested from the manufacturer and was provided. From this additional information provided, it appeared that all relevant studies were included in the network. Given the evidence identified, the ERG's view is that the decision not to perform a network meta-analysis was correct. This does, however, mean that there is no comparison of HCX/F versus ECX in even the general aGC/mGC population.

The second major weakness of the MS was the approach of the narrative synthesis of relevant trials. This contained the argument that a meta-analysis of CF versus ECF regimens, which found an OS advantage for ECF,[8] should be disregarded in favour of the results of two small trials included in the analysis.[6, 7] Whilst there are valid reasons for questioning the approach of the meta-analysis (see section 4.2), the ERG considered that this evidence, which was likely to be conservative to ECF, could not be disregarded. The alternative approach of the MS involved the argument that, since two small RCTs did not show a statistically significant advantage of ECF over CF, this could be regarded as evidence of no advantage. An additional argument was

that the higher dose of CX/F employed in ToGA provided additional efficacy over that used in other trials, giving comparable efficacy to epirubicin-based triplet regimens. The manufacturer therefore contended that the CX/F comparator in ToGA could be considered equivalent to ECX/F (and hence EOX on the basis of the evidence from the REAL-2 trial[4]). The ERG did not consider this argument to be convincing.

Other issues included the fact that the direct evidence presented in the MS is based on a subgroup (the EMEA licensed population) of the single relevant trial (ToGA). This subgroup was defined as a result of advances in the understanding of HER2 testing and therefore has credibility as a distinct population. It is also the case that this population of patients with high HER2 expressing tumours constituted a clear majority of all patients in the ToGA trial (74%). Therefore, while the use of a subgroup as the basis for a submission is potentially problematic, the ERG does not consider it to be a cause for serious concern in this instance. The ERG also noted the fact that ToGA was an open-label trial; this was an inevitable consequence of the therapy assessed and the ethical problems attaching to the use of placebo infusions. The fact that outcome assessors were not blinded was also noted. However, the primary outcome was OS, meaning this is of less concern than would otherwise be the case.

From a cost-effectiveness perspective there were a number of additional potential weaknesses considered by the ERG. These stem largely from the lack of direct comparison of the different regimens incorporated in the economic analysis and the series of assumptions that were then necessary in order to estimate the incremental cost-effectiveness of the relevant regimens. Although the manufacturer undertook a detailed set of sensitivity analyses, several of the model assumptions were not incorporated. The ERG also considered that there were equally plausible assumptions and inputs for several of the key assumptions employed in the manufacturer's 'base-case' analysis. Although the one-way sensitivity analyses undertaken both by the manufacturer and the ERG appeared to demonstrate the ICER for HCX vs ECX remained relatively robust to changes in individual parameters, the

probabilistic sensitivity analysis resulted in a wider range of estimates. Given this level of uncertainty, and the number of assumptions required, scenario analyses could have been used to demonstrate the combined effect of other plausible assumptions and changes in influential parameters. The ERG considered this a weakness of the submission given the number of assumptions required in the model.

The most significant assumptions for which the ERG considered there to be equally plausible alternative estimates to those employed in the manufacturer's base-case analysis included: (i) the relative effectiveness estimates of particular comparators; (ii) the utility values applied during PFS; and (iii) the frequency of cardiac monitoring with trastuzumab and epirubicin. In addition to these assumptions, the ERG also considered that there was insufficient discussion of the logistical issues of undertaking HER2 testing in this population and whether the effectiveness results from the ToGA trial (where parallel testing using IHC and FISH tests was used) could be generalised without any loss in effect due to potential delays that could arise for IHC2+ patients based on the sequential testing approach included in the model.

The ERG undertook a series of alternative 'base-case' analyses to address these perceived weaknesses, varying the key assumptions and altering the cost of the testing strategy. The results of these analyses increased the ICER for the comparison of HCX vs EOX to between £66,982 and £71,636 per QALY.

1.4.3 Areas of uncertainty

As discussed in 1.4.2 it is unclear how HCX/F compares with the standard triplet therapy, particularly ECX or EOX, which HER2 positive patients would currently receive in the UK if they were considered able to withstand it. This uncertainty stems firstly from the lack of evidence as to the comparability of the HER2 positive population with the general mGC population, and secondly from the lack of a network of evidence which would permit the HCX/F versus ECX/EOX comparison in this general population.

It is unclear whether the effect of stopping ToGA early may have been to increase the estimate of effectiveness of HCX/F. There is some evidence to indicate that trials which are stopped early may produce higher estimates of effect for the outcome which triggers the trial termination.[11] However, ToGA was terminated in accordance with a stopping rule based on a planned interim analysis of the primary outcome recommended by an IDMC. These factors reduce the concern which should attach to the termination.

There are several additional areas of uncertainty which remain related to the cost-effectiveness analysis particularly in relation to the potential implications of sequential versus parallel testing for HER2 and the definition of HER2 positive itself. These uncertainties also relate to the overall decision problem and the scope of the appraisal. Since treatment with trastuzumab requires additional diagnostic testing compared to current practice in order to identify HER2 positive patients, the technology being evaluated is actually a combination of two elements: (i) a single diagnostic strategy, based on sequential testing using IHC and FISH tests, to determine HER2 status and (ii) alternative treatment strategies for HER2 positive patients. The diagnostic strategy used by the manufacturer to determine HER2 positivity is based on the EMEA subgroup and the respective license for trastuzumab. However, the diagnostic testing strategy used to determine HER2 status and thus treatment appropriateness will have an impact on the cost-effectiveness of treatment and it should be recognised that other thresholds of defining HER2 positivity and alternative diagnostic tests represent potential additional comparators to those considered here. That is, the single diagnostic strategy being used itself has not been demonstrated as being more cost-effective than other ways of defining HER2 positivity and eligibility for trastuzumab. Although the ERG recognises that such an evaluation probably lies outside the scope of this appraisal, it should be noted as an area of remaining uncertainty.

A related area of uncertainty which the ERG considers to be directly relevant to the scope of the appraisal relates to the potential impact of any delays that could be caused by using sequential rather than parallel testing with IHC and

FISH tests for identifying HER2 positive patients. The ERG considers that the generalisability of the effectiveness results of the ToGA trial to a decision problem based on a different approach to testing with IHC and FISH presents a potentially important source of uncertainty.

1.5 Key issues

As discussed in sections 1.4.2 and 1.4.3, there is no evidence other than the ToGA trial on the efficacy of any therapy in the HER2 positive mGC population.

The MS was not able to establish a network of clinical evidence for a comparison between HCX/F and current standard UK therapy in the general mGC population. The MS subsequently advanced the argument that the absence of individual trial evidence for a statistically significant survival benefit for ECX/F over CX/F constituted evidence for no benefit. This argument involved disregarding the evidence from a meta-analysis of three trials[8] and relying on the individual results of two small trials from this analysis. The ERG regarded this argument as poorly substantiated, and not conservative with respect to ECX/F.

2 BACKGROUND

2.1 Critique of manufacturer's description of underlying health problem

The MS provides a clear summary of the incidence of gastric cancer including the proportion of patients presenting with advanced disease which is treated palliatively.

The proportion of patients with HER2 positive mGC defined by IHC3+ or FISH+ and eligible for the ToGA trial was estimated to be 22.1%.[12] The proportion of patients with IHC2+/FISH+ or IHC3+ tumours (the EMEA licensed population) in the screening programme for the ToGA trial was 17.8%.[13] This is the figure used in the model presented in the MS. The ERG

noted that a figure of 16.6% is cited (MS p33), this appears to represent preliminary analyses. [14]

ToGA was a multinational trial with the great majority of centres located in Asia and Central and South America. The proportion of HER2 positivity in the ToGA trial screening population varies with tumour histology (intestinal 32.2%, diffuse 6.1%, mixed 20.4%)[12] and location (gastric versus GOJ).[15, 16] It is not known whether the distribution of histological subtypes in the ToGA population is representative of the UK mGC population.

2.2 Critique of manufacturer's overview of current service provision

The MS identifies that, in the UK, the most widely used regimen for mGC patients able to tolerate it is a fluoropyrimidine (with increasing clinician preference for capecitabine over 5-FU), a platinum agent (oxaliplatin is increasingly used in place of cisplatin) and epirubicin. The ERG's clinical advisors took the view that the percentage of patients given oxaliplatin is higher than the 6% indicated in the research conducted on behalf of the manufacturer (MS p32), but were otherwise in broad agreement with their overview. However, the MS advances the view that there is no difference in efficacy between doublet regimes consisting of a platinum agent combined with a fluoropyrimidine and triplet regimens which also include epirubicin. It further asserts that the higher dose CX/F regimen used in ToGA would have additional efficacy, increasing it's comparability to a triplet regimen. The ERG considers that this argument is poorly evidenced and could not be substantiated (see section 4.2.2 below).

The clinical effectiveness sections of the MS deal almost exclusively with the technology of trastuzumab for HER2 positive patients. However, current service provision does not include any HER2 testing procedures. If trastuzumab becomes available as a treatment option for those meeting the licensing definition of HER2 positivity, then IHC testing will be required for the whole mGC population, while FISH testing will be required for either the whole population, or for the subset of the population with an IHC2+ result. Based on

the ToGA trial screening population, 11.8% of the mGC population would have an IHC 2+ result, with 5.37% FISH- and therefore ineligible for trastuzumab.[13] The proportion of the population which would require a FISH test depends on whether the tests are conducted in parallel (as was the case in the ToGA trial) or sequentially, as the model presented in the MS assumes. HER2 testing in the mGC population would represent a major departure from current practice and may have the effect of delaying treatment for the approximately 82% (based on ToGA estimates) of the population who are HER2 negative according to EMEA criteria, and particularly for those 5.37% who are IHC2+/FISH-. These implications are further explored in sections 5 and 6.

3 Critique of manufacturer's definition of decision problem

3.1 Population

The NICE scope defines the population as patients with HER2 positive aGC. The population addressed in the MS accords with this. It should be noted that the licensing of trastuzumab is restricted to patients with HER2 positive mGC and this forms the basis for the economic model. Patients in the ToGA trial had inoperable HER2 positive aGC and had not received prior treatment for their advanced disease; the great majority of these had metastatic disease (3% had locally advanced disease). HER2 positive patients were initially defined as those whose tumours were IHC3+ or FISH+.

Advances in understanding of HER2 testing during the ToGA trial resulted in a protocol amendment and a narrower definition of HER2 positivity being adopted. This group with IHC2+/FISH+ or IHC3+ tumours was the population for which the EMEA licensing authorisation for trastuzumab was issued. The submission deals both with the FAS population (IHC3+ or FISH+) and with the EMEA defined subgroup which represented 74% of the FAS population of ToGA.

It should be noted that the trial population in ToGA was substantially younger (median ages 59.0 years (CX/F) and 61.0 years (HCX/F)) than the UK population of aGC patients, only 17% of whom are aged under 65 at death. The population of ToGA also differed substantially from the UK clinical population which is predominantly Caucasian; the ToGA trial contained over 50% of patients recruited from Asian countries.

ToGA is the only trial of any treatment restricted to HER2 positive aGC patients. The trials which were identified in the attempt to produce a network meta-analysis recruited patients with aGC or mGC, but with unknown HER2 status (see Table 3 for details of trial populations).

3.2 Intervention

Trastuzumab is licensed for use in combination with cisplatin and capecitabine or 5-FU for the treatment of patients with HER2 positive metastatic adenocarcinoma of the stomach or GOJ who have not received prior anti-cancer treatment for their metastatic disease. HER2 positive patients are defined in the EMEA authorisation as those whose tumours have HER2 over-expression as defined by IHC2+ and a confirmatory FISH+ result, or IHC 3+, as determined by an accurate and validated assay.

The MS defined the intervention in accordance with the SPC, and the ToGA trial employed it in this licensed indication. Given that the licensed population constitutes a subgroup of the mGC population, the technology assessed should not only be treatment with trastuzumab in its licensed combination for the eligible population of those whose tumours are IHC2+/FISH+ or IHC3+ for HER2 over-expression, but also HER2 status test(s) (IHC and/or FISH) for the whole mGC population. The ToGA trial allowed clinical discretion to determine which fluoropyrimidine (5-FU or capecitabine) was administered in combination with trastuzumab plus cisplatin or with cisplatin alone; the great majority (87%) received capecitabine. This pragmatic element of the trial design takes into account the different contraindications of the fluoropyrimidine options available to UK clinicians. The fact that the percentage of patients receiving 5-FU rather than capecitabine was equal in

the HCX/F versus CX/F arms (12.9% versus 12.7% in the EMEA population) alleviates concerns that the HR may have been affected by the possible superiority of capecitabine to 5-FU for OS. There is some evidence of such superiority from a meta-analysis of the REAL-2 and ML17032 trials (HR 0.87; 95% CI 0.77 to 0.98).[4, 10, 15]

The SPC states that patients eligible for trastuzumab require baseline cardiac assessment and that cardiac function should be further monitored at three monthly intervals during treatment. This was adhered to in the ToGA trial.

Trastuzumab is also licensed for the treatment of HER2 positive metastatic breast cancer and HER2 positive early breast cancer where specific criteria are met (see MS pp 10-11).

3.3 Comparators

The NICE final scope defined relevant comparators as:

Cytotoxic chemotherapy regimens which may include 5-FU or capecitabine in combination with one or more of the following: cisplatin, oxaliplatin, doxorubicin, epirubicin, docetaxel.

The manufacturer cited research conducted by Synovate, from whom Roche purchased the data, which involved sampling 112 patient records in September 2009. Further details of this research were requested which clarified the information presented (see manufacturer's response to queries and clarifications).

This indicated that the treatment with the widest UK use was ECX at 45%, with ECF and EOX accounting for another 7% and 6% respectively. Other treatments used were doublets of fluoropyrimidine and platinum agents, and single agent treatments including docetaxel.

This supports the view of the ERG's clinical advisors that the most relevant comparators in the context of UK clinical practice are the four triplet regimens assessed in the REAL-2 trial[4] of epirubicin plus capecitabine or 5-FU and cisplatin or oxaliplatin (ECF, ECX, EOF and EOX). It should be noted that

oxaliplatin is used outside its licensed indication, but that the ERG's clinical advisors consider EOX to form a substantial part of routine UK clinical practice, being used for more than the 6% of patients indicated in the manufacturer's research.

The comparator in the ToGA trial was CX/F (the choice of capecitabine or 5-FU was at the discretion of the clinician). The MS acknowledges that triplet regimens, and especially ECX, constitute current standard UK practice. The doublet CX/F would normally be used in UK practice for patients considered too frail to withstand triplet therapy, but both agents would be administered at lower doses and on differing schedules than those employed in ToGA.

The MS explored the possibility of using a network meta-analysis to assess the comparison between HCX/F and ECX, but concluded that there was insufficient data from RCTs to permit such an analysis.

3.4 Outcomes

The primary outcome in the MS was overall survival (OS), reflecting the primary outcome of the ToGA trial. Other outcomes were PFS; outcomes related to response rate including the clinical benefit rate; the incidence and severity of adverse events; and QoL. QoL was assessed using the EORTC-C30 for the clinical effectiveness and the EQ-5D for the cost-effectiveness analyses.

No statistical analysis of the EORTC QoL data was available; graphical data were presented in support of the statement that QoL improved over the course of the trial for the FAS population in both the HCX/F and the CX/F groups. The ERG requested that the results of any statistical analyses be supplied, and also that data for the EMEA approved subgroup be provided. The EMEA subgroup data were subsequently provided. The CSR was supplied following a request from the ERG; this provided additional graphical data but reported that only qualitative analysis of the QoL data was planned in the ToGA protocol.

The adverse events data relates to the period for which the ToGA trial reports follow-up; no longer-term safety data are available, although maintenance therapy with trastuzumab is included in the economic model (sections 5 and 6).

3.5 Time frame

The original design of the ToGA trial was predicated on an estimate of median OS in aGC of 7 months. A protocol amendment was implemented on the advice of the IDMC to take account of median OS for the entire trial population being found to be in excess of 12 months. The median duration of survival follow-up at the point of clinical cut-off was 17.1 months (range 0 to 31 months) for the CF/X arm and 18.6 months (range 1 to 34 months) in the HCX/F arm.

The trial was terminated early as a result of the interim analysis which showed clinically significant improvements in the EMEA licensed subgroup of patients with IHC2+/FISH+ or IHC3 tumours. This was done in accordance with a revised stopping rule recommended by the IDMC of a significant OS benefit at an interim analysis performed after 75% of events (345 deaths) or 18 weeks from the first treatment of the last patients randomised if this occurred sooner than 75% of events.

3.6 Other relevant factors

The HER2 positive aGC population has not been identified within previous trials and therefore the efficacy of standard triplet regimens (or indeed any therapy) in this particular group is not known.

The MS is predicated upon the eligibility of the technology for end of life status. This applies to technologies which are licensed for small patient populations; where the indication is for patients with a short life expectancy (normally less than 24 months); and where there is sufficient evidence that the treatment offers an extension to life (normally of at least three months) compared to current standard NHS treatment.

Trastuzumab is licensed for the population of mGC patients who are HER2 positive (IHC3+ or IHC2+/FISH+). Using the estimate of 2,900 patients currently given chemotherapy for mGC (MS p 17) and the HER2 positivity rate for these criteria of 17.8%,[13] 516 UK patients per year would be eligible for treatment, in addition to those treated with trastuzumab for HER2 positive breast cancer. However, the entire population of mGC patients who would normally receive chemotherapy (approximately 2900/year) would require HER2 status testing with IHC, and in some cases also with FISH.

The general population of mGC patients has a median life expectancy of less than 12 months with optimum chemotherapy;[4] the only data on the median life expectancy of those who are HER2 positive comes from the ToGA trial, where it was **(confidential information removed)** months in the control (CX/F) arm. The ToGA trial demonstrated that OS was extended by more than three months in the licensed population treated with HCX/F compared to CX/F. No other data on patients known to be HER2 positive exist. The clinical effectiveness submission was unable to establish a network analysis which would permit the comparison of HCX/F with ECX, ECF or EOX (the current NHS treatments for mGC patients) in the mGC population as a whole (see Figure 1). An argument was advanced that lack of evidence for a statistically significant benefit of epirubicin-containing triplet regimens over CX or CF could be interpreted as evidence of no benefit. This was based on the evidence of small RCTs and the ERG did not consider it to have been substantiated (see section 4.2.2). More clinical evidence will be required to determine the OS benefit of HCX/F compared to ECX, ECF or EOX. The MS used various assumptions in order to develop estimates of cost-effectiveness of HCX/F relative to these comparators. These are noted in Figure 3 and explored in detail in section 5.

4 CLINICAL EFFECTIVENESS

4.1 Critique of manufacturer's approach

4.1.1 Description of manufacturers search strategy and comment on whether the search strategy was appropriate.

The MS described the search strategies used to identify relevant studies on the use of trastuzumab for the treatment of HER2 positive mGC. Full details of strategies are provided in the Appendices to the MS.

Overall the search strategies employed for each of the sections of the submission were appropriate and well-documented. A detailed commentary on the individual searches is provided below. There were some potential weaknesses in the strategies provided. However, it is unlikely that any of these would lead to potentially relevant studies being missed by the searches.

The electronic databases MEDLINE, MEDLINE In Process, EMBASE, EMBASE Alert and BIOSIS were searched using the Dialog DataStar interface to identify clinical studies on the use of trastuzumab for the treatment of HER2 positive mGC. In addition to this, abstracts of conference proceedings, the Roche internal 'Publication Planning' database and clinical sections of the application to the EMEA for the extension of the Herceptin Marketing Authorisation to include aGC were reviewed. Searches were conducted at the end of January/beginning of February 2010 from a start date of 1993, and were documented in Appendix 2 Section 10.2 of the MS, not Section 9.2 as stated in the MS.

Overall the search was comprehensive, and included the use of both indexing terms and free text searching. More exhaustive text word searching could have been used, particularly in relation to the terms used for gastric cancer. For example, the term 'neoplasm' might have retrieved additional studies. The trade name 'Herceptin' was not included in the EMBASE search, and Medical Subject Heading (MeSH) terms had not been used in the search of the Cochrane Library.

4.1.2 Statement of the inclusion/exclusion criteria used in the study selection and comment on whether they were appropriate.

ToGA

The inclusion criteria used in the search for evidence on the direct comparison of HCX/F with CX/F were appropriate and adhered to the NICE scope.

Network analysis

The inclusion criteria for the network analysis detailed in the MS were open to individual interpretation. The ERG requested clarification of their derivation. The manufacturer's response indicated that the interventions initially identified as being relevant to the scope by multiple Roche affiliates were assessed for relevance to the UK by Roche's UK affiliate.

4.1.3 Table of identified studies. What studies were included in the submission and what were excluded.

The one study which directly addressed the scope was the ToGA trial which compared HCX/F with CX/F.[1] There were no identified studies which were subsequently excluded from the MS.

The additional studies listed in Table 2 were identified in the MS as being relevant to a network meta-analysis for the comparison between HCX/F and ECX.

Table 2: Additional studies identified as relevant to a network meta-analysis in the MS

Study	Comparison
REAL-2[4]	[ECX + EOX] versus [ECF + EOF] [ECX + ECF] versus [EOF + EOX]
Kim 2001[6]	CF versus ECF

Tobe 1992[7]	CF versus ECF
Yun 2010[5]	CX versus ECX

Details of these studies, together with the ToGA trial are shown in Table 3. Further details can be found in Figures 5-7 of the manufacturer's response to queries and clarifications.

Table 3: Studies included in the clinical effectiveness sections of the MS.

