Evidence Review Group Report commissioned by the NHS R&D HTA Programme on behalf of NICE

Gemcitabine for metastatic breast cancer

Produced by Southampton Health Technology Assessments Centre

Authors J Jones, Senior Research Fellow
A Takeda, Research Fellow
SC Tan, Research Fellow
K Cooper, Research Fellow
E Loveman, Senior Research Fellow
A Clegg, Director of SHTAC
N Murray, Consultant in Oncology, Southampton University Hospitals Trust

Correspondence to Dr Andrea Takeda
Southampton Health Technology Assessments Centre
Wessex Institute for Health Research and Development
University of Southampton
Mailpoint 728, Boldrewood
Southampton SO16 7PX

Date completed 24th July 2006

Acknowledgements
We thank members of the Resource and Information Service at the Wessex Institute for Health Research and Development, and J. Bryant of SHTAC for acting as internal editor.

Conflicts of Interest:
Dr. N. Murray is involved with a phase 2 trial of gemcitabine and carboplatin for locally advanced or metastatic breast cancer.

Reference to any academic or commercial in confidence data presented in the manufacturer's submission is removed from this document.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AE	Adverse event
ASCO	American Society of Clinical Oncology
BNF	British National Formulary
BPI	Brief pain inventory
BSA	Body surface area
CEA	Cost effectiveness analysis
CEAC	Cost effectiveness acceptability curve
CI	Confidence interval
CUA	Cost utility analysis
D	docetaxel
ERG	Evidence Review Group
GT	Gemcitabine and paclitaxel combination treatment
HTA	Health Technology Assessment
HUI	Health utility index
ICER	Incremental cost effectiveness ratio
ITT	Intention to Treat
JHQG	The B9E-MC-JHQG trial
KPS	Karnofsky Performance Status
MBC	metastatic breast cancer
MS	Manufacturer's submission
OS	Overall survival
PFS	Progression-free survival
PSA	Probabilistic sensitivity analysis
PSS	Personal Social Services
QALY	Quality adjusted life year
QuOROM	Quality Of Reporting Of Meta-analyses
RCT	Randomised Controlled Trial
RSCL	Rotterdam symptom checklist
SG	Standard gamble
SPC	Summary of product characteristics
SR	Systematic review
STA	Single Technology Appraisal
T	Paclitaxel (Taxol) treatment
TtDPD	Time to documented progression of disease
NHS TFR	NHS Trust Financial Returns
TTO	Time trade off
TTP	Time to progression

1 SUMMARY

1.1 Scope of the submission

- The submission's scope is the use of gemcitabine with paclitaxel for the first-line treatment of metastatic breast cancer in patients who have already received chemotherapy treatment with an anthracycline, compared to current standard of care. This reflects the licensed indication, and is an appropriate question for the NHS within the context of the available evidence.

1.2 Summary of submitted clinical effectiveness evidence

- The clinical evidence for gemcitabine with paclitaxel compared with paclitaxel monotherapy as a treatment for MBC comes from the B9E-MC-JHQG trial (referred to here as the JHQG trial), which was published in conference abstracts[1-3] in 2003/4 but has not yet been fully published. The data in the industry submission comes from the as yet unpublished trial, so is mostly marked as Commercial in Confidence. Results from two other published trials are included in the submission to provide a comparison with docetaxel monotherapy[4] and docetaxel/capecitabine combined therapy[5].
- The JHQG trial compared gemcitabine/paclitaxel (GT) with paclitaxel (T) in patients with metastatic breast cancer. The trial by Jones and colleagues[4] compared docetaxel monotherapy with paclitaxel, and the trial by O'Shaughnessy and colleagues[5] compared docetaxel monotherapy with docetaxel/capecitabine combination therapy.
- Overall survival, the primary outcome measure for the JHQG trial, was approximately 3 months longer for the gemcitabine/paclitaxel arm (18.5 months in Albain abstract[1]) (18.6 months in MS) than for the paclitaxel arm (15.8 months).[1] This difference is of borderline statistical significance (p=0.0489), but represents a clinically significant difference to patients. Results from the JHQG trial suggest that gemcitabine added to paclitaxel also improves tumour response and time to documented progression of disease, compared with paclitaxel monotherapy. Haematological serious adverse events were more common in the gemcitabine/paclitaxel arm than in the paclitaxel monotherapy arms.
- In the absence of any formal methods of indirect comparison, there is insufficient robust evidence to compare the relative effectiveness of gemcitabine/paclitaxel with docetaxel monotherapy or docetaxel/capecitabine combination therapy.

1.3 Summary of submitted cost effectiveness evidence

- The cost-effectiveness analysis in the MS uses a Markov state transition model to estimate the effect of treatment with five different chemotherapy regimes, adopting a three year time horizon. Base case results are presented, with docetaxel monotherapy as the comparator for all interventions – assuming that docetaxel is the standard of care for UK practice.
- Additional scenario analyses are presented using alternative comparators and for a price reduction for paclitaxel once the patent expires.
- Treatment effects in the model are derived from pooling data from 15 clinical trials – only three of these are discussed in the clinical effectiveness section of the MS. No formal assessment of trial comparability or any quality assessment has been presented.
- Health state utilities for different stages of disease progression and for patients experiencing treatment-related toxicity are used in the model to derive quality adjusted life expectancy with each treatment.
- The base case cost effectiveness estimate for gemcitabine/paclitaxel, relative to docetaxel, is £17,168 per QALY. When longer survival with docetaxel is assumed, in a sensitivity analysis, the ICER increases to approximately £30,000 per QALY. Probabilistic sensitivity analysis estimates a 70% probability of gemcitabine/paclitaxel being cost effective relative to docetaxel, at an arbitrary threshold willingness-to-pay of £35,000.
- The lack of formal quality assessment or assessment of the comparability of trials included in the input data, and questionable validity of the indirect comparison method adopted, leads to considerable uncertainty over the cost-effectiveness of gemcitabine/paclitaxel. An illustrative analysis using a different method for indirect comparison presented in this report produces an ICER of £45,811 per QALY for gemcitabine/paclitaxel relative to docetaxel.

1.4 Commentary on the robustness of submitted evidence

1.4.1 Strengths

- The structure of the manufacturer's economic model is appropriate for the stated decision problem, and reflects accepted methodology.

1.4.2 Weaknesses

- The manufacturer performed a systematic review, which identified two abstracts (and missed a third) reporting interim results of the JHQG trial. However, the data in the

manufacturer's submission is based on Commercial in Confidence data which is due to be published later this year.

- Although a systematic review was carried out, there is contradiction and a lack of methodological rigour regarding a number of the references included for the economic evaluation. The ERG therefore considers that, although the model's structure is appropriate, selection bias could potentially have affected the data inputs for the economic model.
- The attempted indirect comparison in the clinical effectiveness section simply tabulates data from the JHQG trial and the two comparator trials. It might have been possible to perform a formal statistical indirect comparison of the JHQG trial with that by Jones and colleagues[4] (docetaxel monotherapy vs. paclitaxel), since they have a common comparator arm. However, differences in the trials' patient characteristics may have invalidated such an approach.

1.4.3 Areas of uncertainty

- In the absence of an RCT directly comparing gemcitabine with docetaxel, there does not appear to be sufficient evidence to compare the relative effectiveness of these treatments.

1.5 Key issues

- The evidence for gemcitabine's clinical effectiveness comes from an RCT comparing gemcitabine/paclitaxel with paclitaxel. However, the economic evaluation uses docetaxel as the comparator in the reference case.
- The manufacturer suggests that gemcitabine should be considered as one option for first line therapy for MBC in some patients, but does not appear to advocate that it should replace any of the current taxane treatments.