Trial characteristics	**ToGA.**[1]	**Cunningham 2008 (REAL-2)**[4]	**Kim 2001**[6]	**Tobe 1992**[7]	**Yun 2010**[5]	**Ross 2002**[9]
Population						
N (randomised)	594	1002	121	60	91	580
N (PP)	584	964	NR	43	89	574 (analysed)
Eligible age (mean) years	≥18 (60)	≥18 (63)	NR (56)	≤75 (NR)	≤ 75 (range 33-75)	NR (median 58/59)
% male	76	81	72	61	68	77.4
Ethnicity (%)						
Asian	55	NR	NR	NR	100	NR
Caucasian	32	NR	NR	NR	0	NR
Other	13	NR	NR	NR	0	NR
Prior treatment (%)	**Confidential information removed**	8	No	No	49	NR
Cancer site (%)						
Esophagus	0	34	0	0	0	32.8 (excluded from meta-analysis[8])
GOJ	18	26	0	0	0	21.8
Stomach	82	40	100	100	100	42.3
Cancer stage (%)						
Advanced	3	23	5[†]	NR	NR	40.4
Metastasis	97	77	94[‡]	NR	NR	57.1
Histologically confirmed	Yes	Yes	NR	Yes	Yes	Yes
Tumour Type (%)						
Diffuse	9	NR	NR	NR	NR	NR
Intestinal	75	NR	NR	NR	NR	NR
Mixed	16	NR	NR	NR	NR	NR
HER2 positive	Yes	NR	NR	NR	NR	NR
Outcomes reported						
OS	Yes	Yes	Yes	Yes	No	Yes
PFS (or TTP)	Yes	Yes	Yes	No	Yes	Yes
RR	Yes	Yes	Yes	Yes	Yes	Yes
Adverse events/Toxicity	Yes	Yes	Yes	Yes	Yes	Yes
Quality of life	Yes	Yes	No	No	No	Yes

Intervention	HCX or HCF versus CX or CF (physician choice)	ECX or ECF versus EOX or EOF (4 randomised arms)	ECF versus CX	ECF versus CF	ECX versus CX	ECF versus MCF
Regimens	**HCX/F:** • Trastuzumab (8 mg/kg) day1, then 6mg/kg • Cisplatin (80 mg/m^2) day 1 • Capecitabine (1,000 mg/m^2) twice daily for 14 days OR 5-FU (800 mg/m^2) days 1-5 **CX:** • Cisplatin (80 mg/m^2) day 1 • Capecitabine 1,000 mg/m^2 twice daily for 14 days OR 5-FU (800 mg/m^2/day) days 1-5	**ECX :** • Epirubicin (50 mg/m^2) day 1 • Cisplatin (60 mg/m^2) day 1 • Capecitabine (625 mg/m^2) twice daily throughout **ECF:** • Epirubicin (50 mg/m^2) day 1 • Cisplatin (60 mg/m^2) day 1 • 5-FU (200 mg/m^2) daily throughout **EOX:** • Epirubicin (50 mg/m^2) day 1 • Oxaliplatin (130 mg/m^2) day 1 • Capecitabine (625 mg/m^2 twice daily) throughout **EOF:** • Epirubicin (50 mg/m^2) day 1 • Oxaliplatin (130 mg/m^2) day 1 • 5-FU (200 mg/m^2 daily) throughout	**ECF:** • Epirubicin (50 mg/m^2) day 1 • Cisptatin (60 mg/m^2) day 1 • 5-FU (1,000 mg/m^2) days 1-5 **CF:** • Cisptatin (60 mg/m^2) day 1 • 5-FU (1,000 mg/m^2) days 1-5	**ECF** • Epirubicin (30 mg/m^2) day 2 • Cisplatin (80 mg/m^2) day 1 • 5-FU (425 mg/m^2), day 2-5 **CF** • Cisplatin (80 mg/m^2) day 1 • 5-FU (425 mg/m^2) days 2- 5	**ECX:** • Epirubicin (50 mg/m^2) day 1 • cisplatin (75 mg/m^2) day 1 • Capecitabine (1,000 mg/m^2) twice daily for 14 days **CX:** • Cisplatin (75 mg/m^2) day 1 • Capecitabine (1,000 mg/m^2) twice daily for 14 days	**ECF:** • Epirubicin (50 mg/m^2) every 3 weeks • Cisplatin (60 mg/m^2) every 3 weeks • 5-FU (200 mg/m^2/d) for up to 6 months **MCF:** • Mitomycin (7 mg/m^2, maximum dose 14 mg) every 6 weeks • Cisplatin (60 mg/m^2) every 3 weeks • 5-FU (300 mg/m^2/d) for up to 6 months
Treatment cycle	Every three weeks	Every three weeks	Every four weeks	Every two weeks	Every three weeks	Every three weeks

As can be seen from Table 3, there were some differences between these trials. While the populations were broadly comparable; the regimes and outcome measures were not. In particular one key trial did not report OS.[5] The principal point to be noted, however is that, with the notable exception of REAL-2,[4] the trials were not powered to detect differences in OS.

The network diagram presented (MS p80, reproduced in Figure 1a) showed REAL-2[4] as comparing ECF with ECX and EOX. As shown in Table 3, REAL-2 was a 2 x 2 factorial trial which was powered to assess 2 x 2 comparisons rather than comparisons between individual arms. A secondary analysis did include comparisons between all individual arms.

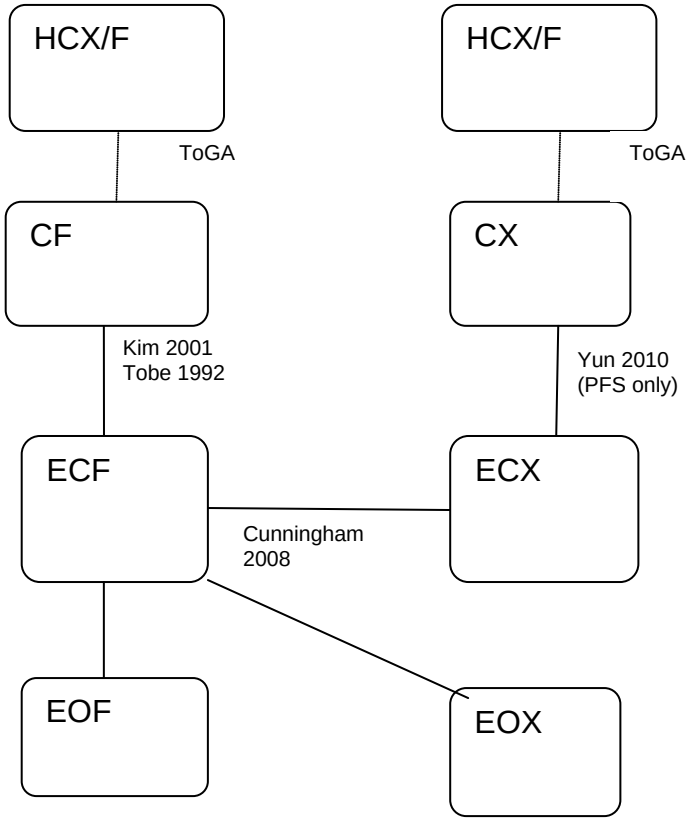
The diagram also showed the individual comparisons [HCF versus CF] and [HCX versus CX]. The analysis of the ToGA trial is for the comparison [HCX or HCF] versus [CX or CF]. Data on the individual comparisons was requested by the ERG. The manufacturer's response stated that no separate analyses were conducted but that the interaction between base chemotherapy and treatment was not statistically significant (p = 0.63) (Response to clarifications and queries). The ERG assumes that this is based on the primary outcome of OS. Given the small numbers of patients treated with 5-FU in each arm of the ToGA trial, the survival advantage for capecitabine over 5-FU suggested by the meta-analysis of Okines et al. (2009)[10] should be borne in mind.

The ERG considers that the network would be more accurately illustrated by Figures 1b or 1c. The assumptions used in order to model the comparisons between HCX or HCF and the triplet regimes ECX, ECF and EOX are shown in Figure 3 (see section 5.1.3).

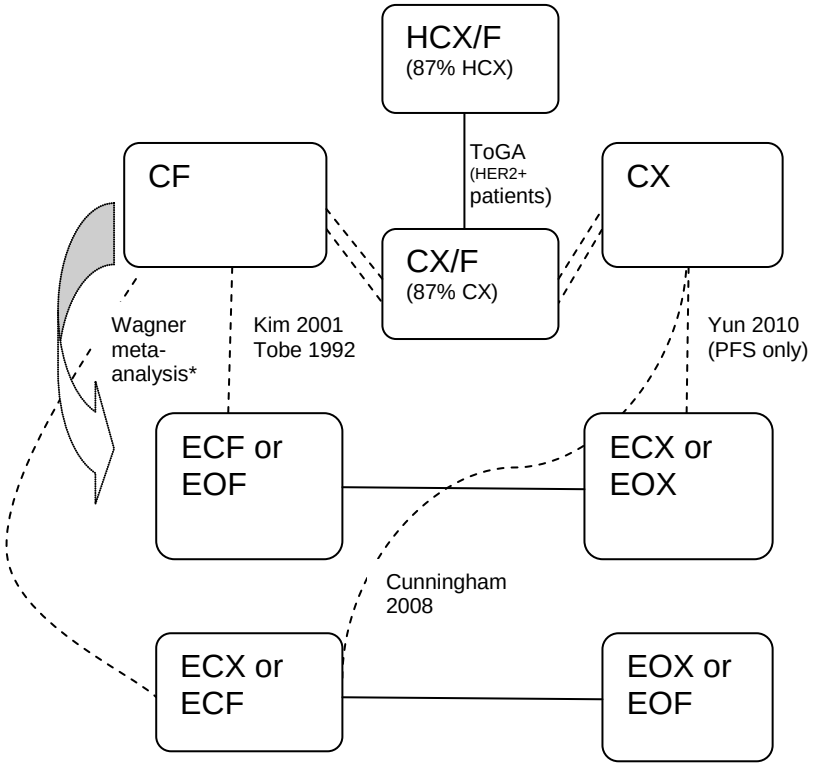
The meta-analysis of Wagner[8] provides a comparison between CF and ECF (see figure 1) although, as the MS notes, the largest trial (N = 334, excluding patients with oesophageal cancer) assessed a comparison between ECF and CF plus mitomycin;[9] the two smaller trials are comparisons of ECF versus CF.[6, 7] Given that the other concerns identified in the MS's critique of Wagner apply with greater force to the use of the individual trial results, the ERG

considers that all trials in the meta-analysis should form part of any network of evidence, and that, in the absence of other evidence, the pooled estimate cannot be dismissed from the assessment of relative efficacy of regimens with and without epirubicin.

In the absence of either a direct comparison or an adequate network of clinical evidence, the MS made a number of assumptions in order to model the cost effectiveness of HCX/F compared to ECX, EOX and ECF. These are shown in Figure 3, and in Table 8, and are explored in detail in sections 5 and 6. Briefly, these involved the assumption that HRs were equal to 1.00 for both PFS and OS, with the exception of an OS benefit of X over F regimes, which was based on an IPD meta-analysis.[10] The ERG were unable to assess the validity of this meta-analysis of two trials evaluating differing regimes, as there was limited reporting of the methodology.[4, 15]



1a) Network of RCTs in aGC populations taken from MS (p 80)



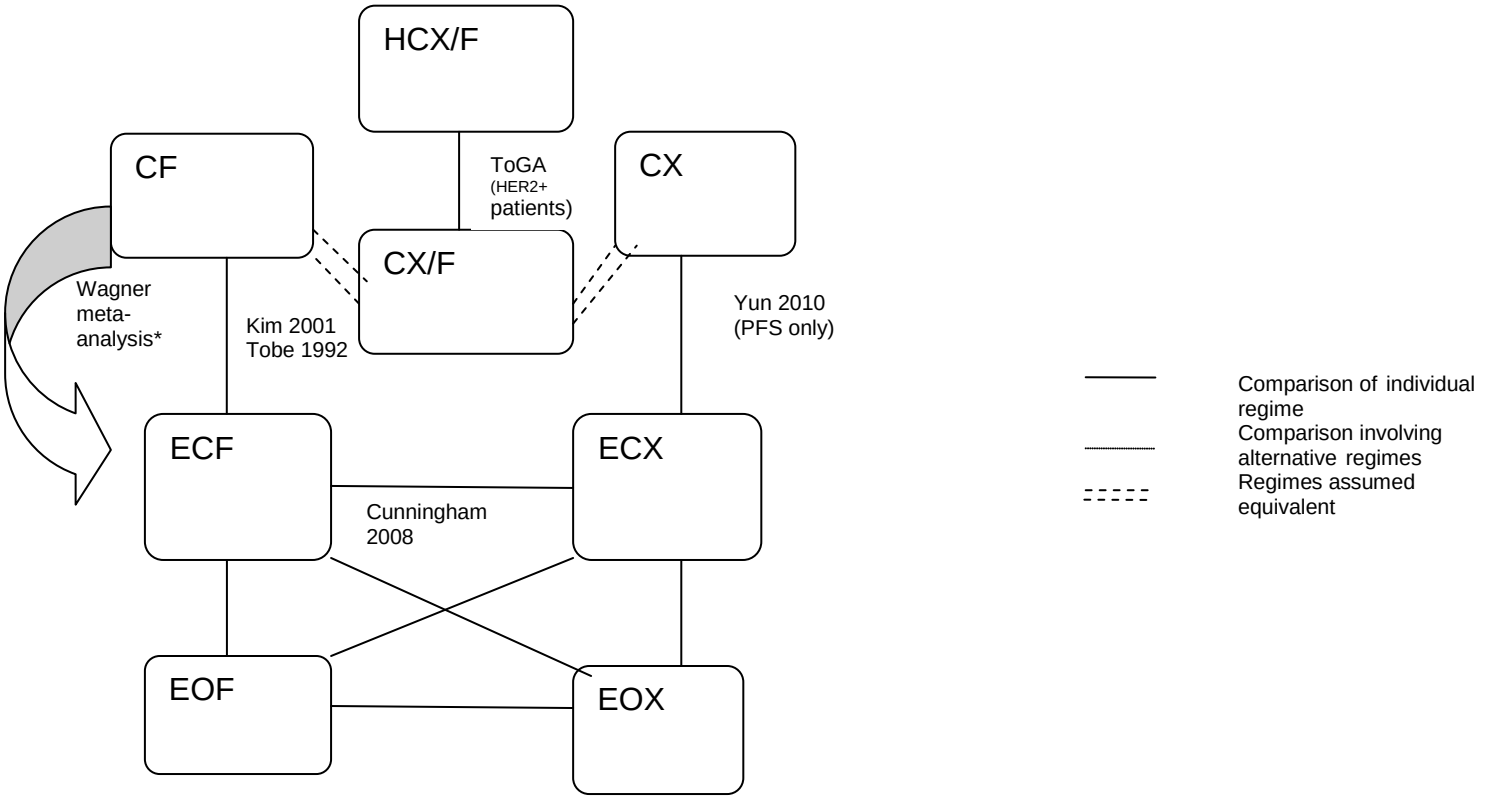
1b) Network of RCTs in aGC populations showing the 2 x 2 comparisons REAL-2 was powered to assess

*Wagner meta-analysis included a trial(Ross 2002) which compared MCF versus ECF

Comparison of individual regime

Comparison involving alternative regimes

Regimes assumed equivalent



1c) Network of RCTs in aGC populations showing the individual arm comparisons from REAL-2.

*Wagner meta-analysis included a trial (Ross, 2002) which compared MCF versus ECF

Figure 1: The network diagram (a) as presented in the MS, (b) showing comparisons REAL-2 was powered to asses, (c) showing individual comparisons between arms in REAL-2

4.1.4 Details of any relevant studies that were not included in the submission

Neither the MS nor the ERG identified any relevant studies which were not included in the direct comparison of HCX/F with a relevant comparator in the population of patients with aGC.

It was not clear from the original MS which studies had been identified but excluded from consideration in the possible network meta-analysis. The ERG requested details of all studies excluded at each stage of the review process. These were subsequently supplied.

The studies in Table 4 (listed in Appendix C2 of the MS, p 207) appear to have been identified but subsequently excluded from consideration. In each case it appears that the study would not have contributed to the network as common comparators were not used.

Table 4. Studies which appear to have been identified but subsequently excluded from consideration for the network meta-analysis.

Study	Comparison
Al-Batran et al. 2008[17]	5-FU plus leucovorin plus oxaliplatin versus 5-FU plus leucovorin plus cisplatin.
Bouche et al. 2004[18]	5-FU plus leucovorin versus cisplatin plus leucovorin versus irinotecan plus leucovorin.
Dank et al. 2008[19]	5-FU plus irinotecan plus folinic acid versus 5-FU plus cisplatin.

4.1.5 Description and critique of manufacturers approach to validity assessment

The validity assessment of the ToGA trial used appropriate criteria. These comprised allocation concealment, randomisation, power calculation, adequacy of follow-up, blinding of assessment, trial design (parallel versus crossover) whether the study was conducted in the UK, representativeness of trial population relative to patients seen in clinical practice, use of dosage

regiments detailed in the SPC, comparability of study groups, appropriateness of statistical analyses, use of an ITT analysis for efficacy data and presence of confounding factors.

However, it was difficult to verify or replicate the assessment as the trial is only reported in a series of abstracts. The ERG requested that the CSR of the trial be provided in order to facilitate this process. On the basis of the manufacturer's response to the ERG's letter of clarification the ERG notes the judgement in the validity assessment in the MS that ToGA was conducted in the UK. Out of a total of 594 patients randomised at 122 centres only 23 patients at six centres were from the UK. The MS also states that an ITT analysis was conducted as specified in the protocol. However, the FAS population which is presented as the primary analysis adheres to these protocol specifications in assessing all patients (n = 584) who were treated at least once, rather than those (n = 594) who were randomised to treatment. The ERG's assessment of key aspects of validity is shown in Table 5. On the basis of the information presented in the CSR, the ERG did not have further concerns with the assessment of validity for the ToGA trial.

Table 5: ERG's validity assessment of the ToGA trial.

Validity issues	
Were methods used to generate allocation sequence adequate?	Yes. Central telephone allocation system used.
Were allocations adequately concealed?	No, open-label trial. Neither investigators nor patients were blind to treatment allocation.
Were the groups similar at baseline?	Yes
Were patients blind to treatment allocation?	No, open-label trial
Were care providers blind to treatment allocation?	No, open-label trial
Were outcome assessors blind to treatment allocation?	No
Were there any unexpected imbalances in dropouts between the groups? If so were they explained/adjusted for?	No
Is there any evidence to suggest that more outcomes were assessed than reported?	No
Did the analysis use ITT analysis appropriately?	No. An analysis of all patients treated with at least one dose was specified in the protocol.

The studies identified in the systematic review process for the network meta-analysis were assessed for validity using the Jadad scale which awards up to five points for the criteria of randomisation, blinding and treatment of withdrawals and dropouts.[20] No assessment of the Tobe trial was reported; presumably because it was reported in abstract form only.[7] Jadad is a widely used scale but has important limitations, not least the fact that it does not assess allocation concealment, nor does it consider whether the studies were adequately powered. This is of considerable importance in the context of the narrative synthesis presented, given the small size of the studies cited as evidence for the comparisons between triplet regimens which are UK standard practice and the doublet(s) used as the comparator in ToGA.

4.1.6 Description and critique of manufacturers outcome selection

The primary outcome was OS, reflecting that of the ToGA trial. Other outcomes were PFS, response rate, QoL and adverse events. These outcomes were relevant to the scope and clinically informative. No statistical analyses were presented for the QoL measures; the ERG requested that these be provided. The CSR was subsequently supplied and confirmed the

manufacturer's statement that only qualitative analyses of these data were undertaken, as specified in the ToGA protocol. The absence of data on the statistical significance of changes or between-group differences in QoL meant that the impact of treatment on this clinically important outcome could not be assessed.

4.1.7 Describe and critique the statistical approach used

As noted in section 4.1.5 the FAS analysis did not include all randomised patients and, therefore, cannot be considered to be a true ITT analysis. However the difference between the randomised population and the FAS population was minimal; 10 randomised patients were excluded from the FAS analysis (four patients were randomised to the HCX/F group, six to the CX/F group). A per-protocol analysis of the FAS population was also supplied for OS in response to a query from the ERG.

The analyses of data from the ToGA trial were reported for both the FAS and the EMEA populations for the outcomes of OS and PFS. Response rate was reported only for the FAS population; the statement that there is no reason to indicate that it would differ in the EMEA population is likely to be conservative. Safety data were reported for the FAS population, which was appropriate. The ERG considered that the decision to focus on data from ToGA, rather than considering data from trials in breast cancer, was reasonable, given the differences in the regimes employed.

4.1.8 Summary statement

The ERG was satisfied that the MS included the only relevant trial in the population of HER2 positive patients with aGC/mGC. The outcome data from the ToGA trial were reported for the FAS and the EMEA approved subgroup for the outcomes of OS and PFS. FAS data were reported for response rate and adverse events. Descriptive QoL data were shown only for the FAS population, equivalent data for the EMEA subgroup were provided in response to the ERG's request. Whilst there are no statistical analyses of the QoL data, the ERG considers it unfortunate that the protocol for a trial in end-of-life treatment did not specify such analyses.

It was not possible to determine from the original MS whether all relevant trials were included for the attempt to construct a network meta-analysis; the ERG requested additional information to establish this. This was subsequently provided, and the ERG was satisfied that all relevant trials were identified. The ERG therefore considered the decision not to conduct a network analysis to be correct, given the difference in trial populations between ToGA and all other trials, the difficulty of establishing direct comparisons between regimens and the fact that one trial did not report OS.[5]

The submission can only, therefore, partially address the decision problem. This is the case because there is no evidence as to the efficacy of current standard treatment (ECX, ECF or EOX) in the known HER2 population, and there was insufficient evidence to establish a network of clinical evidence for the comparison of these regimes with HCX/F in the general mGC population. The attempts in the MS to circumvent this problem by demonstrating the equivalence of ECX/F with CX/F were poorly evidenced. The assumptions required to model the relative cost-effectiveness of the treatments are assessed in sections 5 and 6.

4.2 Summary of submitted evidence

4.2.1 Summary of results

The ToGA trial demonstrated improved OS in HER2 positive patients treated with HCX/F over patients treated with CX/F. In the licensed high HER2 expressing subgroup (74% of the FAS population) the HR was 0.65 (95% CI: 0.51 to 0.83) which corresponded to median survival of 16 months for the HCX/F group versus 11.8 months for the CX/F group. **(Confidential information removed.)**

PFS was also improved for patients treated with HCX/F in the EMEA population (HR 0.64; 95% CI: 0.51 to 0.79) which corresponded to a 2.1 month difference in time to progression/death. In the FAS population there was a statistically significantly higher response rate in the HCX/F group for both partial response (41.8% versus 32.1%, p = 0.01) and for overall response rate (47.3% versus 34.5%, odds ratio 1.70, 95% CI 1.22 to 2.38, p =

0.002). The rate of complete response was also higher in the HCX/F group (5.4% versus 2.4%), but this was not statistically significant (p = 0.06).

No statistical test results were reported for QoL measures; it was reported that in the FAS population both groups showed improvements in QoL over the course of treatment which endured for the duration of PFS. The CSR, which was supplied by the manufacturer at the request of the ERG, confirmed that no statistical results are available for this outcome. It was contended that improved QoL is longer in the HCX/F arm than in the CX/F arm, corresponding to the increased duration of PFS.[3] It is not possible to establish whether this is in fact the case based on the descriptive data presented. QoL data for the EMEA population was also provided at the request of the ERG, but the same uncertainty pertains to the statistical significance of the analysis presented.

Adverse events data were reported for the following categories of events: gastrointestinal, blood and lymphatic, general, metabolism and nutrition, skin and subcutaneous disorders, nervous system disorders, investigated symptoms, renal and urinary disorders, infections. Cardiac adverse events were reported in detail. There were no statistically significant differences between the groups in grade 3 or 4 adverse events. There also appeared to be no statistically significant differences in adverse events leading to treatment discontinuation or dose modification/interruption (Table 8, MS p 86).

The following grade 1 or 2 adverse events were statistically significantly more common ($P < 0.05$) in the HCX/F group: vomiting, diarrhoea, stomatitis, anaemia, fatigue, mucosal inflammation, decreased weight and nasopharyngitis. Constipation was statistically significantly more common in the CX/F group. In accordance with the known safety profile of trastuzumab, more patients in the HCX/F group experienced asymptomatic reductions in left ventricular ejection fraction (LVEF) both of between 10 and 50% (4.6% versus 1.1%) and of greater than 50% (5.9% versus 1.1%). This did not translate into a statistically significant difference in any category of symptomatic cardiac events (MS Table 10, p 90).

4.2.2 Critique of submitted evidence syntheses

The ERG was unable to determine whether all studies potentially relevant to the network meta-analyses were identified in the review process. Details of studies excluded at each stage were therefore requested from the manufacturer. These were supplied and indicated that all relevant studies were included in the network.

Given that all relevant studies were identified, the ERG considered that the decision not to attempt a network meta-analysis was justified. As the MS stated, the ToGA population differs from the general mGC population in that it forms a distinct subgroup and the prognostic impact of HER2 status on outcomes with chemotherapy regimes is unknown. The additional reason cited, that the only trial to assess ECX versus CX used PFS as the primary outcome and did not assess OS,[5] was also considered valid. The ERG also considered that direct comparisons between individual combinations of platinum agents and fluoropyrimidines could not be extracted.

As no network meta-analysis was possible, the MS presented instead a narrative synthesis of the relevant trials. This contained internal contradictions in the value accorded to evidence of particular trials as components of a pooled analysis and as individual comparisons.

The MS advanced the argument that there is evidence for no difference between CX/F and ECX/F. This rested largely on a critique of the meta-analysis by Wagner et al. which found a benefit for triplet regimens including epirubicin over platinum/5-FU alone.[8] This critique was based on a number of premises. These are the fact that the largest trial assesses the comparison between ECF and CF plus mitomycin, rather than between ECF versus CF alone;[9] the fact that one of the trials was published in abstract form only;[6] and the fact that the two trials which did assess ECF versus CF were underpowered to detect a statistically significant difference in OS between the arms.[6, 7]

In place of the pooled OS from Wagner et al. the MS cites the individual trials of Kim et al. (2001)[6] and Tobe et al. (1992)[7] for the comparison of ECF versus

CF and Yun et al (2010);[5] for the comparison of ECX versus CX. These are all small phase II trials which must suffer the constraints which were noted in the critique of the meta-analysis, not least by virtue of being underpowered to detect a statistically significant benefit of treatment.

4.2.3 Summary

The ToGA trial appears to provide evidence of superior overall survival in the HCX/F group compared to the CX/F group in the EMEA approved population. However, there are two reasons to treat this finding with some degree of caution. Firstly, the trial was stopped early, although this was in accordance with an IDMC recommended stopping rule based on a planned interim analysis of the primary outcome. It is nevertheless unclear what the impact of termination may have been on the estimate of effectiveness. Secondly, the finding of significance in the licensed population represents a subgroup analysis. However, this subgroup was defined as a result of advances in the understanding of HER2 testing and therefore has credibility as a distinct population. It is also the case that this population of patients with high HER2 expressing tumours constituted a clear majority of all patients in the ToGA trial (74%). Therefore, while the use of a subgroup as the basis for a submission is potentially problematic, the ERG does not consider it to be a cause for serious concern in this instance.

The ERG was satisfied that there is no available evidence apart from the ToGA trial on the performance of any regimen in HER2 positive patients with mGC. Therefore any assessment of comparisons between HCX/F and current UK standard therapy must rely on comparisons involving CX/F in the general mGC population. The ERG accepted the manufacturer's view that it was not possible to construct a meaningful network analysis for the comparison of HCX/F with ECX/F or EOX/F. However the narrative which was presented in place of such an analysis did not convincingly establish the equivalence of CX/F with ECX/F.

The evidence of the meta-analysis by Wagner et al.[8] is discussed in the MS (p 76) but the finding that triplet ECF combinations provided significantly longer overall survival than CF doublet therapies (HR 0.77, 95% CI: 0.62, 0.91, 3

RCTs, n=501) is disputed. There are valid reasons for questioning this result, notably the fact that the largest trial assessed the comparison between ECF and CF plus mitomycin.[9] An additional reason for caution, not noted in the MS, is that only a subgroup of this trial (patients with gastric or GOJ cancer) is included in the meta-analysis; patients with oesophageal cancer were excluded from the meta-analysis.

Having questioned the validity of the Wagner meta-analysis the MS attempted to construct a network meta-analysis using, in part, the two small trials comparing CF versus ECF, despite having questioned the validity of the evidence they contributed to the pooled estimate derived by Wagner et al. It was concluded that this was not possible, for a number of valid reasons outlined in section 4.2.2 above. In place of a meta-analysis a narrative synthesis of the evidence from individual trials was presented.

The MS uses the lack of individual trial evidence for a statistically significant difference between CF and ECF or between CX and ECX to infer that there is no such difference. This is not a justifiable inference; lack of evidence for a difference is not evidence of no difference. Given that the individual trials were underpowered it is particularly difficult to justify this line of reasoning. It should be noted that pooled estimate of the Wagner meta-analysis for ECF versus CF reflects a consistent direction of effect in the constituent trials (see table 6). The only trial to compare ECX versus CX reported an HR of 0.96 (95% CI: 0.58, 1.57) for PFS; OS was not reported.

Table 6: Estimates of treatment effect on OS from trials comparing ECF with CF

Trial	N	HR (OS): ECF versus CF regimen

Tobe,1992[7]	60	0.57 (95% CI: 0.27 to 1.20)
Kim, 2001[6]	120	0.83 (95% CI: 0.42 to 1.61)
Ross, 2002[9]	334*	0.79 (95% CI: 0.62 to 0.95)**
Pooled estimate[8]	501	0.77 (95% CI 0.62 to 0.95)

*Gastric and GOJ cancer subgroup **ECF versus MCF

If anything, the pooled estimate of the Wagner meta-analysis may be considered to be conservative with respect to ECF, since the comparator in the Ross trial may have derived additional clinical effect from the inclusion of an additional agent (mitomycin).[9] Given that capecitabine-based regimens may be more effective than 5-FU based regimens,[10] and are certainly as effective,[4] it is reasonable to infer that, if ECF is more effective than CF, ECX may also be more effective than CX.

The ERG considers that, as the MS states, there is no evidence as to the efficacy of any chemotherapeutic regimen in the HER2 positive mGC population other than those evaluated in the ToGA trial. The ERG also considers that the relative efficacy of HCX/F compared to the current UK standard treatment (ECX, ECF or EOX) can be inferred from trials conducted in the mGC population as a whole only by employing a series of assumptions which are explored in sections 5 and 6 below. In particular the contention of the MS that these triplet regimes can be regarded as offering no survival benefit over CX/F is not adequately demonstrated; the balance of evidence would indicate that there may be such an advantage (see Table 6). The ERG noted the use of an additional argument, which is that the higher doses used in the ToGA trial (see Table 3) would render the CX regimen more effective than typical doublet therapies, and therefore more comparable to a triplet regimen. As no clinical evidence was presented in support of this assertion the ERG cannot comment on the validity of this hypothesis.

5 ECONOMIC EVALUATION

5.1 Overview of manufacturer's economic evaluation

The manufacturer's initial economic submission to NICE included (references in brackets refer to the manufacturer's submission):

1. A description of the systematic search strategy used to identify existing cost effectiveness studies for trastuzumab in aGC, with full details in an appendix (p97-98, p193-195)
2. A summary of the 'de-novo' economic evaluation conducted by the manufacturer describing the technology, comparators, patient population, model structure, inputs and assumptions and finally the base-case results and sensitivity analysis (p98-174, Figure 18 – 37, Tables 12 – 49).
3. An electronic copy of the Excel model used in the economic evaluation.
4. A detailed series of appendices including full details of the search strategy, details of the assumptions and statistical approaches employed for the extrapolations and other major inputs and detailed results (p183-228).

Following the points of clarification raised by the ERG, a number of addenda were submitted by the manufacturer. These included:

1. A revised electronic copy of the Excel model.
2. Revised base-case cost-effectiveness results with some minor corrections to costs.
3. Exploration of alternative approaches to extrapolating survival estimates and the cost-effectiveness results.
4. Exploration of alternative extrapolation of treatment duration together with an assessment of the impact on the cost-effectiveness results.
5. Additional description of the quality of life data used in the model based on EQ-5D and the mixed modelling methodology.

This section of the ERG report focuses on the economic evaluation submitted by the manufacturer. The economic evaluation is subject to a critical review

based on the manufacturer's report and by examination of the electronic model. The critical review is conducted with the aid of a checklist designed to assess the quality of economic evaluations and a narrative highlighting the key assumptions, possible limitations and any remaining uncertainties. These issues are subsequently explored with additional analyses undertaken by both the manufacturer during the clarification stage, which are found in Section 5.1 and the further analyses by the ERG in Section 6.

The manufacturer's economic evaluation combines clinical, economic and outcome data to determine the cost-effectiveness of trastuzumab as part of combination therapy for mGC in patients that are HER2 positive (IHC2+ with FISH+ or IHC3+). The population used in the economic evaluation was representative of the EMEA subgroup. The evaluation compares treatment effectiveness using clinical trial data from the ToGA trial on time to disease progression (PFS) and overall survival (OS).[1] Patients' quality of life is incorporated by applying utility weights from the ToGA trial and the literature to the modelled health states in order to estimate QALYs. Total costs are calculated by applying unit costs from national databases to estimated resource use. The manufacturer uses a three state transition cohort model with one month cycles over an eight year time horizon. They use the NHS perspective and discount both costs and benefits at 3.5%. Table 7 provides a summary of the manufacturer's economic evaluation, with justifications for key aspects and signposts to the relevant sections of the MS.

The comparators included in the manufacturer's economic analysis were:

- Trastuzumab in combination with cisplatin and capecitabine (HCX)
- Trastuzumab in combination with cisplatin and 5-FU (HCF)
- Epirubicin in combination with cisplatin and capecitabine (ECX)
- Epirubicin in combination with cisplatin and 5-FU (ECF)
- Epirubicin in combination with oxaliplatin and capecitabine (EOX)

The inclusion criteria specified by the manufacturer for the comparator regimens were (i) they were within the final scope of the appraisal issued by NICE and (ii) they are routinely used in the NHS, defined as greater than 10% usage. Usage rates were obtained from a study by Synovate Market Research 2. In addition, clinical experts from a Roche advisory board reported that ECX is the most commonly used regimen in England and Wales and that EOX is also used in a minority of centres. In addition, the experts also reported that 5-FU may be given to patients not suitable for treatment with capecitabine. The MS produced one model and analysed three separate pair-wise comparisons: (i) HCX vs ECX and (ii) HCF vs ECF as primary analyses, and (iii) HCX vs EOX as a secondary analysis.

The MS details what they consider the main assumptions in the economic model on p.115. One of the most significant of these assumptions relates to the treatment effectiveness estimates for PFS and OS applied to the model comparators. The ToGA trial compared the doublet regimens CX/F to the triplet regimens HCX/F, as described previously in section 3.3 of the ERG report, however the doublet regimens CX and CF are not included as separate comparators in the model since it is reported that neither treatment is widely used without E in the UK. Instead the triplet regimens ECX and ECF are used as the main comparators to trastuzumab regimens in the model. Due to the absence of a direct comparison of trastuzumab containing regimens with other triplet therapies it is assumed in the MS that ECX and ECF have the same PFS as that reported for the doublet regimen CX/F in the ToGA trial. It is also assumed that ECX has the same OS as CX/F, but that ECF has a lower survival. The basis for these assumptions has been described and critiqued in earlier sections of the ERG report (4.1.3). From a cost-effectiveness perspective, the main assumption of the model is to assume that adding E to CX/F increases the acquisition cost compared to CX/CF alone but without any additional effectiveness benefit. The impact of this particular assumption and other issues arising from the treatment effectiveness inputs that are related to the economic model are discussed in more detail in later sections of the ERG report (see 5.1.3 and 5.2.3).

Another key issue relates to the overall decision problem itself and the scope of the evaluation. Treatment with trastuzumab requires additional diagnostic testing compared to previous treatments to identify HER2 positive patients. It is important to appreciate that the technology appraised within the economic evaluation is actually a combination of two elements: (i) a single diagnostic strategy, based on sequential testing using IHC and FISH tests, to determine HER2 status and (ii) alternative treatment strategies for HER2 positive patients. The single diagnostic strategy used by the manufacturer to determine HER2 positivity is defined as IHC2+/FISH+ or IHC3+ alone and is based on the EMEA subgroup and the respective license for trastuzumab. However, the diagnostic testing strategy used to determine HER2 status and thus treatment appropriateness will have an impact on the cost-effectiveness of treatment and it should be noted that other thresholds and diagnostic tests have not been considered. It is unknown from the MS how differences in the definition of HER2 positivity effects efficacy of the treatment, the size of the population treated and therefore both the diagnostic costs per patient treated and the potential cost-effectiveness of treatments.

Table 7: Summary of the manufacturer's economic evaluation (and signposts to MS)

	Approach	**Source / Justification**	**Signpost (location in MS)**
Model	Cost-utility analysis using a Markov model		Section 7.3.5.1 p.110
States and events	The model contains 3 states: progression free, progressive disease and dead.	The health states align with the main outcomes reported in the TOGA trial, are typical of previous economic evaluations and facilitate area under the curve approaches to estimate PFS and OS.	Section 7.3.5.1 p. 110
Comparators	Trastuzumab in combination with cisplatin and 5-FU (HCF) or capecitabine (HCX) was compared to: 1. Epirubicin, cisplatin and capecitbine (ECX) 2. Epirubicin, cisplatin and 5-FU (ECF) 3. Epirubicin, oxaliplatin, capecitabine (EOX)	Trastuzumab is assumed to be used in accordance with the recent license extension for mGC. Comparators were included if the combination had a usage of >10% in the NHS. Evidence was presented from a market research study and from a clinical expert advisory board.	Section 7.2.1 p. 100-101 Section 7.3.2 p. 105 p. 107 p. 108 p. 109
Sub groups	No subgroups were considered.	Based on final scope.	Section 7.3.1.3 p.104
Natural History	Based on Markov model. Movements between states were based on the ToGA trial.	An area under the curve model was used to estimate the disease progression and was calculated from PFS and OS estimates from ToGA.	Section 7.3.5.1 p. 110 Figure 21
Treatment effectiveness	Clinical outcomes included PFS and OS. HCF was assumed to be equal to HCX for PFS. The HR for OS for HCF vs HCX was 1.15. ECX was assumed to be equal to CX in ToGA for both PFS and OS. EOX was assumed to be equal to ECX. ECF was assumed to be equal to CF for PFS. The HR for OS for ECF vs ECX was 1.15.	 87% of patients in ToGA received HCX.[1] Based on a meta-analysis by Okines for OS.[10] Based on Yun 2010, PFS HR: 0.96.[5] Based on the REAL-2 trial, oxaliplatin arms vs cisplatin arms OS HR: 0.92.[4] Based on Tobe 2001 and Kim 2002 comparing ECF with FP.[6, 7] Based on a meta-analysis by Okines for OS.[10]	Section 6.6 p.80 p.82 p.81 p.82 p.81 p.82
Health related QoL	Utilities for PFS were estimated from the EQ-5D collected in the ToGA trial. Utilities for progressive disease were from a previous NICE evaluation of a	EQ-5D utility questionnaire was collected every 3 weeks in the ToGA trial until progression. A mixed model was	Section 7.3.7.3 p. 128-129

	treatment for gastrointestinal stromal tumors.	applied to this data to estimate utilities during the PFS period. The utility for progressive disease was not collected in ToGA. The progressive disease utility came from a previous NICE appraisal for a different disease. This was justified by stating that progression free utility from the previous appraisal was the same as from ToGA. They also reference NICE TA91 as having previously assumed transferability of disutility from one tumour type to another.[21]	
Adverse events	Adverse events were included if they were grade 3/ 4 and there was an incidence ≥ 5% observed in any of the arms of the trials. They were costed but were assumed to have no impact on the utility of patients.	Adverse event rates were taken from the ToGA and REAL-2 trials.	Section 7.3.8, p.141
Resource utilisation and costs	Costs were divided into the following categories: HER2 testing, drug acquisition, CVAD installation, drug administration, monitoring during PFS, treatment of adverse events, care costs in PFS after cessation of chemotherapy treatment, and supportive care costs post progression. Resource utilisation was determined by treatment duration in the ToGA trial.	Unit costs were taken from NHS Reference costs 2008/09 and published studies. This follows the treatment effectiveness assumptions.	Section 7.3.8, p.133 to 142
Discount rates	Costs (apart from adverse events) and benefits were discounted at 3.5% per annum	In accordance with the NICE reference case.	Section 7.3.9, p.145
Sensitivity analysis	Probabilistic sensitivity analysis was performed. In addition, a series of one-way sensitivity analyses were performed on treatment effectiveness (including survival curve extrapolation models), utility values, unit costs and various resource use assumptions.	In accordance with the NICE reference case.	Section 7.3.10, p.145 to 149

5.1.1 Literature search

In this section we summarise the literature search undertaken in the MS to identify published economic evaluations of trastuzumab in mGC. The electronic databases MEDLINE, MEDLINE In Process, EMBASE, NHS Economic Evaluation Database (NHS EED) and the OHE Health Economic Evaluations Database (HEED) were searched at the end of December 2009. The MEDLINE and EMBASE databases were searched using the Dialog DataStar interface from 1993 to the most recent date available. The search strategies were provided in Appendix 2 Section 10.7 of the MS.

No previously published economic evaluations of trastuzumab in mGC were identified by the manufacturer searches. Consequently, the manufacturer's de-novo model represents the main source of evidence considered by the ERG.

5.1.2 Natural history

The objective of an economic evaluation is to model the natural history of disease with sufficient detail as to be able to differentiate treatments by clinical, economic and quality of life outcomes. The de-novo analysis presented by the manufacturer uses a state transition cohort model with three stages: (i) Progression Free, (ii) Progressive Disease and (iii) Dead. All patients begin in the Progression Free state and transition through the model with no reversion to less severe states of disease. The model also assumes that the amount of time in the progressive state does not change the likelihood of transitioning to the death state.

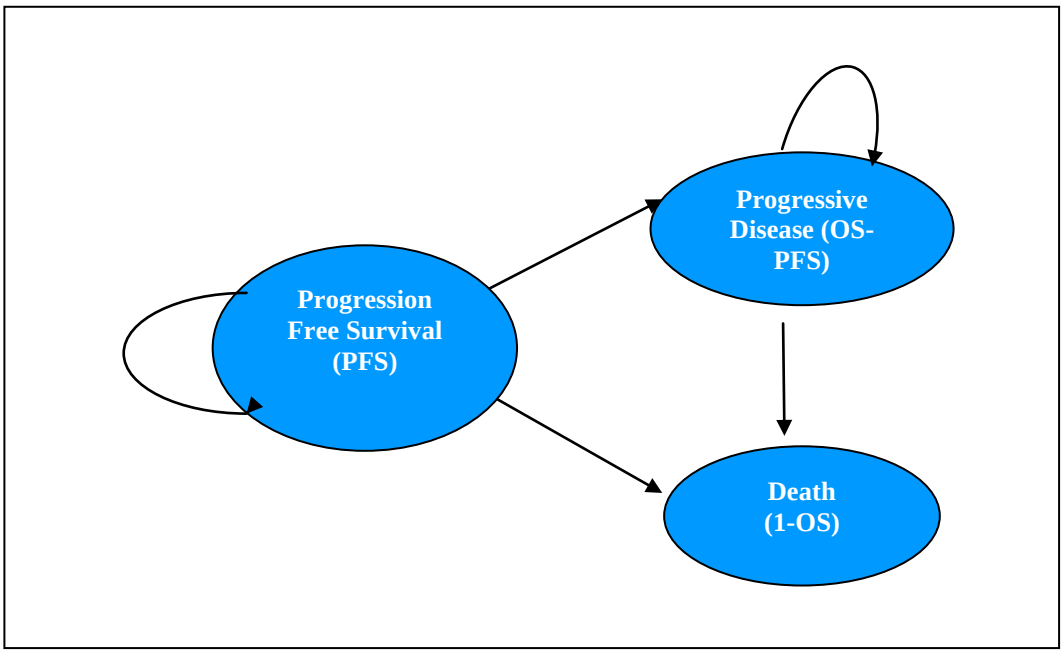


Figure 2: Schema of model design

PFS was estimated from the ToGA trial for the EMEA subgroup. Extrapolation was necessary to estimate PFS beyond the trial length in order to estimate the mean PFS estimates required for the economic analysis. In the main analysis the Kaplan Meier PFS survival estimates were used during the first 12 months and separate Weibull distributions were then used to extrapolate beyond this period due to the low numbers in PFS after 12 months. At 12 months follow-up, 23% of subjects remained progression free in the HCX/F group and 11% remained progression free in the CX/F group. Therefore, at 12 months, the model switches between the KM and Weibull distributions for PFS and an additional assumption was necessary to connect the two distributions. Extrapolation for the period beyond 12 months was calculated by finding the change in the Weibull function $[EXP((\lambda*t^{\gamma})-(\lambda*t_{+1}^{\gamma}))]$ and multiplying by the proportion of patients in PFS at 12 months. Other parametric functions besides the Weibull were also assessed by the manufacturer according to statistical goodness of fit and clinical plausibility. These other functions were rejected in the base-case analysis due to the long 'tails' of the survival distributions that appeared to underestimate progression over a longer time horizon. That is, although other functions appeared to be more statistically appropriate they were less clinically plausible due to the longer time predicted for PFS.

Extrapolation was also required to estimate OS beyond the trial period. OS was estimated from the ToGA trial for the EMEA subgroup. OS was modelled using the Weibull function because it had the best goodness of fit statistics compared to other single parametric survival functions. Hence, in contrast to the approach used for PFS where two separate survival functions were employed (up to 12 months and post 12 months), the estimated Weibull distribution was used to represent the entire OS curve. Other single parametric survival functions were also explored as part of the sensitivity analysis.

The proportion of patients that progress in each cycle of the model was calculated as the difference in those in OS and PFS. Effectively the approach used by the manufacturer is similar to estimating the areas under the respective PFS and OS curves. Time in progression is therefore simply the difference between these areas.

5.1.3 Treatment effectiveness within the submission

Treatment effectiveness included both PFS and overall survival OS. As explained previously in section 4.1.3 of the ERG report, the MS presented a network of evidence between the comparators and argued that it was not possible to perform a robust network meta-analysis. As previously stated in section 4.2.2, the ERG agrees with this assessment. However, in the absence of direct evidence comparing trastuzumab with other interventions used in the NHS, the cost-effectiveness model inevitably requires assumptions to be made about hazard ratios for PFS and OS linking all the comparators considered in the economic model. In the absence of a single trial directly comparing each of the 5 comparators (ECF, EOX, ECX, HCF and HCX), the review of direct and indirect clinical evidence was used to support those assumptions. The network of assumptions linking each of the 5 comparators in the cost-effectiveness analysis is presented in Figure 3, along with the hazard ratios used in the base case model for PFS and OS. The assumptions and the rationale given for them are summarised in Table 8. These are then discussed in more detail in terms of the direct evidence from ToGA and the indirect clinical evidence.