2 INTRODUCTION TO ERG REPORT

This report is a critique of the manufacturer's submission (MS) to NICE from Eli Lilly on the clinical effectiveness and cost effectiveness of gemcitabine with paclitaxel for metastatic breast cancer. It identifies the strengths and weakness of the MS. A clinical expert was consulted to advise the Evidence Review Group (ERG) and to help inform this review.

Clarification on some aspects of the MS was requested from the manufacturer by the ERG via NICE on 31st May 2006. A response from the manufacturer via NICE was received by the ERG on 19th June and this has been included as an Appendix to the ERG report. A CD of additional analyses was also received by the ERG on 26th June 2006. In an attempt to keep

this report concise, minor points of clarification have not been discussed in the main text. Further analyses supplied by the manufacturer and responses to the ERG's key questions are discussed in the text where appropriate.

3 BACKGROUND

3.1 Critique of manufacturer's description of underlying health problem

Breast cancer is classified into four clinical stages. Stages I and II are also known as primary or early breast cancer, and stages III and IV represent advanced breast cancer. Stage IV is metastatic disease, characterised by the spread of secondary tumours to distant sites. A small proportion of incident breast cancers present as stage IV, *i.e.* they have overt metastases at the time of diagnosis.

The manufacturer provides a clear and concise overview of the disease (p. 17-18).The MS states that approximately 40% of patients treated for early breast cancer will relapse and develop metastatic breast cancer (MBC), but does not give any figures for the percentage of patients who present with stage IV disease at first diagnosis. The latter group are described as being unsuitable for treatment with gemcitabine in the flowchart in MS Appendix 1 (p.155), since they will not have received prior anthracycline therapy.

3.2 Critique of manufacturer's overview of current service provision

The MS contains a concise, accurate description of current treatment options for metastatic breast cancer. Taxanes are recommended as first-line therapy for MBC in patients who received treatment with anthracycline drugs at earlier stages of the disease. The MS places gemcitabine as a first line therapy since it is licensed for use in conjunction with paclitaxel (a taxane). Paclitaxel monotherapy, docetaxel monotherapy and docetaxel/capecitabine dual therapy are the licensed taxane first-line treatments for MBC, and the manufacturer identifies these as appropriate comparators. The MS briefly describes other licensed therapies currently used to treat MBC, such as vinorelbine and capecitabine. Since these are second line therapies and the manufacturer is marketing gemcitabine as a first line therapy, it seems appropriate to exclude them as comparators. However, the ERG's clinical advisor indicates that the positioning of particular agents in sequential lines of therapy is rather blurred, and that there is value in looking at other combinations as second line therapy.

Currently, gemcitabine is not widely used in the treatment of MBC. The manufacturer presents some commercial in confidence figures of current and projected estimates of use, but does not provide detailed supporting evidence for these figures. The manufacturer estimates current usage of gemcitabine within the NHS to be around 2% of first-line chemotherapy treatments in the metastatic setting (p.17), but estimates that this will increase to 15%. No supporting evidence is given for the figure of 15%.

4 Critique of manufacturer's definition of decision problem

4.1 Population

The study population described in the decision problem is people with MBC who have received anthracycline-based treatment in the adjuvant/neoadjuvant setting. This reflects UK clinical practice for the treatment of MBC, and is appropriate for the NHS. The MS does not include any further detail on the UK MBC patient population, such as mean age at diagnosis or mean number/location of metastatic sites, against which to compare the characteristics of patients in the clinical trial. The manufacturer states on p. 12 of the MS that it was not considered appropriate to conduct sub-group analysis on the trial population. This was because the trial population was homogeneous, and the study was not powered to detect small differences in sub-groups between treatment arms. The MS does not state whether women with HER2 positive tumours were included in the trial, and the baseline characteristics in table 4 of the MS do not mention this.