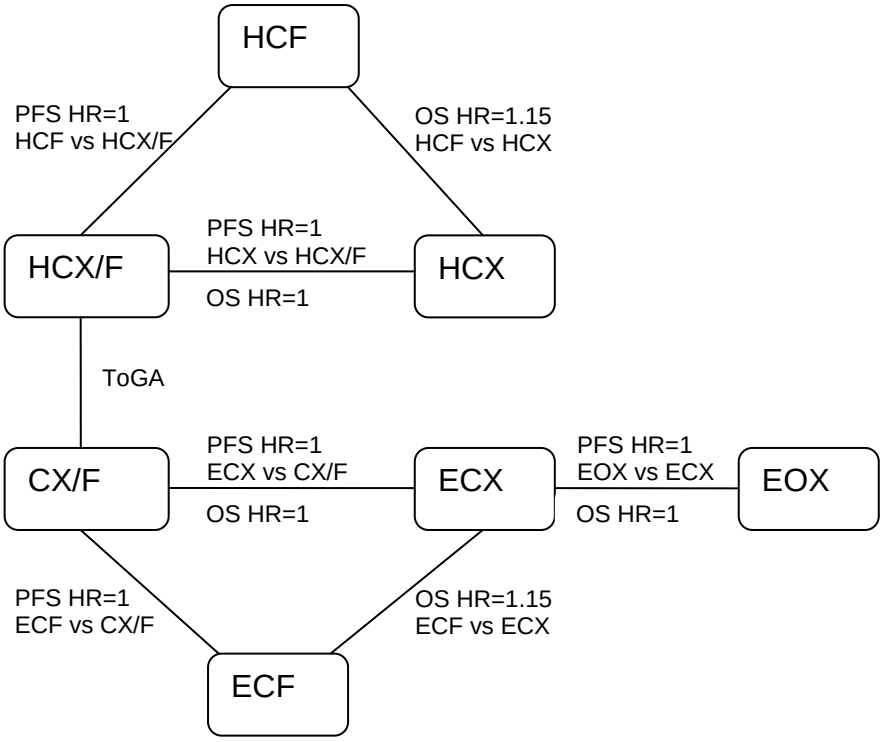


Figure 3: Network of assumptions used in the economic model presented in the MS

Table 8: Assumptions and sources of evidence used to provide effectiveness estimates

	Assumptions and source of effectiveness estimate by outcome	
Comparator	PFS	OS
HCX	Assumed to equal the HCF/X survival curves in the ToGA trial up to 12 months followed by Weibull extrapolation because 87% of patients taking trastuzumab took capecitabine	Assumed to equal the Weibull distribution estimate of the HCF/X survival curve in the ToGA trial because 87% of patients taking trastuzumab took capecitabine
HCF	Assumed equal to HCF/X because no significant interaction for PFS was observed in the ToGA trial between the base chemotherapy and treatment (p-value: 0.6328)	HR of HCF vs HCX = 1.15, based on Okines 2009[10], a meta-analysis comparing capecitabine regimens with 5-FU regimens
ECX	Assumed equal to CF/X in the ToGA trial since there was no evidence of a significant difference (only evidence was Yun 2010[5], PFS HR: 0.96, 95%CI: 0.58, 1.57) and the dose of cisplatin in ECX is lower than the dose of cisplatin in CX The CF/X survival curves in the ToGA trial up to 12 months were used followed by Weibull extrapolation	Assumed equal to CF/X in the ToGA trial since there was no evidence of a significant difference (only evidence was Yun 2010[5], PFS HR: 0.96, 95%CI: 0.58, 1.57) and the dose of cisplatin in ECX is lower than the dose of cisplatin in CX The Weibull distribution estimate of the CF/X survival curve in the

		ToGA trial was used
EOX	Assumed equal to ECX since there was no evidence of a significant difference; only evidence was the REAL-2 trial[4], oxaliplatin arms vs cisplatin arms OS HR: 0.92, 95%CI: 0.8, 1.1	Assumed equal to ECX since there was no evidence of a significant difference; only evidence was the REAL-2 trial[4], oxaliplatin arms vs cisplatin arms OS HR: 0.92, 95%CI: 0.8, 1.1
ECF	Assumed equal to CF/X in the ToGA trial since there was no evidence of a significant treatment effect of adding epirubicin to doublet regimens with cisplatin and either 5-FU or capecitabine; only relevant evidence was Tobe 1992 (HR of OS: 0.57, 95% CI:0.27, 1.2); Kim 2001 (HR of OS: 0.83, 95% CI:0.42, 1.61); and Yun 2010 (HR of PFS: 0.96, 95% CI:0.58, 1.57)[5-7]	HR of ECF vs ECX = 1.15, based on Okines 2009[10], a meta-analysis comparing capecitabine regimens with 5-FU regimens

Direct evidence

One trial (ToGA) evaluated HCF/X and CF/X for this patient population. The PFS survival curves for HCF/X in ToGA for the EMEA subgroup were used to determine PFS for HCX in the cost-effectiveness model. This was justified by the manufacturer based on the high proportion of patients in ToGA (87%) that received capecitabine; the allocation of 5-FU and capecitabine was at the discretion of clinicians. Furthermore, in the manufacturer's clarifications it was stated that no significant interaction was observed in the ToGA trial between the base chemotherapy and treatment (p-value: 0.6328). Though not specified, the ERG assumes this is for the primary outcome of overall survival. For these two reasons, HCF was assumed by the manufacturer to be equal to HCX for PFS.

In the base case analysis for PFS, two separate survival functions were employed. As previously discussed in the 'Natural History' section, Kaplan Meier survival curves were used for the first 12 months and then PFS was extrapolated beyond 12 months using a separate Weibull distribution. A regression analysis was undertaken to estimate the parameters of a Weibull distribution for both the HCF/X and CF/X curves, assuming proportional hazards. Although the doublet regimens (CF and CX) were not included as

comparators to trastuzumab in the economic analysis, the PFS survival curves from this arm of the ToGA trial were linked to assumptions about the relative effectiveness of the comparators that were considered (see Figure 3). The assumptions used to link these comparators to the CF/CX survival curves from the TOGA trial are explained in more detail in the 'Indirect Evidence' section which follows.

For OS, the manufacturer used a single survival function based on a regression analysis to estimate the parameters of a Weibull distribution for both the HCF/X and CF/X curves, assuming proportional hazards. Again, the manufacturer made the assumption that the trastuzumab arm of the ToGA trial was equivalent to a HCX regimen. The HR for OS for HCF vs HCX was based on indirect evidence described below.

As with PFS, a series of additional assumptions and data sources were then employed to link the relative effectiveness of the non-trastuzumab comparator treatment regimens included in the model to the OS data from the CF/CX arm of the ToGA trial.

Indirect evidence

Since ToGA did not directly compare the full range of strategies considered in the economic analysis (HCF, HCX, ECF, EOX, ECX) indirect evidence was reviewed by the manufacturer. The indirect evidence was used to link the regimens evaluated in ToGA to the comparators included in the economic analysis. We have previously reported that the manufacturer assumed that the trastuzumab arm of the ToGA trial was equivalent to the HCX strategy in the economic model. To model a strategy of HCF, the manufacturer further assumed that PFS would be identical to HCX but that OS would be less favourable with HCF (HR of 1.15 compared to HCX), based on a meta-analysis by Okines, 2009 which compared 5-FU and capecitabine regimens.[10]

For the non-trastuzumab comparator regimens a series of additional assumptions were used. Again the manufacturer assumed that the non-trastuzumab arm of the ToGA trial was equivalent to a strategy of CX. Although CX was not considered as a separate treatment strategy in the

economic analysis, this doublet regimen provided the basis for linking to the comparator strategies that were included. The manufacturer assumed that there was no difference in the relative effectiveness of ECX compared to CX for either PFS or OS, since there was no evidence of a significant difference in treatment effect from their review of indirect evidence. The only evidence comparing PFS of ECX and CX cited by the manufacturer was Yun 2010 (PFS HR of ECX vs CX: 0.96; 95%CI: 0.58, 1.57).[5] However, it should be noted that this trial was small with an overall sample size of 89. This assumption of equal effectiveness was also reported to have been validated by discussions with eight oncologists. Also supporting the assumption of equal effectiveness, the oncologists noted that the dose in the CX regimen in the ToGA trial was higher than in the UK current practice, suggesting that this may make the CX regimen as assessed in ToGA more comparable to the ECX regimen assessed in the economic model. The survival curve for PFS for ECX was therefore based directly on the Kaplan Meier curve for CF/X in the ToGA trial up to 12 months and then extrapolated using a Weibull distribution assuming proportional hazards between HCF/X and CF/X. For OS, the estimated Weibull distribution was used for the entire curve.

The manufacturer also assumed that there was no difference in treatment effect between EOX and ECX for either PFS or OS, since there was no evidence of a significant difference in treatment effect based on their review. The only evidence cited was the REAL-2 trial (Cunningham 2008), which showed that for OS the HR for EOF and EOX compared to ECF and ECX was not statistically different (HR= 0.92; 95%CI: 0.8, 1.1).[4]

Finally, the manufacturer assumed no difference in PFS for ECF compared to CF because there was no evidence of a significant treatment effect of adding epirubicin to doublet regimens with cisplatin and either 5-FU or capecitabine. The only evidence considered relevant comparing ECF with CF that was cited were Tobe 1992 (OS HR of ECF vs CF: 0.57, 95% CI: 0.27-1.2) and Kim 2001 (OS HR of ECF vs CF: 0.83, 95% CI: 0.42-1.61).[6, 7] The only evidence comparing ECX with CX cited was Yun 2010 (PFS HR of ECX vs CX: 0.96; 95%CI: 0.58, 1.57).[5] For OS, a HR of 1.15 of ECF vs ECX was used based on

a meta-analysis by Okines, 2009, which compared capecitabine regimens with 5-FU regimens.[10]

In summary, the manufacturer assumed that PFS was equivalent for all the non-trastuzumab comparator regimens (ECX=ECF=EOX) and, in turn, that these were equivalent to the CX/CF arm of the ToGA trial. For OS the manufacturer assumed that ECX and EOX were both equivalent to the CX/CF arm of the ToGA trial but that ECF had a HR of 1.15 (i.e. increased mortality) compared to the CX/CF arm, and HCF had a HR of 1.15 compared to HCX due to a treatment effect of capecitabine over 5-FU.

5.1.4 Health related quality of life

Health related quality of life was estimated for PFS and progressive disease (PD) health states. Utilities were applied to the population in each health state during each cycle to determine the overall QALY estimates for each regimen.

Utility values for patients who were in PFS were calculated using results from the EQ-5D data collected at baseline and then every 3 weeks until progression in the ToGA trial. A total of 431 patients with 3,256 data points were used to estimate PFS utilities. A mixed model was fitted controlling for time from baseline and treatment. In the final model treatment was dropped as it was not statistically significant. The model estimated a baseline utility of 0.7292 with a positive coefficient for time. This coefficient suggests that the quality of life of patients who remain progression-free improves over time.

Since EQ-5D was not collected post progression from the ToGA trial, an external literature based estimate was used to inform the quality of life estimate applied to the progressive disease state. The utility value of 0.577 for progressive disease was taken from a recent NICE appraisal of sunitinib, a second line treatment of gastrointestinal stromal tumors.[22]

5.1.5 Resources and costs

Costs were divided into the following categories: HER2 testing, drug acquisition, drug administration, monitoring during PFS, treatment of adverse events, care costs in PFS after cessation of chemotherapy treatment, and

supportive care costs post progression. These categories are discussed in more detail below.

HER2 testing

The licensed indication for trastuzumab is that patients test HER2 positive. HER2 positive for the EMEA population from the ToGA trial was defined as IHC2+/FISH+ or IHC3+. It was assumed in the economic evaluation that HER2 testing using IHC and FISH would be done in sequence. That is, the IHC test would be given initially to all mGC patients and then only patients that are IHC2+ would then receive an additional FISH test. The proportion of eligible patients, assumed in the model, that are IHC2+ (66%) and thus requiring a FISH test was obtained from Bang 2009.[13] The unit costs for the IHC and FISH tests are presented in Table 9.

The proportion of mGC patients that are eligible for trastuzumab based on the combined testing strategy (IHC2+/FISH+ or IHC3+ = 17.8%) was also obtained from Bang 2009.[13] These estimates were used by the manufacturer to calculate the mean total test costs per mGC patient eligible to be treated with trastuzumab. The number of IHC tests per mGC patient eligible for trastuzumab was 1/0.178=5.61. The number of FISH tests per mGC patient eligible for trastuzumab was 0.66. Thus the mean total cost of HER2 testing was calculated as:

$$(5.61 \times £68) + (0.66 \times £133) = £466.67$$

Table 9: Unit costs of IHC and FISH tests

Test	Cost (£)	Source
IHC	68	Average price of Biomedical and UCL prices
FISH	133	Average price of Biomedical and UCL prices

Drug acquisition

The drug acquisition costs for each drug for every treatment cycle were calculated by multiplying the total quantity of each drug administered for each patient per cycle with the unit price of each drug listed in Table 10. The unit costs of each product were obtained from the BNF58, accessed Jan 10.[23] Although not explicitly reported in this MS, the price of capecitabine was reduced by 10% from the BNF58 quote, which is consistent with a previous submission by the manufacturer in which it was claimed that this was in response to the Pharmaceutical Price Regulation Scheme. For the HCX and HCF regimens, the total quantity of each drug administered per cycle per patient was calculated as the total quantity consumed per cycle in the ToGA trial adjusted for wastage. For the regimens ECF, ECX and EOX the total quantity of each drug administered was taken from REAL-2.

The manufacturer assumed that 80% of centres would vial share implying no wastage. Wastage was calculated for the 20% of centres with no vial sharing by rounding up the dose consumption up to the nearest whole vial or tablet packet required to be at least as great as the mean dose per treatment cycle, using the smallest vial or tablet quantities.

The total quantity consumed per patient per cycle was less than the protocol doses for some of the drugs. The proportion of the protocol dose actually consumed represents the dose intensity. The protocol doses, the dose intensities and the total acquisition costs of the drugs per cycle are presented in Table 11.

The treatment duration for each individual drug in the combined regimens HCF and HCX was calculated from the ToGA trial, using the respective Kaplan-Meier curves for time to off treatment for each individual drug. These were sufficiently complete for every individual drug apart from trastuzumab. The proportion of patients with PFS remaining on trastuzumab was extrapolated beyond month 19 in the model using ordinary linear regression.

In a similar manner to PFS and OS, additional assumptions were required in order to estimate the treatment duration curves of the comparator regimens

assessed in the model. The treatment duration curves for the regimens EOX and ECX were based on the regimen CX/F in the ToGA trial. The treatment duration curve for ECF was based on the regimen CX/F in the ToGA trial. The treatment duration of epirubicin and oxaliplatin was assumed to be equal to the treatment duration of cisplatin.

Table 10: Unit costs of evaluated drugs (BNF58 accessed Jan 10), Table 23, p.133 in MS

Product	£/mg
Trastuzumab	2.7160
Oxaliplatin non proprietary	2.9950
5FU non proprietary	0.0128
Capecitabine	0.0044
Cisplatin non-proprietary	0.5036
Epirubicin non-proprietary	1.6133

Table 11: The recommended dose, dose intensity and total cost per cycle for each drug in each treatment regimen

Drug	Recommended Dose (per cycle per protocol)	Dose intensity	Total cost per cycle (£)
ECF			
Epirubicin	86mg	93%	**128**
Cisplatin	103mg	91%	**48**
IV 5-FU	7,207mg	93%	**86**
ECX			
Epirubicin	86mg	89%	**125**
Cisplatin	103mg	92%	**48**
Capecitabine	45,043mg	88%	**181**
EOX			
Epirubicin	86mg	89%	**125**
Oxaliplatin	223mg	93%	**647**
Capecitabine	45,043mg	88%	**181**
HCF			
Cisplatin	137mg	86%	**59**
IV 5-FU	6,864mg	93%	**82**
Herceptin	376mg	100%	**1064**
HCX			
Cisplatin	137mg	86%	**59**
Capecitabine	48,046mg	82%	**179**
Herceptin	376mg	100%	**1064**

Drug administration

The frequency of administration of each drug per cycle and the dose is presented in Table 12. Although there were 5 main treatment regimens in the economic analysis, some patients in ToGA who started on triple regimens were adjusted to doublet and then single regimens over time and some of

those who started on doublet regimens were adjusted to a single regimen over time. As a consequence the MS calculated drug administration costs for 10 different combinations of drugs as listed in Table 13.

Table 12: Time period of drug administration per 21 day cycle and dose by treatment regimen

	Time period of drug administration and dose					
Treatment regimen	Tastuzumab	Capecitabine	5-FU	Cisplatin	Oxaliplatin	Epirubicin
HCX	Day 1 8mg/kg/1st cycle then 6mg/kg/cycle	Days 1-14 1000mg/m2		Day 1 80mg/m2		
HCF	Day 1 8mg/kg/1st cycle then 6mg/kg/cycle		Days 1-5 800mg/m2	Day 1 80mg/m2		
ECF			Days 1-21 200mg/m2	Day 1 60mg/m2		Day 1 50mg/m2
ECX		Days 1-21 625mg/m2 x 2		Day 1 60mg/m2		Day 1 50mg/m2
EOX		Days 1-21 625mg/m2 x 2			Day 1 130mg/m2	Day 1 80mg/m2

The resource use per cycle for each treatment regimen and their unit costs are presented in Table 13. Significant assumptions for the administration of the treatment regimens are as follows:

- All patients receiving 5-FU required a CVAD to be installed before drug administration could proceed. This was assumed to require a separate hospital visit. This was only required once regardless of the number of cycles administered. The unit cost was obtained from the NHS reference costs 2008/09 HRG QZ14A.[24] Inflated to current prices, this came to £542.

- *Line insertion.* For treatment regimens that include IV 5-FU, a central venous access line is inserted before treatment begins. This central line

remains in place until all cycles of this treatment have finished or until failure of the central line. The line insertion at the start of treatment was costed and patients were assumed not to stay overnight. The cost of replacement of any line insertion that has failed was not costed.

- *Line extraction.* It was assumed that the central line extraction at the end of treatment would coincide with a routine visit, and was therefore not costed.

- *Pharmacy preparation costs.* Drugs were dispensed once per cycle. The pharmacy time required to dispense capecitabine and prepare oxaliplatin was obtained from Millar 2008.[25] The time of 12 minutes was assumed to be the same for the other drugs in the base case. The cost of pharmacist time was taken from the PSSRU, 2009.[26]

- *Pump.* It was assumed that a disposable 5 day elastomeric 'balloon' pump was used for the delivery of 5FU. The cost of a pump was based on the average price of two pumps provided by Baxter quoted on their website. The number of pumps required was 1 in the HCF regimen because 5-FU was only given over a 5 day period, and was 3 in the ECF regimen because 5-FU was given over a 21 day period, which implies that the balloon pump could be used for 7 days. The manufacturer assumed that if the regimen ECF was reduced to F due to toxicity then only 1 pump would be required.

- *Drug delivery and hospital attendances.* For every treatment regimen, the patient attends the hospital only once (unless adverse events occur) and this is on the first day of every cycle. The first chemotherapy and trastuzumab infusions are given in hospital on the first day. For the ECF regimen, patients require 2 district nurse home visits to replace the pump and flush the CVAD; and the HCF regimen requires 1 district nurse home visit to disconnect the pump and flush the CVAD. The manufacturer assumed that if the regimen ECF was reduced to F due to toxicity then only 1 district nurse home visit would be required instead of two.

If infusion is required on the first day then a day case is costed. If only capecitabine is required then no day case is costed. The cost of a day case involving a chemotherapy combination was obtained from the NHS Reference costs. The cost of a day case for receiving trastuzumab alone was obtained from Ward 2006.[27] It was assumed that the cost of a day case for receiving 5-FU alone was the same as that for trastuzumab. The cost of receiving both trastuzumab and 5-FU was assumed to be 20% greater than the cost of receiving trastuzumab alone.

- *Transport costs.* Transport costs were not explained in the MS. The model assumes that 30% of patients on any treatment regimen require hospital transport.

Table 13: Units of resource use per cycle for 10 treatment regimens and the unit costs for each unit of resource use (adapted from the Excel file of the model in the MS).

Unit costs (£)	Resource	Resource use per cycle									
		ECX	EOX	EOF	HCX	HCF	H	HX	X	HF	F
9	Pharmacy preparation & dispensing for infusion drugs	2	2	3	2	3	1	1		2	1
9	Pharmacy preparation & dispensing for oral drugs	1	1		1			1	1		
30	Patient transport	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
38.5	Ambulatory pump			3		1				1	1
39	District nurse visit			2		1				1	1
268	Day case for chemotherapy combinations	1	1	1	1	1					
161	Day case for 5-FU & trastuzumab									1	
134	Day case for 5-FU or trastuzumab alone						1	1			1

Monitoring during PFS

Monitoring during PFS consisted of routine consultations with an oncologist and additional cardiac monitoring. Cardiac monitoring was assumed to be done using a MUGA scan or an echocardiogram. The MS stated that routine consultations with an oncologist take place every three weeks during chemotherapy therapy and every six weeks while on maintenance trastuzumab therapy (and also every six weeks during PFS when a patient is not receiving chemotherapy). The cost of a routine consultation was obtained from the NHS Reference costs.[26]

Cardiac monitoring was assumed to take place once every cycle with epirubicin and once every three months with trastuzumab. These assumptions were employed by the manufacturer based on the respective product licenses reported in the SPC. In the MS 33% of patients were assumed to receive a MUGA scan and 67% were assumed to receive an echocardiogram. The average cost per patient assumed for cardiac monitoring was therefore £133.The unit costs and sources of the MUGA scan and echocardiogram are presented in Table 14.

Table 14: unit costs of IHC and FISH and sources

Test	Cost (£)	Source
MUGA scan	258	Ward 2006[27]
Echocardiogram	79	NHS Reference costs (HRG DA02) inflated to 2010 price year[24]

Adverse events

The inclusion criteria for adverse events in the economic analysis were: (1) grade 3/4 and (2) an incidence ≥ 5% observed in any of the arms of the trials. It was assumed that only grade 3/4 adverse events had any significant cost implications. The adverse events for the regimens HCX and HCF were obtained from ToGA and the adverse events for ECF, ECX, and EOX were obtained from REAL-2.[4] The costs of treating adverse events were taken from

published studies and the NHS Reference costs 2008/09.[24] The costs of adverse events were not discounted and hence were assumed by the manufacturer to be incurred during the initial 12 month period.

The MS reported in Table 31 (reproduced as Table 15 below) that REAL-2 did not report anorexia events. However, the model includes anorexia rates for ECF, EOX and ECX, and the ERG is it is not clear where these estimates came from.

The costs and sources for the adverse events are presented in Table 16.

Table 15: Adverse event rates by regimen (reproduced from Table 31, p.141 in MS)

Adverse event	HCX/F	ECX	ECF	EOX
Anaemia	12.24	10.50	13.10	8.60
Anorexia	6.46	-	-	-
Diarrhoea	9.18	5.10	2.60	11.90
Febrile neutropenia	5.10	6.70	9.30	7.80
Hand-foot syndrome	1.36	10.30	4.30	3.10
Nausea and vomiting	13.61	7.70	10.20	11.40
Neutropenia	26.87	51.10	41.70	27.60

Table 16: Unit cost for treatment of adverse events (reproduced from Table 30, p.140 in the MS)

Adverse event	Unit cost (£'s) inflated to 2010 costs	Reference / comment
Anaemia	582	Agrawal 2006[28]
Anorexia	132	LRIG 2006 Erlotinib[29]
Diarrhoea	237	LRIG 2006 Erlotinib
Febrile Neutropenia	3,272	Ref costs 2008/9[24]
Neutropenia / granulocytopenia	140	LRIG 2006 Erlotinib
Palmar-plantar erythrodysaesthesia syndrome (Hand and Foot)	156	York CRD 2004, September 2004
Vomiting / Nausea	728	Ref costs 2008/9

Second line treatments

It was assumed that there was no difference in the use of second line treatments between the treatment regimens. This was based on similar proportions of patients receiving second line treatment and the mix of drugs between the treatment arms in ToGA.

Supportive care costs post progression

Supportive care costs were included for patients in the progressive disease state. The cost of supportive care (£542 per month) was obtained from the NICE Advanced breast cancer guideline (CG81).[30]

Indexing to the current price year

The price year was 2010. Most costs were inflated to 2010 using the PSSRU 2009 cost index.[26] Inflation for the final year was projected.

5.1.6 Discounting

Both costs and benefits were discounted at an annual rate of 3.5%.

5.1.7 Sensitivity analyses

Probabilistic sensitivity analysis was performed. Distributions were applied to utilities, unit costs, monthly supportive care costs, adverse event probabilities, survival curves parametric parameters, and PFS monthly Kaplan-Maier estimates. In addition, a series of one-way sensitivity analyses were performed. The main analyses are described below and then summarised in Table 17.

- ***Treatment effectiveness***

In the base case the manufacturer assumed that HR=1 for both PFS and OS of ECX vs CX. In a sensitivity analysis (SA), the manufacturer varied the PFS and OS HR of ECX vs CX from 0.96 to 1.04 (Table 41, p.156). The 0.96 comes from the PFS HR in Yun 2010.[5] A HR of 1.04 was chosen as a high value to be symmetrical about 1.

In the base case the manufacturer assumed that HR=1 for both PFS and OS of EOX vs ECX. In a SA, the manufacturer varied the HR of EOX vs ECX for PFS and OS from 0.92 to 1.09 (Table not numbered, p.163).

In the base case, the treatment effectiveness for HCF/X for PFS was determined by using the Kaplan Meier curve for the first 12 months and extrapolating using a Weibull distribution. In a SA, a low estimate of PFS was

derived using a full Weibull model. A high estimate was derived using a log-logistic model.

For OS in the base case, a Weibull model was used for the full curve. In a SA, a low estimate of OS was derived using a log logistic model. A high estimate was derived using a log normal model.

- ***Health related quality of life***

The base case utility values were varied up and down by 10%.

- ***Resources and costs***

In the base case, the percentage of patients requiring transport was 30%. In a SA, this was varied from 0 to 50%. In the base case, the proportion of centres vial sharing was 80%. In a SA, this was varied from 50 to 100%. In the base case, the extrapolation of the proportion of patients on trastuzumab treatment from month 19 was assumed to be linear. In a SA, this proportion was assumed to be constant from month 19.

Most unit costs were varied up and down by 40%.

Table 17: Summary of the scenarios and assumptions included in the manufacturer's sensitivity analysis.

Scenarios considered	Assumptions / Analysis
HR of ECX vs CX	PFS and OS varied from 0.96 to 1.04 based on Yun 2010 (HR=0.96)
HR of EOX vs ECX	PFS and OS varied from 0.92 to 1.09 based on REAL-2 (HR=0.92)
Modelling the PFS of HCF/X	A low estimate of effectiveness used a full Weibull model and a high estimate used a log logistic model
Modelling the OS of HCF/X	A low estimate of effectiveness used a log logistic model and a high estimate used a log normal model
Utility values	Varied up and down by 10%
% of patients requiring transport	Varied from 0 to 50%
Proportion of centres vial sharing	Varied from 50 to 100%
Unit costs	Varied up and down by 40%

5.1.8 Model validation

The internal validation and debugging of the model was reported to have been performed by a health economist who had not been involved in the development of the model. Extreme tests were also reported to have been done to check the plausibility of model outcomes.

5.1.9 Manufacturer's clarifications

The manufacturer adjusted the economic model in the response to clarifications in order to account for different monitoring costs while on chemotherapy, and particularly epirubicin, compared to receiving trastuzumab alone. Consultations with an oncologist are assumed in the MS to be every 3 weeks during chemotherapy treatment compared to every 6 weeks on trastuzumab treatment alone. Cardiac monitoring is assumed in the MS to be once every cycle during epirubicin treatment and once every 3 months otherwise. The revised results are reported in section 5.3 along with the original results.

5.2 Critique of approach used

The ERG has considered the methods applied in the manufacturer's economic evaluation in the context of a detailed checklist reported in Appendix 1 which is used for quality assessing decision analytic models.[31] Table 18 compares the manufacturer's submission to that of the NICE reference case.[32]

Table 18: NICE reference case checklist

Attribute	Reference Case	Included in submission	Comment on whether de-novo evaluation meets requirements of NICE reference case
Comparator(s)	Alternative therapies including those routinely used in NHS	Yes	Included all comparators with more than 10% market use according to their survey.
Perspective - costs	NHS and PSS	Yes	NHS and PSS costs have been taken into account.
Perspective - benefits	All health effects on individuals	Yes	QALY benefits to treated individuals were considered.
Time horizon	Sufficient to capture differences in costs and outcomes	Yes	The economic model has a time horizon of 8 years. Fewer than 0.01% are projected to survive beyond this period.
Synthesis of evidence	Systematic review	Yes	Systematic review was used for clinical effectiveness of comparator treatments. No formal indirect comparison was considered possible due to inadequate evidence. However, significant assumptions on treatment effect were subsequently employed. It is unclear if all other inputs to the model were systematically reviewed, particularly utilities.
Outcome measure	QALYs	Yes	EQ-5D data was collected in the ToGA trial and converted to utilities. The utility estimate for disease progression was obtained from NICE TA179, 2009.[21]
Health states for QALY measurement	Described using a standardised and validated instrument	Yes	Derived from EQ-5D data.
Benefit valuation	Time Trade Off or Standard Gamble	Yes	TTO
Source of preference data	Sample of public	Yes	Societal tariffs from EQ-5D and sample of general public for the TTO trial.
Discount rate	Health benefits and costs	Yes	Benefits and costs have both been discounted at 3.5%.
Equity	No special weighting	Yes	No special weighting was undertaken.
Sensitivity analysis	Probabilistic sensitivity analysis	Yes	Probabilistic sensitivity analysis was undertaken but there were not distributions for all of the hazard ratios, which were the main assumptions.

5.2.1 Literature search

The listed electronic databases searched were appropriate, using adequate time periods. Although the EconLIT database was among the listed databases to be searched for cost-effectiveness information, no search strategy for EconLIT was provided in Appendix 2 Section 10.7. Hence, the ERG cannot verify that the reported search results are correct for this specific database. In addition, the interface used for the searches of the NHS EED and HEED databases was not stated, nor was the date range searched.

The search strategies for the MEDLINE and EMBASE databases were extensive, and used an economics search filter to identify cost-effectiveness data on trastuzumab for gastric cancer. The strategies for NHS EED and HEED were less complex, and could have made greater use of additional free text search terms such as ‘stomach’, or the truncation of ‘gastric’ to retrieve terms such as ‘gastro-oesophageal’.

Although the ERG has not been able to verify all the searches undertaken by the manufacturer and has minor issues about the search terms employed, the ERG considers it unlikely that any previously published studies have been excluded by the manufacturer.

5.2.2 Natural history

In this section we critique the model structure, particularly for how well the natural history of disease is reflected by the model. As mentioned previously the proportion of patients with progressive disease is calculated as the difference in the overall survivors and those who remain progression free. Due to the use of the Weibull function in calculating OS and the use of Kaplan Meier estimates for the PFS (first 12 months) there are some initial time points where the model estimates a negative number of patients in the progressive disease health state. This is due to the manufacturer using different survival functions at different time points and ignoring the correlation between the separate endpoints. This negative number of patients then contributes negative utilities and negative costs to the overall results. However, this logical inconsistency only occurs in the first two cycles of the model and

depends on the distributions chosen for PFS and OS in both arms. Consequently, the overall effect on the ICER is expected to be minor. However, this logical inconsistency has been formally addressed by the ERG as part of the additional work reported in section 6.

The model structure assumes implicitly that patient history does not affect transitions from progressive disease, for example death does not depend on how long a patient has had progressive disease. Using Weibull modelling indicates that the probability of mortality is not constant over time but the Markov structure imposes a structural assumption that the probability of mortality can only be dependent upon the overall time elapsed in the model and not the time elapsed in a particular state. This structural assumption could have been overcome by modelling time from progression to death and using tunnel states to better reflect the impact of time in a particular state. However, while this approach might better characterise the natural history and would also address the correlation between PFS and OS mentioned above, the general approach used by in MS is considered unlikely by the ERG to generate any serious bias.

The manufacturer tested different distributions for modelling both PFS and OS. A combination of the Kaplan-Meier and Weibull estimates were used to model PFS. Although a number of different parametric distributions were also tested for modelling OS, this did not include an equivalent approach to that used for PFS, i.e. Kaplan-Meier with a Weibull extrapolation. The modelling of PFS in this manner is justified by using as much real life data as possible. While a similar justification holds for OS as well, this approach was not used by the manufacturer. While the ERG does not consider that this is a serious omission from the manufacturer's submission, the potential impact of using a common approach to modelling both PFS and OS is explored using additional analyses undertaken by the ERG and reported in section 6.

The model evaluates metastatic gastric patients who are HER2 positive as defined by being IHC 3+ or IHC2+ and FISH+. The costs of testing and the effectiveness in the proportion of patients identified as HER2 are central to the cost-effectiveness of trastuzumab. However, the MS does not explore the

impact of changing the proportion of patients found to be HER2 positive nor do they provide effectiveness results for different thresholds of HER2 positivity. Instead the population is assumed to be fixed when in reality the manufacturers are not only evaluating a new treatment but also a testing strategy. Without additional data on the effectiveness of treatment by different definitions of HER2 positivity, the ERG is unable to further evaluate the potential impact of this assumption. However, it is plausible that a range of alternative testing strategies are possible (e.g. IHC testing alone, parallel versus sequential testing etc) and the relative cost-effectiveness of the specific testing strategy considered has not been demonstrated. Despite these potential concerns, the ERG recognises that the manufacturer has evaluated the cost-effectiveness of trastuzumab in accordance with its product license which states the HER2 positivity should be defined using the approach evaluated by the model.

Although the use of a single diagnostic approach appears justifiable based on the scope and the current license for trastuzumab, the ERG considers that the impact of variability in the rate of HER2 positivity could have been explored in more detail by the manufacturer. That is, even with a single diagnostic testing strategy, the rate of HER2 positivity could vary according to different population characteristics etc. The potential impact of this is explored by the ERG in section 6.

5.2.3 Treatment effectiveness

The MS conducted a systematic review of the literature for relevant clinical evidence. As previously stated in section 4.1.8, the ERG was satisfied that the MS included the only relevant trial in the population of HER2 positive patients with mGC. However, it was not possible to determine from the original MS whether all relevant trials were included for the attempt to construct a network meta-analysis. In the absence of direct evidence comparing all the treatment regimens included in the analysis, the MS referenced indirect evidence and used expert opinion in order to determine the treatment effectiveness of the different treatment regimens. The network of model assumptions between the comparators was presented earlier in Figure 3, and the assumptions and

rationale were summarised in Table 8. These assumptions are now critiqued in more detail by the ERG.

The treatment effectiveness of HCX and CX/F

The ToGA trial compared HCF/X with CX/F. The Weibull model extrapolation beyond month 12 for PFS and for the full treatment curve for OS of HCF/X and CX/F were produced using regression analysis. This analysis modelled the hazard rate for all patients in the trial as a function of treatment arm (HCF/X or CX/F) assuming that the treatment arm had a constant treatment effect over the duration of treatment; this is the proportional hazards assumption. The manufacturer reported in the response to clarifications that the plot of the log negative log of S(t) vs log of time demonstrated that the proportional hazard assumption was appropriate. These plots were not shown. Martingale and deviance residual plots were presented. The full Martingale plots were not complete, but the deviance plots appeared to suggest that the assumption of proportional hazards is appropriate for the specific comparisons presented in the ToGA trial.

The hazard ratio of HCF vs HCX

As explained in section 5.1.3, it was assumed that for PFS the treatment effectiveness of HCF was equal to that for HCF/X. For OS it was assumed that the HR of HCF vs HCX was 1.15 based on a meta-analysis by Okines 2009.[10] If for OS HCX was assumed to be more effective than HCF using secondary data, then it may be equally plausible that it is more effective for PFS also; it may be the extension of PFS that results in an extension of OS. The different assumptions given for the hazard ratios for PFS and OS were not considered by the ERG to be adequately justified, and it would have been helpful to explore the effect of assuming equal effectiveness at either a HR of 1 or 1.15 in sensitivity analyses. The ERG explores alternative assumptions in section 6.

The hazard ratio of ECX vs CX

The MS assumed that the effectiveness of ECX was equal to CX for both PFS and OS. The only evidence cited was by Yun 2010, PFS HR: 0.96, 95% CI: 0.58, 1.57.[5] The trial had an overall sample size of 89 and it is unlikely to be powered to detect a significant effect of treatment. It was argued that the lack of evidence of significant differences in treatment effectiveness justified the assumption of equal effectiveness and this assumption was reported to be validated by 8 oncologists. However, the method of validation was not stated nor whether there was consensus amongst all the oncologists. Moreover, as stated in section 4.2.3, the ERG considers that equal effectiveness cannot be inferred from a lack of evidence. Furthermore, the logical extension of the manufacturer's assumption that ECX is equivalent to CX in terms of effectiveness, would have been to include a doublet CX regimen as a comparator in the cost-effectiveness analysis because it might also be cheaper. However, a doublet CX regimen was not included and hence the incremental cost-effectiveness of trastuzumab has not been compared against a logical additional strategy which naturally arises from the manufacturer's base-case assumptions.

Alternatively, it may be that ECX is more effective than CX despite the lack of significant difference reported in Yun 2010. An argument in favour of ECX being more effective than CX with the same dose of cisplatin is described in section 4.2.3. A more conservative approach might have been to use the HR estimates for both PFS and OS to reflect a marginal improvement in efficacy of ECX evaluated in the economic model compared to the CX regimen in the ToGA trial.[5] The ERG explores the potential impact of these uncertainties in section 6.

The hazard ratio of EOX vs ECX

The MS assumed that the PFS and OS for EOX was the same as those for ECX. As explained in section 5.1.3, this was because of the lack of evidence of a significant difference in treatment effect. This was based on the REAL-2 trial, which showed that for OS the HR of EOF and EOX compared to ECF

and ECX was 0.92 (95%CI: 0.8, 1.1).[4] In the manufacturer's response to clarifications, it is argued that the 2 by 2 comparison is more robust than the individual comparison. However, doubling the sample size does not compensate for an inappropriate comparison. From Table 2 reported in Cunningham 2008, the result of dividing the HR of EOX vs ECF by the HR of ECX vs ECF is a HR of 0.87 of EOX vs ECX for both PFS and OS.[4] The ERG explores the use of this alternative hazard ratio in section 6.

The hazard ratio of ECF vs ECX/CF

The MS assumed that for PFS ECF was equal to CF. As explained in section 5.1.3, this was because of the lack of evidence of a significant difference in treatment effect. For OS it was assumed that the HR of ECF vs ECX was 1.15 based on a meta-analysis by Okines 2009.[10] If for OS ECX was assumed to be more effective than ECF using secondary data, then it may be that it is more effective for PFS also; again, it may be the extension of PFS that results in an extension of OS. The different assumptions given for the hazard ratios for PFS and OS were not justified by the manufacturer, and it would have been helpful to explore the effect of assuming equal effectiveness at either a HR of 1 or 1.15 in sensitivity analyses. However, the ERG notes that adopting the same effectiveness for PFS as for OS (either HR of 1 or 1.15) would not change the cost-effectiveness conclusions as ECF would still be more expensive and therefore dominated by ECX.

5.2.4 Health related quality of life

The search undertaken by the manufacturer to identify gastric cancer utilities identified one study with published utility values. However, this study did not conform to the NICE reference case and the utilities were deemed inappropriate for use in the analysis. This study reports utility values of 0.75 for gastric cancer patients on primary chemotherapy and 0.58 for those on best supportive care. A more recent trial of chemotherapy for gastric cancer was identified by the ERG. This study used the EQ-5D HUI to report the utility for PFS on treatment. The mean utility for a patient who was progression free was 0.76 or 0.66, depending on the treatment regimen[33]. This study was not

identified by the manufacturer and thus was not considered when choosing ranges for the sensitivity analyses.

Despite not identifying the study identified by the ERG, the baseline utility calculated from the ToGA trial appears broadly similar to these external estimates. The ERG also compared the estimates applied in the model to the general UK population norms for EQ-5D and calculated that the baseline utilities from the ToGA trial to be 0.07 lower than the equivalent utility values for the UK general population aged 55-64. Although the ERG considers that the baseline estimate for progression free utility appears reasonable based on the limited published evidence that exists for mGC and do imply a decrement compared to UK norms, there are a series of additional assumptions which are subsequently applied which require further consideration. These assumptions primarily relate to how the baseline utility is modelled over time and the potential impact of adverse events.

A mixed model was fit to the observed data to estimate the change in utility over time. It is not clear from this analysis the magnitude of the missing datapoints or the reasons for missingness. The MS reports that utility scores were collected on the first day of recorded progression for 168 of the 431 patients who began utility collection. If data were not missing at random then it is possible that there is selection bias in terms of the patients who completed follow-up questionnaires for QoL. It is possible that only the relatively healthier patients continue to contribute to the follow-up timepoints in which case the subsequent estimates may overestimate the underlying QoL of the progression free cohort.

Regardless of potential selection bias issues, the ERG considers the base-case approach employed by the manufacturer to be relatively optimistic due to the impact of the time covariate in the mixed model. Table 19 displays utilities estimated from the mixed model at baseline and at different theoretical time points up to 4 years. Although only 1 in 10000 patients are assumed to remain progression free in the model until 4 years, the equivalent utility values applied in the model at this point would be 0.9365. This estimate is 0.2073

higher than at baseline and 0.1365 higher than UK population norms of the same age.

Table 19. Estimated PFS utility over time

Time	Utility
Baseline	0.7292
1 month	0.7335
6 months	0.7551
1 year	0.7811
2 years	0.8329
4 years	0.9365

The ERG considers that a more conservative approach for the base-case analysis might have been to assume the baseline QoL is maintained until progression. That is, QoL neither improves nor declines over time while in the PFS state. Although this approach was considered by the manufacturer as part of their sensitivity analysis, it is plausible that even this approach might be considered optimistic if the impact of adverse events is not considered to have been adequately captured in the EQ-5D follow-up assessments which could arise if the degree of missing data is related to the occurrence of adverse events. Furthermore, even if the impact of adverse events have been appropriately captured by the EQ-5D data, since the QoL for the general population as a whole is naturally declining over time due to the impact of ageing. The ERG considers that the manufacturer should have explored the effect of modelling a declining utility over time for both PFS and PD in either one-way or probabilistic sensitivity analysis.

The manufacturers also assume that QoL for PFS are identical for the different regimes. This assumes that there is no differential utility according to the mode of delivery or the adverse events of the different regimes. However, even if it is considered reasonable to assume the same utility while patients are on different treatments, the fact that the duration of treatments are different and that patients stay on treatment with trastuzumab longer may mean that the assumptions employed may be potentially optimistic.

5.2.5 Resources and costs

In general, the ERG considered that resource use was adequately determined using appropriate sources, and the unit costs were also derived from appropriate sources. The main limitation is that the treatment duration was based on the same assumptions as those used to determine the treatment effectiveness. These assumptions were based on the argument that the lack of evidence of a significant difference in treatment effect. The assumptions about effectiveness also effect resources used and costs because, for example, additional PFS is likely to lead to an increase in the mean time on treatment. In addition, there are several additional points that need to be considered:

HER2 testing

HER2 testing was assumed to be done in sequence as part of the economic model. This raises a number of important issues that relate to the cost assumptions but may also be central in generalising from the effectiveness estimates from the ToGA trial to routine clinical practice when sequential testing is used. The potential impact of variability in HER2 positivity rates also raises further issues.

From a costing perspective the ERG notes that the MS does not incorporate the potential costs of repeat tests that may arise due to test failures. In the manufacturer's clarifications it was reported that out of 3,812 samples there were 485 FISH failures and 176 IHC failures. Any additional costs that these may incur should have been incorporated in the MS. Of the FISH failures, 125 were due to insufficient or no tumour tissue and 360 were due to sample specific failure. Of the IHC test failures, 156 were due to insufficient or no tumour tissue and 20 were due to sample specific failure. The manufacturer argued that the insufficient or no tumour tissue reason was a result of the ToGA trial protocol and is unlikely to occur in the routine setting. It is not clear what percentage of patients would require a second biopsy, if any. If some patients do require a second biopsy this may incur additional costs to the system and time delays to patient treatment. If there is a technical failure the

manufacturer argued that the cost of retesting would be absorbed by the company performing the test and would not affect the cost to the NHS. It is unknown if technical failure would result in a delay to patient treatment.

Perhaps most importantly the treatment effectiveness estimates in the model are derived from the ToGA trial which employed parallel testing using IHC and FISH as part of the trial protocol. The advantage of parallel testing is that it is likely to minimise any additional delays in starting trastuzumab therapy for patients who require confirmatory FISH tests to determine eligibility. It is a key assumption of the manufacturer's economic model that the use of sequential testing will not result in an additional delay in the start of trastuzumab treatment compared to the approach used as part of ToGA protocol (or, alternatively, that any additional delay caused by sequential testing will not adversely impact on either the absolute or relative effectiveness estimates derived from the ToGA trial).

This specific issue is further discussed in section 6 in relation both to the effectiveness and cost inputs. The ERG did not consider it possible to appropriately model the potential negative impact on effectiveness inputs given the lack of existing evidence on the impact that additional treatment delays could cause due to the use of sequential testing. However, an exploratory analysis is undertaken to the cost inputs, effectively altering the diagnostic strategy to be based on a parallel rather than sequential approach. Although this will increase the overall cost assumptions, the ERG considers this an appropriate scenario to consider since it is likely to more closely to match the effectiveness assumptions employed by the manufacturer.

It was also estimated by the manufacturer that 17.8% of mGC patients are eligible for trastuzumab (calculated from Table 3, Bang 2009).[13] However, the percentage varied by country and by patient characteristics. For example, in the UK subgroup of patients the percentage was 26% (n=132). HER2 positive also varied both by tumour site (33.2% in GOJ and 20.9% in stomach) as well as by histological subtype (32.2% intestinal vs 6.1/20.4% in diffuse/mixed). The MS did not consider the impact of population heterogeneity in the economic analysis, which could have been explored, producing different cost-

effectiveness estimates for different identifiable groups. To address this issue, the ERG has explored the effect of varying the trastuzumab eligibility rate in section 6. It should also be noted that the MS also did not evaluate the uncertainty around the mean estimate of the HER2 positive percentage in the PSA.