4.2 Intervention

Gemcitabine is an antimetabolite, and works by preventing normal cellular division. It is licensed for use in combination with paclitaxel for the treatment of metastatic breast cancer, in patients who received anthracyline-based treatment in the adjuvant/neoadjuvant setting. Paclitaxel is a taxane, so is one of the drugs preferred for first line treatment of breast cancer.

The JHQG trial reported in the MS reflects the licensed indication for this drug as first line therapy. Currently, gemcitabine is not frequently used in UK clinical practice. Its place in the treatment pathway as a first-line therapy for MBC is justified by its combination with a taxane, but it is not as widely used as taxane monotherapies or other taxane combination therapies. Gemcitabine is currently more likely to be used as a later line of therapy, once other treatments have been unsuccessful. Commercial in confidence figures presented by

the manufacturer suggest that it had only 1% of the market share in 2005. The manufacturer predicts that this will increase to 3% in 2006, but presents no data to support this. The manufacturer's 2006 predictions are that 55% of eligible patients would receive taxane monotherapies, 3% would receive gemcitabine with paclitaxel, 24% would receive other taxane combination therapies, and that 18% would receive other chemotherapies. However, the ERG's clinical advisor indicated that 79% use of taxane or taxane combination is too high. Taxanes are much less likely to be used in the elderly, but will be the treatment of choice for the majority of younger, fitter patients. The statement of clinical effectiveness (p.67) states that: "it is not anticipated that GT will be used in all patients as a substitute for the standard of care, but will be used in a small proportion of patients, who are younger, fitter and looking for a higher level of efficacy than a monotherapy can provide, without the toxicity usually associated with a combination regimen." The manufacturer does not provide any rationale for selecting this particular patient group as the target population for gemcitabine. The ERG's clinical advisor indicates that the standard of care would be docetaxel (and exceptionally docetaxel/capecitabine), and that the stated target patient group is appropriate for gemcitabine/paclitaxel treatment. The MS suggests that gemcitabine should be considered as one option for first line therapy for MBC in some patients, but does not appear to advocate that it should replace any of the current treatments.

4.3 Comparators

The manufacturer restricts possible comparators to those treatments which are licensed to be used as first line therapy for anthracycline pre-treated MBC. The manufacturer has restricted the comparator to taxane drugs, specifically to paclitaxel monotherapy, docetaxel monotherapy, and docetaxel/capecitabine combination therapy. These are all valid comparators, and are appropriate for the first-line treatment of MBC within the NHS. The manufacturer has specifically excluded vinorelbine and capecitabine monotherapies as comparators, since they are only used as second-line treatments for MBC (p.12-13). This reflects NICE guidance on the second-line use of vinorelbine[6] and capecitabine.[7] Since treatments are not always used in such clearly delineated lines of therapy in clinical practice, there could be some value in comparing gemcitabine/paclitaxel with second line therapies.

Trastuzumab was excluded as a comparator as it is only given to women with HER2 positive tumours. However, it is not clear from the MS whether the JHQG trial included any women with HER2 positive tumours. All patients with HER2 positive disease should be receiving trastuzumab as a minimum, but it would not be a valid direct comparator for gemcitabine. Trastuzumab in combination with another agent could be a valid combination to compare against gemcitabine/paclitaxel with trastuzumab.

4.4 Outcomes

The MS lists five clinical outcome measures and two measures of cost effectiveness on p.13. These are all appropriate and clinically meaningful outcomes, and there are no other valid outcomes which the ERG would have expected to be included. Clinical outcome measures are: overall survival (primary outcome); time to disease progression; tumour response; health related quality of life [measured by brief pain inventory (BPI) and Rotterdam Symptom checklist (RSCL)] and adverse effects. Incremental cost per quality-adjusted life year and per life year gained are used as measures of cost-effectiveness.

5 CLINICAL EFFECTIVENESS

5.1 Critique of manufacturer's approach

5.1.1 Description of manufacturer's search strategy and comment on whether the search strategy was appropriate.

5.1.1.1 Clinical effectiveness searches

The sources used by the manufacturer for the search (Embase, Medline, Medline in Process, NICE, Cochrane, NCCHTA, ASCO, NHS CRD, Internal databases, internet), are appropriate and comprehensive. Additional databases that could have been used to obtain the clinical evidence are Biosis and Web of Science, although it is unlikely that they would have yielded any additional key results. The manufacturer has documented the use of ASCO, which is the key source of information for sourcing ongoing cancer trials. The search documentation could have been widened or clarified to include mention of sources such as the national research register, controlled clinical trials, clinicaltrials.gov, in order to track any ongoing trials.

The search strategies in Appendix 6 are transparent, fully documented and reproducible. The ERG reproduced components of the search on 23rd May 2006. The main search (Search 1, MS p.175-176) yielded similar results, but the ERG identified 457 citations with the paclitaxel search (after amending to take account of extra references since November 2005), compared with 84 in the manufacturer's search (p.178). The manufacturer's Embase search was from 1988, whereas the ERG's was from 1980, but searches were otherwise as similar as was feasible. A brief scan of the identified references suggested that none of the 'extra' references were relevant to the systematic review.

The MS (p.27) states that the search included data up until the 28th November 2005. For the sake of completeness, the ERG considers that an update search should have been re-run for all the study drugs.

5.1.1.2 Cost effectiveness searches

The searches for cost-effectiveness studies are not clearly described in the MS. The searches described in the clinical effectiveness section (p. 26) appear to have covered cost-effectiveness, since the reviewers identified studies from these which were only applicable to the economic model. However, the cost-effectiveness section (p. 69) then describes a separate search (dated 8th September 2005) of all the key databases. This search is not well documented, and only basic keywords are included in table 19 of the MS (p.70). The citations identified by this search are different from those identified in the earlier stage of the review, and appear to have been used to inform the design of the economic model.

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

The MS describes an appropriate method of identifying and screening references for inclusion in the systematic review. Three independent reviewers applied pre-specified inclusion/exclusion criteria to citations identified by the searches, and discussed any unclear references until agreement was reached.

The MS specified the following inclusion criteria for the systematic review of the literature (p.28):

1. study design –original studies reporting final results of phase III clinical trials;
2. interventions –gemcitabine/paclitaxel, docetaxel/capecitabine, paclitaxel monotherapy or docetaxel monotherapy;
3. population –patients with MBC who have been treated and failed on prior anthracycline treatment in an adjuvant or neoadjuvant setting;
4. outcome measures - no outcome measures were specified in the inclusion/exclusion criteria.

Phase I and II trials, observational studies, letters to the editor and editorials were excluded from the systematic review. The manufacturer did not state whether published systematic reviews would be considered in the review, and did not state clearly whether conference abstracts would be included or excluded. The specified inclusion/exclusion criteria were appropriate and reflect the information given in the decision problem.

Searches were divided into Search 1 (Medline and Embase) and Search 2 (company database, NICE, Cochrane database etc.). QUOROM flow charts for these searches are provided in figures 2 and 3 (p.32) of the MS, suggesting that a total of 15 studies were to be included in the review. However, table 2 (p.30) lists only 11 studies identified for inclusion and table 3 (p.31) lists four other studies which were excluded. The ERG therefore assumes that the flow charts give the number of studies identified as 'shortlisted' rather than the final total of included studies.