Drug acquisition

The ERG considers that the costs of drug acquisition were adequately calculated and unit costs were appropriately sourced in the MS. Significant assumptions noted by the ERG include the following:

- The MS extrapolated the proportion of patients on trastuzumab treatment beyond month 19. In the base case a linear regression was applied. In SA a constant proportion was applied from month 19 until progression. The extrapolation was for a small proportion of patients (approximately 7%, read from Figure 27, p.138 in the MS) and the results did not significantly change. The ERG considers this SA to be adequate.

- The treatment duration of epirubicin and oxaliplatin was assumed to be equal to the treatment duration of cisplatin. There was no evidence to support this.

- The relative dose intensities (the % of protocol dose actually administered) were used to calculate the actual treatment costs used in the model. For the HCX and HCF regimens, the dose intensities were derived from the drugs actually administered in ToGA. For the ECF, ECX and EOX regimens, the dose intensities were obtained from REAL-2. This assumes comparable populations in the ToGA and REAL-2 trials.[1, 4]

- The proportion of centres that vial share: In the base case it was assumed that 80% of centres shared vials and packets of drugs and that there would be no wastage in these centres. This was varied between 50% and 100% in the sensitivity analysis. The results did not

change significantly. In the manufacturer's response to the clarifications, the assumption that there would be no wastage in packets of drugs in any centres was tested. Capecitabine was the only drug provided in packet form. The results did not change significantly.

Drug administration

Although there were 5 treatment regimens in the analysis, some patients in ToGA who started on triple regimens were adjusted to doublet and then single regimens over time and some of those who started on doublet regimens were adjusted to a single regimen over time. As a consequence the MS calculated drug administration costs for 10 different combinations of drugs which were previously listed in Table 13.

The number of cycles per month was the same for each of these drug combinations. This was calculated as the average number of cycles per month across the individual drugs in the different regimens in ToGA. When calculating the drug acquisition costs for different drug combinations, the specific cycles per month for each drug were used. The manufacturer argued that a global average is better for applying costs to treatment combinations. It would have been more consistent to calculate different cycles per month for each drug combination. The MS assumption favours trastuzumab. However, the ERG expects that this assumption does not have a significant effect on the results, although it was accounted for when the ERG produced a revised base case model in section 6.

In the event that the regimen of ECF had to be reduced to F due to toxicity, the MS assumed that the number of pumps required per cycle would fall from 3 to 1. This seems unlikely unless the regimen of providing 5-FU changed. This assumption is favourable to ECF. In the event that the regimen of ECF had to be reduced to F due to toxicity, the manufacturer assumed that the number of district nurse home visits required per cycle would fall from 2 to 1. Again, this seems unlikely unless the regimen of providing 5-FU changed. This assumption is favourable to ECF. Both of these issues were also

accounted for when the ERG produced a revised base case model in section 6.

Monitoring during PFS

Cardiac monitoring was assumed to take place once every cycle with epirubicin and once every 3 months with trastuzumab. The clinical experts advising the ERG considered that in the UK cardiac monitoring is not considered routine practice for epirubicin (or would only be routinely undertaken after the accumulation dose exceeds a particular amount over time). Since the average unit cost of a cardiac test was £133, this was considered to by the ERG to be a potentially important assumption that was not considered within the sensitivity analyses in the MS. The ERG investigates the impact of assuming common cardiac monitoring for epirubicin and trastuzumab, in section 6.

Adverse events

Although adverse events were not discounted appropriately, the ERG considers this is unlikely to have had a significant effect on the results. As mentioned in section 5.1.5, adverse events for HCX and HCF were taken from the ToGA trial. The adverse events for ECX, ECF and EOX were taken from the REAL-2 trial.[4] A small percentage of patients in ToGA remained on trastuzumab beyond the trial (approximately 7%, read from Figure 27, p.138 in the MS). It is possible that the cost of adverse events for trastuzumab was underestimated given the censoring of trastuzumab treatment duration in ToGA. Treatment duration in ToGA was discussed in section 5.1.3. The ERG will explore applying the adverse event rates of ECX to HCX in section 6, which has the effect of increasing the adverse event rate for HCX.

Two minor errors in the model were also identified by the ERG including using slightly wrong sample numbers to calculate the probabilities of adverse events, and not correctly inflating a few of the adverse events costs to the year 2010. Again, these were accounted for when the ERG produced a revised base case model in section 6.

5.2.6 Sensitivity analysis

The MS reported probabilistic sensitivity analysis. However, no probability distributions were allocated to the hazard ratios. No one-way SA was done to evaluate a range of hazard ratios. The hazard ratios were the most important assumptions in the model.

5.3 Results included in manufacturer's submission

The MS presents the results of the incremental analysis including each of their comparators on a cost-effectiveness plane (see Figure 4). The manufacturer also evaluated two separate pair-wise comparisons in the primary analysis. These were HCX vs ECX and HCF vs ECF (see section 5.1). The MS state that the pair-wise comparison between HCF and ECF is appropriate for patients for whom capecitabine is unsuitable. The manufacturer compared HCX vs EOX in a secondary analysis. The manufacturer provided updated results in response to the clarifications requested due to an error in the model. This was discussed in section 5.1.9. The incremental cost-effectiveness ratios (ICERs) for these 3 revised pair-wise comparisons are reported in Table 20. The pair-wise ICERs range from £40,711 (HCX vs EOX) to £51,927 (HCX vs ECX) per QALY.

Table 20: Mean ICERs (£/QALY) per patient (revised results)

Comparison	Mean ICERs (£/QALY)
HCX vs ECX	£51,927
HCF vs ECF	£50,838
HCX vs EOX	£40,711

For comparison, the original results, prior to ERG clarifications are listed in Table 21.

Table 21: Mean ICERs (£/QALY) per patient (original results)

Comparison	Mean ICERs (£/QALY)
HCX vs ECX	£53,010
HCF vs ECF	£52,363
HCX vs EOX	£41,795

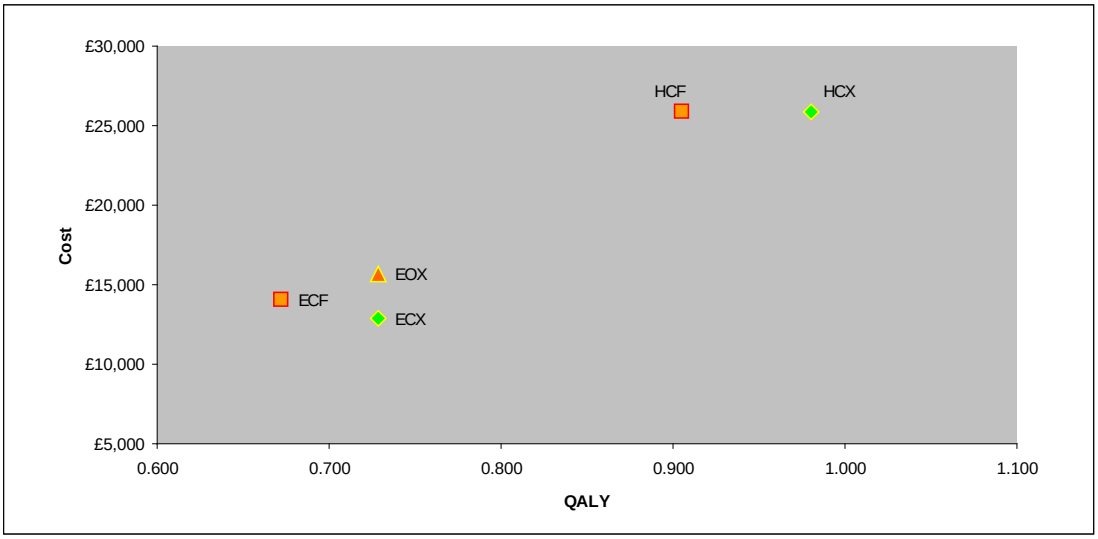


Figure 4: The cost-effectiveness plane, showing the expected costs and QALYs of each regimen (copied from Figure 12 in the manufacturer's response to clarifications)

The MS presented a cost-effectiveness acceptability curve (CEAC), but the probabilities of being cost-effective of thresholds of £20,000 and £30,000 were not reported in the text. From the CEAC graph (Figure 33, p.159) and the cost-effectiveness plane scatter plot (Figure 32, p.158) it appears that the probability of HCX is cost-effective at £30,000 is 0% and at £50,000 is 42%.

In the one-way sensitivity analysis the lowest value of the ICER of HCX vs ECX in the updated results was £48,337. The highest value of the ICER of HCX vs ECX from any of the one-way sensitivity analyses in the manufacturer's updated results was £54,901.

From the probabilistic sensitivity analysis the ICER of HCX vs ECX ranged from £37,180 to £95,238 per QALY at the 2.5 and 97.5 percentiles. The PSA results demonstrate that by varying several parameters simultaneously the uncertainty in the model is much more influential. However, extreme estimates only occur in a small percent of the simulated population.

5.4 Comment on validity of results presented with reference to methodology used

In the manufacturer's base case all the regimens apart from HCX and ECX are dominated either strictly (that is, a comparator is both more expensive and

less effective than either HCX or ECX) or by extension (that is, a comparator is both more expensive and less effective than providing a combination of HCX and ECX). This means that the ICER of £51,927 for the comparison of HCX vs ECX is the relevant ICER for HCX in the manufacturer's base case.

To be a full analysis, all the relevant comparators should be included. It was argued in section 5.2.3 that the assumption of equal effectiveness of ECX with CX makes it necessary to evaluate the regimen CX, since it may be cheaper. The analysis is therefore only a partial analysis, and the ERG explores the impact of including the comparator CX on the cost-effectiveness ratio of HCX in section 6.

The manufacturer presented a CEAC. However, as the uncertainty surrounding the hazard ratios is not included in the model, the CEAC does not fully represent the decision uncertainty.

5.5 Summary of uncertainties and issues

Table 22: Summary of uncertainties and issues identified in Section 5.2

Topic, uncertainty or issue	*Likely consequences for the results and conclusions*	*Additional analysis by manufacturer*	*Additional analysis by ERG*
Comparators			
The trial comparator CX/F was not included in the model	Minor	None	Section 6.3.2 (1)
Natural History			
Distributional Assumptions of PFS and OS	Minor	Section 7.3.5.8	Section 6.3.1
No correlation between PFS and OS	Minor	None	Section 6.3.1
Model Population			
Proportion of patients HER2 positive	Minor	None	Section 6.3.1
Definition of HER2 positive	Major	None	Section 6.3.1
Efficacy			
Differential treatment effects as measured by the hazard ratios	Major	None	Section 6.3.2
Adverse events			
No utility decrement associated with treatment	Minor	None	Section 6.3.3 (3)
Calculation of AEs	Minor	None	Section 6.2
Estimate of AEs for trastuzumab arm	Minor, favours trastuzumab	None	Section 6.3.4 (4)
Health related QoL			
Utility of progressive patients	Minor	Section 7.3.10.4	Section 6.3.3 (1)
Increasing utility over time	Minor	Section 7.3.10.4	Section 6.3.3 (2)
Resource utilisation and costs			
Treatment durations assumed to be equal	Minor	None	None
HER2 testing done sequentially	Favours trastuzumab	None	Section 6.3.4 (2)
Costs of test failures	Minor, Favours trastuzumab	None	Section 6.3.4 (1)
Dosing frequency was assumed to be the same	Favours trastuzumab	None	Section 6.2
Changes in resource use due to switch to monotherapy	Minor, conservative for trastuzumab	None	Section 6.2
Cardiac monitoring of epirubicin	Favours trastuzumab	None	Section 6.3.4 (3)
Sensitivity analysis			
Did not test hazard ratios for treatment effect, PSA and CEAC do not fully reflect the decision uncertainty	Underestimates uncertainty	None	Section 6.3.2

6 Additional 'exploratory' or other work undertaken by the ERG

6.1 Overview

This section presents the results from additional cost-effectiveness analyses undertaken by the ERG. We begin by presenting a revised model correcting for a number of minor calculation errors and inconsistencies identified during

the critique and examination of the Excel model provided as part of the MS. This revised model is then used to undertake a series of additional univariate sensitivity analyses to determine the robustness of the ICER results to alternative assumptions applied in the economic model. These alternative assumptions have been employed to address some of the remaining uncertainties and issues outlined in Section 5 of the ERG report and previously summarised in Table 22. Finally, the ERG presents the results from a series of alternative scenarios. These scenarios are considered by the ERG to represent alternative and equally plausible 'base-case' approaches which explore the combined impact on the ICER of altering a range of alternative assumptions.

6.2 ERG Revisions to Manufacturer's Base-Case Model

The ERG corrected a series of relatively minor calculation errors and logical inconsistencies in the Excel model provided as part of the MS. These programming changes made by the ERG included the following:

(i) inflating AE costs to the current price year (2010) to be consistent with other cost inputs applied in the model;
(ii) changing the probabilities of AEs to match the original source;
(iii) applying logical constraints to rectify the logical error noted due to the use of separate distributions for PFS and OS; and
(iv) adjusting the resource use assumptions to be consistent throughout the model.

A detailed description of the individual programming changes made by the ERG is provided in Appendix 2.

The impact of these programming changes is presented in Tables 23 to 26. These tables report the 3 separate pair-wise comparisons of the ICER (HCX vs ECX, HCF vs ECF and HCX vs EOX) together with a fully incremental analysis of the ICER comparing all the interventions simultaneously.

Table 23: Pair-wise ICER comparison between HCX and ECX

Strategy	*Mean Costs*	*Mean QALYs*	*ICER*
HCX	***£26,891***	***0.980***	***£49,005***
ECX	***£14,559***	***0.729***	***-***

Table 24: Pair-wise ICER comparison between HCF and ECF

Strategy	*Mean Costs*	*Mean QALYs*	*ICER*
HCF	***£26,933***	***0.905***	***£47,907***
ECF	***£15,753***	***0.672***	***-***

Table 25: Pair-wise ICER comparison between HCX and EOX

Strategy	*Mean Costs*	*Mean QALYs*	*ICER*
HCX	***£26,891***	***0.980***	***£40,942***
EOX	***£16,588***	***0.729***	***-***

Table 26: Incremental comparison of ICER including all strategies

Strategy	*Mean Costs*	*Mean QALYs*	*ICER*
HCX	***£26,891***	***0.980***	***£49,005***
HCF	***£26,933***	***0.905***	***Dominated***
ECX	***£14,559***	***0.729***	***-***
EOX	***£16,588***	***0.729***	***Dominated***
ECF	***£15,753***	***0.672***	***Dominated***

The full incremental comparison shows that HCF, ECF and EOX are dominated, leaving the most relevant comparison between HCX and ECX. These changes resulted in the ICER decreasing from £53,010 per QALY to £49,005 per QALY. The pair-wise comparisons are meant to provide the cost-effectiveness of switching to trastuzumab in the small population of those for whom capecitabine is unsuitable. The model revisions undertaken by the ERG lower the ICER comparing HCF and ECF to £47,907 per QALY.

6.3 Additional ERG Sensitivity Analyses

The following sections report the results of the additional sensitivity analyses undertaken by the ERG to address other remaining uncertainties and issues highlighted in section 5.

6.3.1 Natural History

The ERG identified 2 main uncertainties related to natural history. These were related to: (i) the distributional assumptions employed for PFS and OS and (ii) variation in input parameters due to population heterogeneity (e.g. the rate of HER2 positivity in different populations). These are now considered in turn.

Distributional assumptions

As discussed in Section 5.1.2, the base case PFS is modelled using the Kaplan Meier data in the first 12 months and a Weibull distribution thereafter;

OS is modelled using the Weibull distribution only. We explored the robustness of the ICER to different distributions for PFS and OS. Table 27 provides the results of the ICER of HCX vs ECX for each of the possible combinations of PFS and OS distributions. The base case ICER result (£49,005) is highlighted in bold. In addition to the main distributional options provided in the manufacturer's original model we also modelled the OS data using a similar approach to that employed by the manufacturer for PFS, namely the use of a Kaplan-Meier for the first 12 months followed by a Weibull distribution.

The results indicate that the ICER for the comparison of HCX vs ECX ranges from £39,830 to £54,287 per QALY depending upon the combination of survival distributions employed. The lowest ICER from this range is obtained assuming a Kaplan Meier-Weibull distribution for PFS and an exponential distribution for OS, and the highest ICER is obtained assume a log logistic distribution for PFS and Gompertz distribution for OS.

Table 27: ICERs for HCX vs ECX using different combinations of distributional assumptions for overall survival and PFS

		Overall survival distributions					
		KM-Weibull	Weibull	Exponent	Log Logistic	Log Normal	Gompertz
PFS distributions	KM-Weibull	£46,869	**£49,005**	£39,830	£44,394	£40,717	£52,980
	Weibull	£47,619	£49,741	£40,711	£45,158	£41,493	£53,712
	Exponent	£47,215	£49,221	£40,629	£44,879	£41,372	£52,291
	Log Logistic	£49,254	£51,435	£41,904	£46,511	£42,721	£54,287
	Log Normal	£48,292	£50,379	£41,305	£45,825	£42,148	£53,410
	Gompertz	£48,088	£50,210	£41,170	£45,621	£41,943	£54,173

From our additional analysis and given the manufacturer's justifications of clinical plausibility and goodness of fit the ERG is satisfied with the distributional assumptions used for PFS and OS in the base-case analysis reported in the MS.

Variation in input parameters due to population heterogeneity

The previous comparison of the ICER for different combination of survival distributions is based on the licensed EMEA population. Although this

approach to modelling and the use of data is consistent with the license definition of the eligible population for trastuzumab, this approach raises two potential sources of uncertainty.

Firstly, the licensed EMEA population represents a subgroup of the main ToGA trial and the definition of HER2 positivity altered during the course of the ToGA trial. In the clinical effectiveness section, the ERG highlighted that while the use of a subgroup as the basis for a submission is potentially problematic, it was not considered to be a cause for serious concern. However, given that some concerns about the validity of using subgroup data remain, it seems reasonable as part of a sensitivity analysis to explore the potential cost-effectiveness using the main FAS population. The difference between the EMEA subgroup and the main FAS population is that the EMEA subgroup excluded patients who were IHC2+ and FISH-. The ERG also considers that a sensitivity analysis using the FAS population provides an additional indication of the potential robustness of the ICER estimate to potential variation in the use of HER2 testing and the definition of HER2 positivity that could arise in actual clinical practice. Consequently, as an additional sensitivity analysis, the ERG has used survival data from the FAS population as the basis for estimating the ICER estimates and also explored the range of ICERs using alternative distributional assumptions for PFS and OS. Although Kaplan-Meier plots were not provided for the FAS population, the ERG explored all other distributions that were provided by the manufacturer to the ERG during the clarification stage. The ICERs comparing HCX vs ECX in the FAS population ranged from £52,790 to £68,458 per QALY. The ICER based on the Weibull distribution for PFS and OS, which was argued to provide the most appropriate distribution in the EMEA population, was £62,576 per QALY. This analysis demonstrates that the estimate of the ICER is potentially sensitive to the definition of HER2 positive and thus the population considered eligible for treatment.

The second source of uncertainty is also related to the potential impact of population variation. As discussed in Section 5.2.5 under the heading ‘HER2 testing’, the ERG previously noted that there is a potentially wide range

around the manufacturer's estimate of 17.8% for HER2 positivity. This range may depend on the country as well as the disease characteristics. The estimate for HER2 positivity will impact the ICER estimates; higher HER2 positive rates should result in more favourable ICERs for trastuzumab and correspondingly lower HER2 positive rates should result in less favourable ICER estimates. Variation in the rate of HER2 positive was not considered by the manufacturer as part of the sensitivity analysis they undertook. In the absence of this, the ERG has undertaken an additional sensitivity analysis to assess the impact that altering the HER2 positive rate has on the ICER estimates. The ERG tested a range of 5% to 30% for HER2 positive in the model. If the HER2 positive rate was 30% this would result in a more favourable ICER of £48,395 per QALY for the comparison between HCX and ECX. However, using a lower estimate of 5% increases the ICER to £52,866 per QALY. The results of this additional sensitivity analysis undertaken by the ERG suggest that the ICER estimates appear relatively robust to the rate of HER2 positivity applied in the model.

6.3.2 Treatment Effectiveness

The ERG identified a number of potential uncertainties related to the assumptions made for the comparator regimens included in the cost-effectiveness analysis. The following sections examine particular comparator regimens and explore the robustness of the ICER estimates to alternative parameter assumptions made by the ERG.

1) CX comparator

As discussed in section 5.2.3, one of the most important assumptions in the manufacturer's model is that the HR of ECX vs CX is 1 for both PFS and OS based on no evidence of a significant difference. The cisplatin dose is higher in the CX regimen (based on the ToGA trial) than in the ECX regimen (based on the REAL-2 trial). However, the CX regimen was not included as a separate comparator in the cost-effectiveness analysis because the estimated usage in the UK according to a survey was 0%. However, the ERG noted in section 5.2.3 that a logical extension of the manufacturer's assumption that ECX is equivalent to CX in terms of effectiveness, would have been to include

a doublet CX regimen as a comparator in the cost-effectiveness analysis because it might also be associated with lower costs. The ERG therefore undertook an additional sensitivity analysis to incorporate the doublet regimen of CX. For this regimen the ERG undertook 2 separate analyses: (i) using a cisplatin dose equal to that used in REAL-2 trial and (ii) using a dose equal to that used in the ToGA trial. Using the REAL-2 trial cisplatin dose, the ICER for the HCX vs CX comparison was £53,775 per QALY. Using the ToGA trial cisplatin dose, the ICER for the same comparison was £53,567 per QALY. These ICERs for HCX are just under £5k higher than the base-case ICER reported by the manufacturer for the HCX vs ECX comparison.

2) ECX vs CX

There are no trials comparing ECX to CX with a higher dose of cisplatin. However, the exclusion of CX in favour of ECX in the survey in the UK suggests a preference for ECX over CX. Since CX is cheaper than ECX, perhaps ECX is believed to be more effective than CX. The only trial evidence for ECX vs CX comes from Yun 2010, showing a PFS HR: 0.96, 95%CI: 0.58, 1.57. This trial took place in the Republic of Korea and few details of the population were reported, so it is possible this population is not comparable with the scope population. Furthermore, the overall sample was small (n=89). Nevertheless, this is the only evidence available for this specific comparison. It is not known what the correlation between PFS and OS hazard ratios are for this comparison, so the ERG tested a HR of 0.96 for ECX vs CX for PFS only, and for PFS and OS. A HR of 0.96 for ECX vs CX for PFS only increased the ICER of HCX vs ECX from £49,005 to £49,754 per QALY. A HR of 0.96 for ECX vs CX for PFS and OS increased the ICER of HCX vs ECX from £49,005 to £52,709 per QALY.