The information given on p. 28 suggests that searches were carried out for clinical and cost effectiveness reviews, and that screening of identified studies then sorted references into those used for the clinical effectiveness review and those used for analysis of cost effectiveness. Table 2 lists 11 references, of which 4 were to be included in both the clinical-effectiveness review and the economic evaluation, and 7 of which were only for inclusion in the economic evaluation. However, a separate search (dated 8 September 2005) is described in the cost-effectiveness section (p. 69), and a different set of papers is identified. Section 3.1.2 describes 7 studies identified by this separate search for cost effectiveness studies, none of which appear in table 2. None of these 7 studies appear to have been used in the evaluation of cost-effectiveness. Table 21 (p. 85) shows median overall survival from a number of studies. These studies include papers explicitly excluded from the review in table 3 (p.31), such as that by Winer and colleagues 2004[8], and other papers which were not mentioned in either of the sections discussing the systematic review (e.g. Extra and colleagues[9]). This will be addressed in more detail in Section 6.4.4.

5.1.3 Identified studies

Two abstracts of an RCT comparing gemcitabine/paclitaxel combination versus paclitaxel (registration trial) were included in the systematic review[1;3]. However, these only present interim results from the trial, and therefore do not actually meet the inclusion criteria for the systematic review, which stated that only studies reporting final results of phase III clinical trials were to be included. No RCTs comparing gemcitabine with the other relevant comparators (docetaxel or capecitabine/docetaxel in combination) were identified. The MS therefore included two additional RCTs to form indirect comparisons. One RCT reported docetaxel versus paclitaxel[4], and the other reported capecitabine/docetaxel combination versus docetaxel.[5] The patients in these trials do not appear to be directly comparable with the JHQG population, as they had received prior chemotherapy for MBC whereas the JHQG patients received the study treatment as first line MBC therapy.

Table 1 Characteristics of included studies

Study	Methods	Participants	Outcomes
JHQG trial, reported in 2 abstracts[1;3]	*Design*: interim analysis of an RCT *Interventions*: Group A: 1250 mg/m^2 gemcitabine + 175 mg/m^2 paclitaxel Group B: 175 mg/m^2 paclitaxel *Number of centres*: 98 *Median duration of treatment*: 6 cycles (group A), 5 cycles (group B) *Sponsor*: Eli Lilly	*Inclusion criteria*: MBC previously treated with anthracyclines; no prior chemotherapy for metastatic breast cancer; score ≥70 on activities of daily living scale (Karnofsky Performance Status (KPS)) *Numbers*: 529 participants. Group A: 267 Group B: 262 *Median age (range):* Group A 53.5 (26-83) years, Group B 52.9 (26-75) years	*Primary outcome*: overall survival *Secondary outcomes*: time to documented progression of disease; progression-free survival; tumour response rate; duration of response; quality of life, measured by Brief Pain Inventory (BPI) and Rotterdam Symptom Checklist (RSCL); adverse events/toxicites *Length of follow-up:* median follow-up 15.6 months for overall survival outcome
Jones et al.[4]	*Design*: RCT *Interventions*: Group A: 100 mg/m^2 docetaxel Group B: 175 mg/m^2 paclitaxel *Number of centres*: 53 *Median duration of treatment*: 6 cycles (group A), 4 cycles (group B) *Sponsor*: Aventis pharmaceuticals	*Inclusion criteria*: adenocarcinoma of the breast and disease progression after one prior chemotherapy regimen for locally advanced or MBC, or with locally advanced or MBC that progressed during or within 12 months of completing an adjuvant or neoadjuvant chemotherapy regimen. Prior anthracycline required, unless contraindicated. *Numbers*: 449 Group A: 225 Group B: 224 *Median age (range):* Group A: 56(22-93) Group B: 54 (28-82)	*Primary outcomes*: objective response rate and toxicity *Secondary outcomes*: duration of response; time to progression, overall survival, quality of life *Length of follow-up: median duration 5.1 years*
O'Shaughnessy et al.[5]	*Design*: RCT *Interventions*: Group A: 1250 mg/m^2 capecitabine + 75 mg/m^2 docetaxel Group B: 100 mg/m^2 docetaxel *Number of centres*: 75 *Minimum duration of treatment*: 6 weeks *Sponsor*: Roche	*Inclusion criteria*: unresectable locally advanced or metastatic disease, prior anthracycline therapy. Patients were excluded if they had received 3 or more chemotherapy regimens for locally advanced or MBC. *Numbers*: 511 Group A: 255 Group B: 256 *Median age (range):* Group A 52 (26-79) Group B 51 (25-75)	*Primary outcome*: time to disease progression *Secondary outcomes*: safety, quality of life *Length of follow-up: at least 15 months*