3) EOX vs ECX

Another important assumption in the model is that the HR of EOX vs ECX is 1 for both PFS and OS based on no evidence of a significant difference. There is no evidence comparing EOX and ECX in the trastuzumab licensed (EMEA) population. The only evidence comes from the REAL-2 trial which has a UK

study sample and more than 200 patients per arm.[4] It is not known what proportion of this sample was HER2 positive as defined by the EMEA population. From the REAL II trial publication, an estimate of the HR for EOX vs ECX for both PFS and OS can be derived; this is 0.87 for both PFS and OS. The ERG therefore explored the robustness of the ICER results to assuming a HR of 0.87 for EOX vs ECX for PFS only, OS only, and for both PFS and OS. There is no difference in results when only the HR for PFS is 0.87. When the HR for OS is 0.87, EOX is no longer dominated by ECX. EOX is the new comparator for HCX. A HR of 0.87 for OS only increased the ICER of HCX to £50,745 per QALY, comparing to the next best option which in this case is EOX (see Table 28). A HR of 0.87 for PFS and OS increased the ICER to £54,114 per QALY, compared to the next best option which in this case is EOX (see Table 29). The relative position of EOX on the cost-effectiveness plane is shown in Figure 5. These analyses do not affect the pair-wise comparison between HCF and ECF.

Table 28: Incremental comparison of ICER including all strategies when the HR of EOX compared to ECX is 0.87 for OS only

Strategy	***Mean Costs***	***Mean QALYs***	***ICER***
HCX	***£26,891***	***0.980***	***£50,745***
HCF	***£26,933***	***0.905***	***Dominated***
EOX	***£17,273***	***0.739***	***£43,696***
ECX	***£14,559***	***0.729***	***-***
ECF	***£15,753***	***0.672***	***Dominated***

Table 29: Incremental comparison of ICER including all strategies when the HR of EOX compared to ECX is 0.87 for PFS and OS

Strategy	*Mean Costs*	*Mean QALYs*	*ICER*
HCX	***£26,891***	***0.980***	***£54,114***
HCF	***£26,933***	***0.905***	***Dominated***
EOX	***£17,212***	***0.802***	***£36,452***
ECX	***£14,559***	***0.729***	***-***
ECF	***£15,753***	***0.672***	***Dominated***

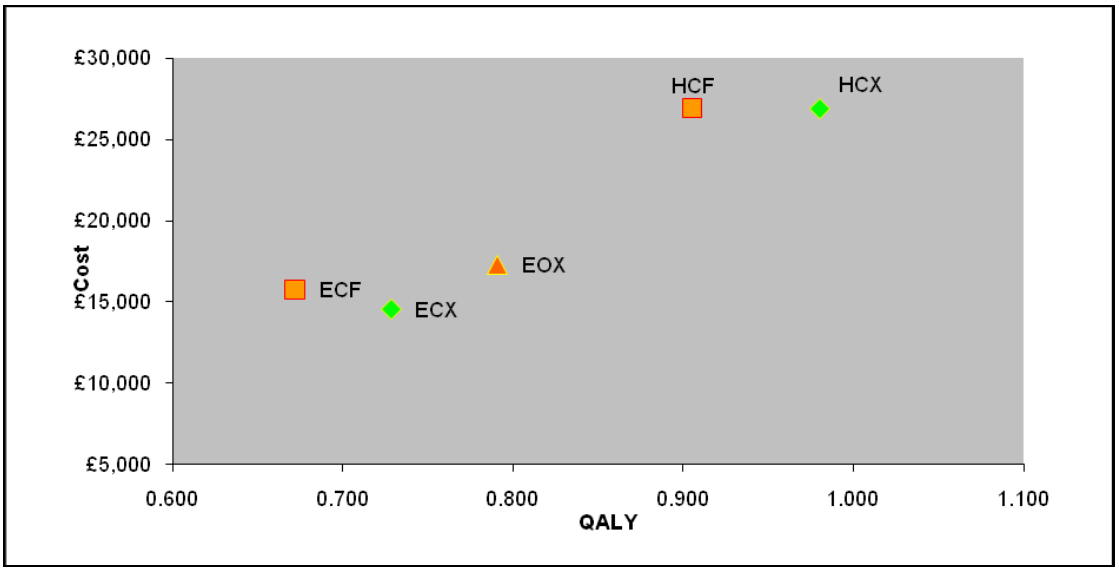


Figure 5: The cost-effectiveness plane comparing all strategies when the HR of EOX compared to ECX is 0.87 for PFS and OS

4) HCF vs HCX

In the base case, a HR of 1 for HCF vs HCX was assumed for PFS but a HR of 1.15 was assumed for OS based on a meta-analysis by Okines 2009. The correlation between PFS and OS is not known, but these assumptions appear inconsistent to the ERG. HCX already dominates HCF in that HCF is less effective and more costly. This result will not change by making the HR for

HCF equal HCX for PFS and OS since the regimens will have the same effectiveness but HCF will remain more costly.

The results of the univariate analyses on the indirect comparisons between regimes resulted in ICERs ranging from £49,005 to £54,114 per QALY. This modest difference in results suggests that the model is robust to changes in the HR between regimens.

6.3.3 Utility

In Section 5.2.4 the ERG identified a number of potential uncertainties related to the utility estimates and assumptions employed in the manufacturer's base-case analysis. To address some of the uncertainties the ERG undertook 3 additional sensitivity analyses:

1) Due to the lack of published data on the utilities for the progression free and progressive disease states, the ERG increased the range of the one-way sensitivity analysis around the base case estimates from 10% (used by the manufacturer) to 20%. Lowering the utility estimate by 20% for both of these states increased the ICER to £60,724.

2) In the manufacturer's base-case analysis it was assumed that utilities in the progression free state are increasing over time based on the results of the mixed-model analysis of the EQ-5D data from the ToGA trial. The ERG identified 2 potential issues with this approach: (i) selection bias could be a potential issue over the follow-up period that EQ-5D data was collected if the cause of missing data was related to severity of disease or the incidence of adverse events; (ii) assuming an increase in QoL does not appear to reflect the impact of ageing over time reflected in the natural deterioration in QoL of the UK general population. To explore this issue the ERG undertook a separate sensitivity analysis in which the utilities in both the PFS and PD states decreased at the same rate as those from an equivalent age group based on the UK general population norms for EQ-5D. The ERG calculated this utility decrement to be 0.003502 per year. Applying this

decrement to the model increased the ICER for HCX vs ECX to £51,309 per QALY.

3) In a final separate test we considered a scenario in which progression free patients had differential utilities based on whether they were on treatment or not. We assumed treated patients had a 5% decrement in their utility compared to those not receiving treatment; this analysis resulted in an ICER of £50,123 per QALY.

The results from the 3 separate analyses produced relatively small changes to the ICER estimates unless progression free and progressive health states were both lowered substantially. The ERG thus considers that the results are relatively robust to this source of uncertainty.

6.3.4 Resource Use

The ERG tested a number of resource assumptions that were not considered by the manufacturer. Four alternative assumptions were considered.

1) The ERG explored the impact of assuming an increase in the cost of HER2 testing to account for IHC and FISH test failures. These were not accounted for in any of the analyses presented by the manufacturer. The ERG based the percentage of test failures on data provided by the manufacturer during the clarification stage. Assuming 6% of IHC tests had to be re-done and 9.4% of FISH test were re-done, the ERG estimated a very small ICER increase to £49,128 per QALY for the comparison between HCX and ECX.

2) The ERG also considered issues related to the timing of IHC and FISH tests. In the base case the manufacturer assumes tests are undertaken sequentially with only those with IHC2+ undergoing a FISH test. The ERG previously noted that it was not considered possible to adjust effectiveness assumptions to take into account the potential impact of any additional delay caused due to the use of a sequential testing strategy compared to the parallel approach used in the ToGA trial. However, any additional delay over and above that which arose in the ToGA trial is likely to increase the ICER estimates reported here.

As an alternative approach, the ERG considers it informative to undertake an exploratory sensitivity analysis to the cost inputs, effectively altering the diagnostic strategy to be based on a parallel rather than sequential approach. Although this will increase the overall cost assumptions, the ERG considers this an appropriate scenario to consider since it is more likely to increase the generalisability of the ToGA trial to a real-life setting. Assuming that the IHC and FISH tests would be undertaken in parallel will result in the number of FISH tests being equal to the number of IHC tests. The resulting ICER for this analysis is £51,618 per QALY for the comparison between HCX and ECX.

3) In the manufacturer's base case it is assumed that the frequency of cardiac monitoring is greater when the patient is receiving epirubicin than when receiving trastuzumab. However, the ERG clinical advisors suggested that cardiac monitoring may not routinely be undertaken when treating patients with epirubicin. Consequently, the ERG undertook an additional sensitivity analysis which assumed that the monitoring frequency for epirubicin is equal to that for trastuzumab. This increased the ICER to £50,816 per QALY.

4) The probability of adverse events for trastuzumab was calculated from ToGA. It is possible that the average number of adverse events that occur during trastuzumab treatment is underestimated since some patients remained on treatment beyond the end of the trial. The ERG undertook an additional sensitivity analysis which assumed that the adverse event costs would be the same for HCX as for ECX. This resulted in an ICER of £52,384 per QALY.

Across the separate sensitivity analyses undertaken by the ERG for resource use and costs the ICER estimate for HCX vs ECX ranged from £49,128 to £52,384. The individual changes suggest that the ICER results appear relatively robust to a range of alternative assumptions.

6.4 Alternative ERG 'base case' scenarios

The results from Section 6.3 indicate that the base-case ICER estimates presented by the manufacturer appear relatively robust to a range of separate assumptions used by the ERG to address a range of uncertainties relating to assumptions and model inputs. However, each of these analyses has been considered separately, consequently the combined impact of altering several of these parameters may be considerably greater.

To explore the potential impact of altering a range of separate assumptions simultaneously the ERG undertook two 'alternative base case' analyses which altered key assumptions of the manufacturer's model for which alternative estimates or assumptions were considered equally plausible to those employed by the manufacturer. These were as follows:

1) The ERG changed the HR of EOX compared to ECX to those calculated from the REAL-2 trial as described above and in Section 6.3.2.

2) The ERG changed the HR of ECX compared to CX to 0.96 for both PFS and OS as described in Section 6.3.2.

3) The ERG changed the utility to incorporate a decrement due to aging as described in Section 6.3.3.

4) Finally, the ERG changed the cardiac monitoring for epirubicin to be the same as trastuzumab as described in Section 6.3.4.

The combined impact of making these three changes to the model changed the comparator to EOX and increased the ICER of HCX vs EOX to £66,982 per QALY (see Table 30).

Table 30: Incremental comparison of ICER including all strategies for the ERG's selected scenario analysis

Strategy	*Mean Costs*	*Mean QALYs*	*ICER*
HCX	***£26,922***	***0.957***	***£66,982***
HCF	***£26,964***	***0.886***	***Dominated***
EOX	***£16,935***	***0.808***	***£37,538***
ECX	***£14,254***	***0.736***	***-***
ECF	***£15,283***	***0.660***	***Dominated***

A second scenario was also considered by the ERG, combining the changes noted above with the additional costs associated with using parallel as opposed to sequential testing. This analysis resulted in an ICER of HCX vs EOX of £71,637 per QALY (see Table 31).

Table 31: Incremental comparison of ICER including all strategies for the ERG's selected scenario analysis including costs associated with parallel

Strategy	*Mean Costs*	*Mean QALYs*	*ICER*
HCX	***£27,616***	***0.957***	***£71,637***
HCF	***£27,658***	***0.886***	***Dominated***
EOX	***£16,935***	***0.808***	***£37,538***
ECX	***£14,254***	***0.736***	***-***
ECF	***£15,283***	***0.660***	***Dominated***

7 Discussion

7.1 Summary of clinical effectiveness issues

The ToGA trial is the only evidence available as to the efficacy of any treatment in the subpopulation of HER2 positive mGC patients. The EMEA licensed population of patients who are IHC3+ or IHC2+/FISH+ constitutes a subgroup (74%) of the FAS of this trial. The ToGA trial showed evidence of increased OS, for HCX/F in both the EMEA and the FAS populations compared to CX/F. CX/F is not used in an NHS context at the doses assessed in the ToGA trial. The primary issue is the lack of direct evidence which would enable the comparison of overall survival in HER2 positive mGC patients treated with HCX/F to those treated with ECX/F or EOX which are NHS standard therapy.

In the absence of evidence from other trials in known HER2 positive patients, the outcome data from trials in the overall mGC population were considered for indirect comparisons. Such an analysis is predicated upon the assumption that HER2 positive patients do not have a significantly different prognosis to the total mGC population. Having identified the extant relevant trials, the manufacturer decided that it was not possible to construct a meaningful network meta-analysis of clinical data from these trials which would allow the HCX/F versus ECX/EOX comparison to be assessed. The ERG considered that this decision was correct.

Since a network meta-analysis was not viable, the MS presented a narrative synthesis of the evidence from relevant trials. This included a critique of the meta-analysis by Wagner,[8] which found evidence of the superiority of ECF over CF without epirubicin. This critique focused on the fact that the largest trial assessed a comparison between ECF and CF plus mitomycin[9] and that the other two trials were small and one was published in abstract only.[6, 7] These two small trials were subsequently cited by the MS as showing no evidence of a statistically significant difference between ECF and CF, from which evidence of no difference was inferred. Given that the MS clearly notes that these trials were underpowered (p76) this inference must be considered

to be a logical fallacy. In particular the ERG considers that this assumption was not conservative with respect to ECF, in view of the consistent direction of effect shown by the trials included in Wagner and the fact that the MCF comparator in Ross is likely to show more, rather than less efficacy than CF alone. Given that ECX/CX may be superior to ECF/CF with respect to OS[10] and is certainly comparable,[4] it may also be the case that ECX is more effective than CX.

There is, therefore, a lack of clinical evidence for the efficacy of the ToGA comparator (CX/F) compared with current NHS therapy (ECX or EOF) in either the HER2 positive population or in the total mGC population. The ERG was also concerned by the lack of statistical data for QoL from the ToGA trial. The graphical data which were available did not allow the impact of treatments on an outcome with high clinical relevance to an advanced cancer population to be assessed.

Any assessment of the clinical and cost-effectiveness of treatment with HCX/F compared with UK standard therapy relies on a number of assumptions, which are discussed in section 7.2 below.

7.2 Summary of cost effectiveness issues

The manufacturer's economic evaluation combined clinical, economic and outcome data to determine the cost-effectiveness of trastuzumab as part of combination therapy for mGC in patients that are HER2 positive (IHC2+ with FISH+ or IHC3+). The evaluation compared two separate trastuzumab combination regimens (HCX and HCF) with three epirubicin combination regimens (ECX, EOX and ECF). The population used in the economic evaluation was representative of the EMEA subgroup. The manufacturer used a three state transition cohort model (progression-free, progressive disease and death) with one month cycles over an eight year time horizon. The evaluation compared treatment effectiveness using clinical trial data from the ToGA trial on time to disease progression (PFS) and overall survival (OS) combined with assumptions derived from literature and expert opinion. Patients' quality of life was incorporated by applying utility weights from the

ToGA trial and the literature to the modelled health states in order to estimate QALYs.

The manufacturer compared the cost-effectiveness of all 5 regimens simultaneously and demonstrated that 3 of these regimens were ruled out by either dominance (EOX and ECF) or by extended dominance (HCF). Of the 2 remaining (non-dominated) regimens, ECX appeared both less costly and less effective than HCX. The incremental cost-effectiveness ratio (ICER) of HCX vs ECX was £53,010 per QALY. This ICER was subsequently altered by the manufacturer to £51,927 per QALY during the clarification stage, following minor corrections to their original costing estimates. The probability that HCX was cost-effective at £30,000 was 0%. The findings were reported by the manufacturer to be robust across a wide range of alternative assumptions.

After revision for corrections identified by the ERG, the ICER of HCX vs ECX fell to £49,005 per QALY. The ERG noted a number of alternative assumptions which were considered equally plausible to those used by the manufacturer in their base-case analysis. When the ERG altered these assumptions the comparator changed to EOX and the ICER comparing HCX vs EOX increased to £66,982 per QALY.

A detailed critique of the manufacturer's initial submission and revised model following points for clarification was undertaken by the ERG. The economic model structure was considered appropriate for the decision problem, and the general approach employed by the manufacturer to estimate lifetime cost-effectiveness was deemed appropriate and met the requirements of the NICE reference case approach. However, the ERG identified a number of model assumptions that had not been subjected to sensitivity analysis and a few alternative and equally plausible assumptions to those used in the model. Adjusting the revised base case model to test each individual assumption, the ICER mostly increased, but the model results remained relatively robust to a range of additional one-way sensitivity analyses.

The three most significant assumptions for which the ERG considered there to be equally plausible alternatives were:

(i) the relative effectiveness estimates of particular comparators; (ii) the utility values applied during PFS; and (iii) the frequency of cardiac monitoring with trastuzumab and epirubicin. In addition to these assumptions, the ERG also felt that there was insufficient discussion of the logistical issues of undertaking HER2 testing in this population and whether the effectiveness results from the ToGA trial (where parallel testing using IHC and FISH tests was used) could be generalised without any loss in effect due to potential delays that could arise for IHC2+ patients based on the sequential testing approach included in the model.

The ERG undertook a series of alternative 'base-case' analyses to address these perceived weaknesses, varying the key assumptions and altering the cost of the testing strategy. The results of these analyses changed the comparator to EOX and increased the ICER for the comparison of HCX vs EOX to between £66,982 and £71,637 per QALY.

Several key areas of uncertainty remain concerning the potential implications of sequential versus parallel testing for HER2 (e.g. whether delays for additional testing of IHC2+ patients could have an impact on patient outcomes) and the definition of HER2+ itself and hence eligibility for trastuzumab.

7.3 Implications for research

It appears unlikely that the comparison between HCX/F and triplet regimens (particularly ECX/F) in HER2 positive aGC patients will be explored in an RCT. However, one possible avenue of research would be to retrospectively HER2 type preserved tumour samples from patients treated with standard triplet therapies (for example those enrolled in REAL-2[4]) and assess the clinical outcomes for the HER2 positive subgroup. While lacking the rigour of data obtained from an RCT, this information would permit some comparison of survival in comparable patient populations treated with triplet therapies versus HCX/F.

References

1. Van Cutsem E, Kang Y, Chung H, Shen L, Sawaki A, Lordick F. Efficacy results from the ToGA trial: A phase III study of trastuzumab added to standard chemotherapy (CT) in first-line human epidermal growth factor receptor 2 (HER2)-positive advanced gastric cancer (GC). In: *2009 ASCO Annual Meeting*; 2009. 2009. Abs No. LBA4509

2.Aaronson N, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez N, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993;85:365-76.

3. Satoh T, Leon J, Lopez RI, Ferry DR, Bang Y, Van Cutsem E, et al. Quality of life results from a phase III trial of trastuzumab plus chemotherapy in first-line HER2-positive advanced gastric and GE junction cancer. In: *2010 Gastrointestinal Cancers Symposium* 2010. 2010.. Abs No. 7

4. Cunningham D, Starling N, Rao S, Iveson T, Nicolson M, Coxon F, et al. Capecitabine and oxaliplatin for advanced esophagogastric cancer. *N Engl J Med* 2008;358:36-46.

5. Yun J, Lee J, Park SH, Park JO, Park YS, Lim HY, et al. A randomised phase II study of combination chemotherapy with epirubicin, cisplatin and capecitabine (ECX) or cisplatin and capecitabine (CX) in advanced gastric cancer. *Eur J Cancer* 2010;46:885-91.

6. Kim TW, Choi SJ, Ahn JH, et al. A prospective randomized phase III trial of 5-fluorouracil and cisplatin (FP) versus epirubicin, cisplatin, and 5-fu (ECF) in the treatment of patients with previously untreated advanced gastric cancer (AGC). *Eur J Cancer* 2001;37:S314.

7. Tobe T, Nio Y, Tseng CC, et al. A randomized, comparative-study of combination chemotherapies in advanced gastric-cancer - 5-fluorouracil and cisplatin (fp) versus 5-fluorouracil, cisplatin, and 4'-epirubicin (fpepir). *Anticancer Res* 1992;12:1983-88

8. Wagner AD, Unverzagt S, Grothe W, Kleber G, Grothey A, Haerting J, et al. *Chemotherapy for advanced gastric cancer*. In: Cochrane Database of Systematic Reviews 2010, Issue 3. Art. No.: CD004064. DOI: 10.1002/14651858.CD004064.pub3; 2010.

9. Ross P, Nicolson M, Cunningham D, Valle J, Seymour M, Harper P, et al. Prospective randomized trial comparing mitomycin, cisplatin, and protracted venous-infusion fluorouracil (PVI 5-FU) with epirubicin, cisplatin, and PVI 5-FU in advanced esophagogastric cancer. *J Clin Oncol* 2002;20:1994-2004.

10. Okines AFC, Norman AR, McCloud P, Kang YK, Cunningham D. Meta-analysis of the REAL-2 and ML17032 trials: evaluating capecitabine-based combination chemotherapy and infused 5-fluorouracil-based combination chemotherapy for the treatment of advanced oesophago-gastric cancer. *Ann Oncol* 2009;20:1529-34.

11. Bassler D, Briel M, Montori VM, Lane M, Glasziou P, Zhou Q, et al. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. *JAMA* 2010;303:1180-87.

12. Bang Y, Chung H, Xu J, Lordick F, Sawaki A, Lipatov O, et al. Pathological features of advanced gastric cancer (GC): Relationship to human epidermal growth factor receptor 2 (HER2) positivity in the global screening programme of the ToGA trial. In: *2009 ASCO Annual Meeting*; 2009. 2009. Abs No. 4556

13. Bang YJ, Chung HC, Xu JM, Lordlick F, Sawaki A, Lipatov O, et al. Pathological features of advanced gastric cancer: relationship to human epidermal growth factor receptor 2 positivity in the global screening programme of the ToGA trial. In: *45th ASCO Annual Meeting*; 2009 29 May - 2 June; Orlando, Florida, USA. 2009. Poster 4556.

14. Chung HC, Bang YJ, Xu JM, Lordick F, Sawaki A, Lipatov O, et al. Human epidermal growth factor receptor 2 [HER2] in gastric cancer [GC]: results of the ToGA trial screening programme and recommendations for HER2 testing. In: *ECCO 15 and ESMO 34th Multidisciplinary Congress*; 2009 20-24 September; Berlin. 2009. Poster 6511.

15. Kang YK, Kang WK, Shin DB, Chen J, Xiong J, Wang J, et al. Capecitabine/cisplatin versus 5-fluorouracil/cisplatin as first-line therapy in patients with advanced gastric cancer: a randomised phase III noninferiority trial. *Ann Oncol* 2009;20:666-73.

16. Lordick F, Schulze T. Molecular prognostic factors and new systemic therapies in gastric cancer. [German]. *Onkologe* 2008;14:389-95.

17. Al-Batran SE, Hartmann JT, Probst S, Schmalenberg H, Hollerbach S, Hofheinz R, et al. Phase III trial in metastatic gastroesophageal adenocarcinoma with fluorouracil, leucovorin plus either oxaliplatin or cisplatin: a study of the Arbeitsgemeinschaft Internistische Onkologie. *J Clin Oncol* 2008;26:1435-42.

18. Bouche O, Raoul JL, Bonnetain F, Giovannini M, Etienne PL, Lledo G, et al. Randomized multicenter phase II trial of a biweekly regimen of fluorouracil and leucovorin (LV5FU2), LV5FU2 plus cisplatin, or LV5FU2 plus irinotecan in patients with previously untreated metastatic gastric cancer: a Federation Francophone de Cancerologie Digestive Group Study - FFCD 9803. *J Clin Oncol* 2004;22:4319-28.

19. Dank M, Zaluski J, Barone C, Valvere V, Yalcin S, Peschel C, et al. Randomized phase III study comparing irinotecan combined with 5-fluorouracil and folinic acid to cisplatin combined with 5-fluorouracil in chemotherapy naive patients with advanced adenocarcinoma of the stomach or esophagogastric junction. *Ann Oncol* 2008;19:1450-7.

20. Jadad AR, Moore RA, Carroll D. Assessing the quality of reports of randomized clinical trials: is blinding necessary. *Control Clin Trials* 1996;17:1-12.

21. National Institute for Health and Clinical Excellence. *Sunitinib for the treatment of gastrointestinal stromal tumours*. London: NICE; 2009. Available from: http://guidance.nice.org.uk/TA179

22. Chabota I, LeLorierb J, Blacksteinc ME. The challenge of conducting pharmacoeconomic evaluations in oncology using crossover trials: The example of sunitinib for gastrointestinal stromal tumour. *Eur J Cancer* 2008;44:972-77.

23. Joint Formulary Committee. *British national formulary*. London: Pharmaceutical Press; 2009.

24. *NHS reference costs 2008/9*. London: Department of Health; 2010. Available from: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_111591

25. Millar DR, et al. A service evaluation to compare secondary care resource use between XELOX and FOLFOX-6 regimens in the treatment of metastatic colorectal cancer from a UK National Health Service (NHS) perspective. In: *ISPOR 11th Annual European Congress*; 2008; Athens, Greece. 2008. Poster EE8.

26. Curtis L. *Unit costs of health and social care 2009*. Canterbury: Personal Social Services Research Unit; 2009. Available from: http://www.pssru.ac.uk/pdf/uc/uc2009/uc2009.pdf

27. Ward S, Pilgrim H, Hind D. *Trastuzumab for the treatment of primary breast cancer in HER2 positive women: a single technology appraisal*. Sheffield: University of Sheffield, School of Health and Related Research; 2006. Available from: http://www.nice.org.uk/nicemedia/pdf/Breastcancer_trastuzumab_erg.pdf

28. Agrawal S, et al. Assessing the total costs of blood delivery to hospital oncology and haematology patients. *Curr Med Res Opin* 2006;22.

29. Liverpool Reviews and Implementation Group. *Erlotinib for the treatment of relapsed non-small cell lung cancer: Report commissioned by NHS R&D HTA Programme*. Liverpool: Liverpool Reviews and Implementation Group; 2006.

30. National Institute for Health and Clinical Excellence. *Advanced breast cancer: diagnosis and treatment*. London: NICE; 2009. Available from: http://guidance.nice.org.uk/CG81/Guidance/pdf/English

31. Philips Z, Bojke L, Sculpher M, Claxton K, Golder S. Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. *Pharmacoeconomics* 2006;24:355-71.

32. National Institute for Health and Clinical Excellence. *Guide to the methods of technology appraisal*. London: NICE; 2008.

33. Curran D, Pozzo C, Zaluski J, Dank M, Barone C, Valvere V. Quality of life of palliative chemotherapy naive patients with advanced adenocarcinoma of the stomach or esophagogastric junction treated with irinotecan combined with 5-fluorouracil and folinic acid: results of a randomised phase III trial. *Qual Life Res* 2009;18:853–61.

Appendix 1: Quality Assessment using the Philips economic modelling checklist

Table 32: Quality Assessment of Economic Model

Quality criterion	*Question(s)*	*Response (√, X, or NA)*	*Comments*
S1	Is there a clear statement of the decision problem?	√	Appendix A
	Is the objective of the evaluation and model specified and consistent with the stated decision problem?	√	
	Is the primary decision-maker specified?	√	NHS and Personal Social Services
S2	Is the perspective of the model stated clearly?	√	
	Are the model inputs consistent with the stated perspective?	√	
	Has the scope of the model been stated and justified?	√	
	Are the outcomes of the model consistent with the perspective, scope and overall objective of the model?	√	
S3	Is the structure of the model consistent with a coherent theory of the health condition under evaluation?	X	Adverse events were not related to the treatment duration. There is a lack of correlation between PFS and OS.
	Are the sources of data used to develop the structure of the model specified?	√	State that health states are typical of other economic evaluations of metastatic oncology.
	Are the causal relationships described by the model structure justified appropriately?	X	No justification for the lack of correlation between PFS and OS was provided.

S4	Are the structural assumptions transparent and justified?	X	No discussion of why Markov model was used.
	Are the structural assumptions reasonable given the overall objective, perspective and scope of the model?	√	Although did not consider utilities or adverse events on or off chemotherapy explicitly. Correlation between PFS and OS did not seem to be a major issue.
S5	Is there a clear definition of the options under evaluation?	√	
	Have all feasible and practical options been evaluated?	X	CX was not evaluated and it was assumed to be equal in effectiveness to ECX. Neither was CF.
	Is there justification for the exclusion of feasible options?	√	CX/F are have less than 10% usage by the NHS according to market research.
S6	Is the chosen model type appropriate given the decision problem and specified causal relationships within the model?	√	
S7	Is the time horizon of the model sufficient to reflect all important differences between options?	√	
	Are the time horizon of the model, the duration of treatment and the duration of treatment effect described and justified?	√	
S8	Do the disease states (state transition model) or the pathways (decision tree model) reflect the underlying biological process of the disease in question and the impact of interventions?	√	
S9	Is the cycle length defined and justified in terms of the natural history of disease?	X	One month cycle length unjustified.

D1	Are the data identification methods transparent and appropriate given the objectives of the model?	X	It is unclear how the utility for the progressive disease health state was chosen.
	Where choices have been made between data sources, are these justified appropriately?	√	The assumption that treatment effectiveness was equal between comparators was justified by lack of evidence proving significant difference.
	Has particular attention been paid to identifying data for the important parameters in the model?	√	
	Has the quality of the data been assessed appropriately?	√	
	Where expert opinion has been used, are the methods described and justified?	X	Only the number of oncologists consulted was reported.
D2	Is the data modelling methodology based on justifiable statistical and epidemiological techniques?	√	
D2a	Is the choice of baseline data described and justified?	√	
	Are transition probabilities calculated appropriately?	NA	An under-the-curve modelling approach was taken.
	Has a half-cycle correction been applied to both cost and outcome?	√	
	If not, has this omission been justified?	NA	
D2b	If relative treatment effects have been derived from trial data, have they been synthesised using appropriate techniques?	NA	
	Have the methods and assumptions used to extrapolate short-term results to final outcomes been documented and justified?	√	PFS and OS were extrapolated beyond trial data using parametric survival functions.
	Have alternative extrapolation assumptions been	√	Different parametric distributions were modelled and

	explored through sensitivity analysis?		tested.
	Have assumptions regarding the continuing effect of treatment once treatment is complete been documented and justified?	NA	No assumptions based since treatment effect is based on actual survival curves.
	Have alternative assumptions regarding the continuing effect of treatment been explored through sensitivity analysis?	NA	
D2c	Are the costs incorporated into the model justified?	√	
	Has the source for all costs been described?	√	
	Have discount rates been described and justified given the target decision-maker?	√	
D2d	Are the utilities incorporated into the model appropriate?	X	Assume increasing utilities while being treated by chemotherapy.
	Is the source for the utility weights referenced?	√	
	Are the methods of derivation for the utility weights justified?	√	
D3	Have all data incorporated into the model been described and referenced in sufficient detail?	X	All data has been described in sufficient detail except for the rates of anorexia for treatment regimens ECF, ECX, EOX.
	Has the use of mutually inconsistent data been justified (i.e. are assumptions and choices appropriate)?	X	It is assumed that the OS hazard ratio for HCF vs HCX was 1.15 yet the hazard ratio for PFS for HCF vs HCX was assumed to be 1. The same applies for ECF vs ECX.
	Is the process of data incorporation transparent?	√	
	If data have been incorporated as distributions, has the	X	The distributions are described but not justified.

	choice of distribution for each parameter been described and justified?		
	If data have been incorporated as distributions, is it clear that second order uncertainty is reflected?	√	
D4	Have the four principal types of uncertainty been addressed?	X	Have not discussed structural uncertainty or heterogeneity.
	If not, has the omission of particular forms of uncertainty been justified?	X	
D4a	Have methodological uncertainties been addressed by running alternative versions of the model with different methodological assumptions?	√	Test survival distributions.
D4b	Is there evidence that structural uncertainties have been addressed via sensitivity analysis?	X	
D4c	Has heterogeneity been dealt with by running the model separately for different subgroups?	X	Have not tested different age groups or definitions of HER 2 positivity.
D4d	Are the methods of assessment of parameter uncertainty appropriate?	X	The uncertainty surrounding the assumptions of hazard ratios based on a lack of evidence was not incorporated in the probabilistic sensitivity analysis so the cost-effectiveness acceptability curves are meaningless.
	If data are incorporated as point estimates, are the ranges used for sensitivity analysis stated clearly and justified?	X	The ranges were not stated clearly in the MS though they were in the model, and some of the effectiveness
C1	Is there evidence that the mathematical logic of the model has been tested thoroughly before use?	X	We found some calculation errors in the model. Including a negative number of patients in the progressive disease state.

C2	Are any counterintuitive results from the model explained and justified?	X	The negative number of patients in the disease progression state was not explained.
	If the model has been calibrated against independent data, have any differences been explained and justified?	NA	
	Have the results of the model been compared with those of previous models and any differences in results explained?	NA	No other models were identified in their review of the literature

Appendix 2: Details of ERG models

Revised Model

The ERG developed a revised model as explained in section 6.2. This included correcting for minor errors and logical inconsistencies. The details of the revisions are listed below.

- Inflate AE costs to 2010 – changed ICER
    - Calculated inflated AE costs – sheet: Admin-Pharm Mon, cells: G3:J91
    - Input new inflated costs in tables – sheet: AE events, cells: D28:D31, D33:D35
- Update N for ECF/ECX/EOX – did not account for table foot notes
    - Changed formula to match table 3 pg 44 of Cunningham 2008 – sheet AE events, cell F16 – from 234 to 236, cell F19 – from 234 to 236, cell F22 – from 234 to 236, cell N5 – from 234 to 229, cell N8 – from 234 to 229, cell N11 – from 234 to 229, cell N16 – from 227 to 232, cell N19 – from 227 to 232, cell N22 – from 227 to 232
- Logical constraint on progression
    - Changed formulas – sheets: HCX, HCF, EOX, ECX, ECF, cell: O6:O186 – from '=R6-L6' to '=IF(L6>R6,0,R6-L6)', cell: P6:P186 – from '=S6-M6' to '=IF(M6>S6,0,S6-M6)', cell: Q6:Q186 – from '=T6-N6' to '=IF(N6>T6,0,T6-N6)'
- Cycles per month by actual treatment
    - Recalculated average by treatments in regimen – sheet: Dose Table, cells A24:J24
    - Replaced cycle means(1.29) with new calculations – sheet: Admin-Pharm-Mon, cells: C6:L6
- Ambulatory pump changing F Mono to be the same as ECF – sheet: Admin-Pharm-Mon, cell: L16 – from 1 to 3
- District Nurse Visit changing F Mono to be the same as ECF – sheet: Admin-Pharm-Mon, cell: L17 – from 1 to 2
- Correct cell X6 on sheet EOX – results not calculated from this cell
    - Changed formula to =SUM(U6:V6) +W6+ c_ae_EOX
    - from =SUM(U6:V6) +W6+ c_ae_comx

ERG scenario analyses

The ERG conducted several scenario analyses as described in section 6. Excel files were created to perform these analyses. The changes made to the ERG revised model to create each files are detailed below, and the scenario analyses associated with these files are listed.

File 'Herceptin Gastric NICE model.NH1'

This file was used to evaluate different distributional assumptions for PFS and OS.

- Sheet: HCX, cells: R6:R18 changed to '=CHOOSE(distn,EXP(-olnw*$B6^ognw), EXP(-olne*$B6^ogne),(1 / (1 + olnl*$B6^ognl)), IF($B6 = 0,1, (1 - NORMDIST(((LN($B6) - olnn) / ognn),0,1,TRUE))), EXP((olngo/ogngo)*(EXP(ogngo*$B6)-1)), 1, 'KM OS'!$**B**4)'
- Sheets:EOX, ECX and ECF, cells: O6:O18 changed to '=CHOOSE(distn,EXP(-(olcw*HR_OS_ECF)*$B6^ogcw), EXP(-(olce*HR_OS_ECF)*$B6^ogce),(1 / (1 + olcl*($B6*HR_OS_ECF)^ogcl)), IF($B6 = 0,1, (1 - NORMDIST(((LN($B6*HR_OS_ECF) - olcn) / ogcn),0,1,TRUE))), EXP(((HR_OS_ECF*olcgo)/ogcgo)*(EXP(ogcgo*$B6)-1)),1, 'KM OS'!$**G**4)'
- Sheets: HCX and HCF, cells R19:R186 changed to '=CHOOSE(distn,EXP(-olnw*$B19^ognw),EXP(-olne*$B19^ogne),(1/(1+olnl*$B19^ognl)),IF($B19=0,1,(1-NORMDIST(((LN($B19)-olnn)/ognn),0,1,TRUE))),EXP((olngo/ogngo)*(EXP(ogngo*$B19)-1)),1,**(EXP((olnw*$B18^ognw)-(olnw*$B19^ognw)))*R18)**
- Sheets: EOX, ECX and ECF, cells: O19:O186 changed to '=CHOOSE(distn,EXP(-(olcw*HR_OS_EOX)*$B19^ogcw), EXP(-(olce*HR_OS_EOX)*$B19^ogce),(1 / (1 + olcl*($B19*HR_OS_EOX)^ogcl)), IF($B19 = 0,1, (1 - NORMDIST(((LN($B19*HR_OS_EOX) - olcn) / ogcn),0,1,TRUE))), EXP(((HR_OS_EOX*olcgo)/ogcgo)*(EXP(ogcgo*$B19)-1)),1,**(EXP((olcw*$B18^ogcw)-(olcw*$B19^ogcw)))*O18)**'
-

File 'Herceptin Gastric NICE model.NH2'

This file was used to evaluate the model using survival curves for the FAS population instead of the EMEA population.

- The variance-covariance matrices for the FAS population from the Excel spreadsheet embedded in A5 of the manufacturers' responses to clarifications was copied and pasted over the original data in the spreadsheets labelled Exponential, Gompertz, Weibull, Log Logistic.
-

File 'Herceptin Gastric NICE model.T1'

This file was used to evaluate different hazard ratios for HCF vs HCX for PFS and OS.

- Sheet: HCF, cells: L7:L18 changed to '=L6*(1-HR_PFS_HCF*(HCX!L6-HCX!L7)/HCX!L6)'
- Sheet: HCF, cells: L19:L186 changed to (with the text in bold added to the formula) '=CHOOSE(Distn_PFS,EXP(-(plnw*HR_PFS_HCF)*$B19^pgnw),EXP(-(plne*HR_PFS_HCF)*$B19^pgne),(1/(1+plnl*($B19*HR_PFS_HCF)^pgnl)),IF($B19=0,

1,(1-NORMDIST(((LN($B19*HR_PFS_HCF)-plnn)/pgnn),0,1,TRUE))),EXP(((plngo*HR_PFS_HCF)/pgngo)*(EXP(pgngo*$B19)-1)),1,(EXP((plnw**HR_PFS_HCF** *$B18^pgnw)-(plnw**HR_PFS_HCF** *$B19^pgnw)))*L18)'

The formula '=L6*(1-HR_PFS_HCF*(HCX!L6-HCX!L7)/HCX!L6)' is derived as follows:

The hazard ratio, φ, is

$$\varphi = \frac{h_F}{h_X}$$

Where h_F is the hazard rate for HCF and h_X is the hazard rate for HCX.

For continuous functions the hazard rate is

$$h(t) = \frac{-S'(t)}{S(t)}$$

For discrete functions, it is approximated by

$$h(t_0) = \frac{-\left(S(t_1) - S(t_0)\right)}{S(t_0)}$$

The survival functions for HCF and HCX are stated as S_F and S_X, respectively. Using the above formulae, the value of the survival function for HCF at time $t_1, S_F(t_1)$, is obtained as follows:

$$h_F = \varphi \times h_X$$

$$\frac{S_F(t_0) - S_F(t_1)}{S_F(t_0)} = \varphi \times \frac{S_X(t_0) - S_X(t_1)}{S_X(t_0)}$$

$$S_F(t_1) = S_F(t_0) \times \left(1 - \varphi \times \left(\frac{S_X(t_0) - S_X(t_1)}{S_X(t_0)}\right)\right)$$

File 'Herceptin Gastric NICE model.T2'

This file was used to evaluate a CX comparator with the effectiveness of ECX and the same cisplatin dose as in REAL-2, but without the costs of epirubicin.

- Sheet: Drug cost, Cells: E25 changed to 0
- Sheet: Admin-Pharm-Mon, Cells: C28,D28,E28 changed to 0

File 'Herceptin Gastric NICE model.T3'

This file was used to evaluate a CX comparator with a cisplatin dose as in ToGA.

- Sheet: Drug cost, Cells: E25 changed to 0
- Sheet: Admin-Pharm-Mon, Cells: C28,D28,E28 changed to 0
- Sheet: Regimen Drug Costs, Cells: D6, F6 changed to 86%, Cells: D14, F14 changed to 137

File 'Herceptin Gastric NICE model.T4'

This file was used to evaluate a hazard ratio of 0.96 for ECX vs CX.

- Sheet: ECX, cells: I7:I18 changed to '=I6*(1-HR_PFS_ECX*('KM PFS'!J4-'KM PFS'!J5)/'KM PFS'!J4)'
- Sheet: ECX, cells: I19:I186 changed to (with the text in bold added to the formula) '=CHOOSE(Distn_PFS,EXP(-(plcw*HR_PFS_ECX)*$B19^pgcw),EXP(-(plce*HR_PFS_ECX)*$B19^pgce),(1/(1+plcl*($B19*HR_PFS_ECX)^pgcl)),IF($B19=0,1,(1-NORMDIST(((LN($B19*HR_PFS_ECX)-plcn)/pgcn),0,1,TRUE))),EXP(((plcgo*HR_PFS_ECX)/pgcgo)*(EXP(pgcgo*$B19)-1)),1,(EXP((plcw**HR_PFS_ECX***$B18^pgcw)-(plcw**HR_PFS_ECX***$B19^pgcw)))*I18)'

File 'Herceptin Gastric NICE model.T5'

This file was used to evaluate a hazard ratio of 0.87 for EOX vs ECX.

- Sheet: EOX, cells: I7:I18 changed to ' =I6*(1-HR_PFS_EOX*(ECX!I6-ECX!I7)/ECX!I6)'
- Sheet: EOX, cells: I19:I86 changed to (with the text in bold added to the formula) ' =CHOOSE(Distn_PFS,EXP(-(plcw*HR_PFS_EOX)*$B19^pgcw),EXP(-(plce*HR_PFS_EOX)*$B19^pgce),(1/(1+plcl*($B19*HR_PFS_EOX)^pgcl)),IF($B19=0,1,(1-NORMDIST(((LN($B19*HR_PFS_EOX)-plcn)/pgcn),0,1,TRUE))),EXP(((plcgo*HR_PFS_EOX)/pgcgo)*(EXP(pgcgo*$B19)-1)),1,(EXP((plcw**HR_PFS_EOX***$B18^pgcw)-(plcw**HR_PFS_EOX***$B19^pgcw)))*I18)'

File 'Herceptin Gastric NICE model.U1'

This file was used to evaluate testing a range of +/-20% from the base case utility values for PFS and OS.

- Sheet: One-way, cell D5 changed to '=C5***0.8**'
- Sheet: One-way, cell E5 changed to '=C5***1.2**'
- Sheet: One-way, cell D8 changed to '=C8***0.8**'
- Sheet: One-way, cell E8 changed to '=C8***1.2**'
- Re-run sensitivity analysis

File 'Herceptin Gastric NICE model.U2'

This file was used to evaluate a daily decrement to utility in PFS.

- Sheet: Model Inputs, cell E43 changed to '=IF(uPFS_Treat_yes = 1, -0.000141, -0.0000095890410958904) * uPFS_Day_Yes'
- Sheet: HCX , cells AH6:AH186 changed to '=Q6*IF(psa=0,u_prog**+uPFS_Day*B6*day2mon**,pu_prog**+uPFS_Day*B6*day2mon**)'
- Sheet: HCF, cells AG6:AG186 changed to '=Q6*IF(psa=0,u_prog**+uPFS_Day*B6*day2mon**,pu_prog**+uPFS_Day*B6*day2mon**)'
- Sheets: EOX, ECX and ECF, cells AC6:AC186 changed to '=N6*IF(psa=0,u_prog**+uPFS_Day*B6*day2mon**,pu_prog**+uPFS_Day*B6*day2mon**)'

File 'Herceptin Gastric NICE model.U3'

This file was used evaluate different utilities for being on and off treatment.

- Sheets: HCX and HCF, column AG to

'=N6*MIN(IF(psa=0,u_PFS+uPFS_Day*B6*day2mon,pu_pfs+puPFS_Day*B6*day2mon),uPFS_Cap)
***(1-MAX(E6,H6,K6))**
+N6*MIN(IF(psa=0,u_PFS+uPFS_Day*B6*day2mon,pu_pfs+puPFS_Day*B6*day2mon),uPFS_Cap)
***MAX(E6,H6,K6)*0.95**'

- Sheets: EOX, ECX and ECF, column AB to

'=K6*MIN(IF(psa=0,u_PFS+uPFS_Day*B6*day2mon,pu_pfs+puPFS_Day*B6*day2mon),uPFS_Cap)
***(1-MAX(E6,H6))**
+K6*MIN(IF(psa=0,u_PFS+uPFS_Day*B6*day2mon,pu_pfs+puPFS_Day*B6*day2mon),uPFS_Cap)
***MAX(E6,H6)*0.95**

File 'Herceptin Gastric NICE model.R1'

This file was used to evaluate a range of HER2 positivity costs.

- Sheet: Admin-Pharm-Mon, cell D73 change to '=**1/5%**'
- Sheet: Admin-Pharm-Mon, cell D73 change to '=**1/30%**'

File 'Herceptin Gastric NICE model.R2'

This file was used to evaluate the extra cost of reducing a percentage of IHC and FISH tests due to test failure.

- Sheet: Admin-Pharm-Mon, cell D72 change to '=3280***1.06**/(212+373)'
- Sheet: Admin-Pharm-Mon, cell D73 change to '=388***1.094**/585'

File 'Herceptin Gastric NICE model.R3'

This file was used to evaluate the cost of parallel testing of IHC and FISH rather than sequential testing.

- Sheet: Admin-Pharm-Mon, cell D73 change to '=**D72**'

File 'Herceptin Gastric NICE model.R4'

This file was used to evaluate a reduced frequency of cardiac monitoring while on epirubicin to the same frequency as applies with trastuzumab.

- Sheet: Admin-Pharm-Cost, Cells: C28:E38 changed to 0.33

File 'Herceptin Gastric NICE model.R5'

This file was used to evaluate increased adverse event costs for HCX by changing the adverse event costs of HCX to equal those for ECX.

- Sheet: AE events, cell:G4 changed to '=c_ae_ECX'