5.1.3.1 Appropriateness of included studies

The MS attempts to assess comparability of the groups within the trials included for indirect comparisons with the patients in the registration trial. The MS states that the patient characteristics were similar, except for the line of treatment. Both of the additional included trials (Jones and colleagues[4] and O'Shaughnessy and colleagues[5]) had a higher proportion of patients with previous chemotherapeutic treatments for MBC. Table 14 in the MS (p. 59) indicates that approximately one third of people in the docetaxel vs. paclitaxel trial[4] and just under two-thirds of patients in the capecitabine/docetaxel vs. docetaxel trial[5] had received prior chemotherapy for MBC. By contrast, only 0.4% of the patients in the JHQG trial had received prior chemotherapy for MBC. The ERG also notes that the study by Jones and colleagues[4] had a higher proportion of participants with locally advanced breast cancer than the trials of the other comparisons. Over 65% of the patients in the study by O'Shaughnessy and colleagues[5] had three or more metastatic sites, compared with figures closer to 40% for the other two trials' patients.

These factors restrict the degree of similarity between the participant groups, and may influence any resulting analyses. The ERG requested further information on the heterogeneity of patients in the included trials. The manufacturer replied that this was not possible due to the differences in the reporting of key characteristics, and supplied tables of sites and numbers of metastases, and performance score measured on the Karnofsky Performance Status indicator. Patients in the JHQG trial had visceral, lung, liver, non-visceral only and 'other' metastases, whereas those in the trial by O'Shaughnessy and colleagues had lymph node, lung, bone, skin and liver metastases. However, the ERG's clinical advisor indicated that these are just differences in terminology rather than differences in sites of metastasis. No figures are presented in this additional table for the number of patients experiencing metastases at these sites, and data are missing for the trial by Jones and colleagues. The additional information from the manufacturer includes the number of metastatic sites for each study, but again gives no actual figures for the number of patients within each category. KPS score was similar between the three studies included for the clinical effectiveness review.

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

5.1.4.1 Ongoing studies

The manufacturer identified one ongoing phase III trial by Chan and colleagues[10] from which additional evidence is likely to be available in the next 6-12 months. This study of docetaxel/gemcitabine combination versus docetaxel/capecitabine combination has

presented initial results, which were used in the assessment of cost-effectiveness presented in the MS. Final results from this study are awaited. The MS also lists three ongoing trials which investigate gemcitabine/paclitaxel (p.17).

Searches of controlled-trials.com by the ERG identified eight additional ongoing or recently completed RCTs with gemcitabine combination therapies for MBC. Of these, three are recorded as due to complete in the next 12 months and another appears to be the study by Chan and colleagues[10]. The ERG has relied on the data provided on controlled-trials.com and provides brief details (with clinicaltrials.gov reference number) below.

1. An Eli Lilly sponsored phase II trial of gemcitabine/paclitaxel versus gemcitabine/docetaxel (NCT00191672). This study started in December 2003 and is now no longer recruiting; the expected completion date is June 2007. Women with advanced or MBC are included and participants may have had one chemotherapy treatment for advanced or metastatic disease.

2. A phase II/III RCT sponsored by Eli Lilly, comparing docetaxel/gemcitabine with docetaxel/capecitabine with crossover to the alternative agent (NCT00191152). The study started in Feb 2002 and is expecting to recruit 442 participants; the expected completion date is March 2007. Participants are included if they have had up to one prior course of chemotherapy for metastatic disease.

3. A randomized phase II study comparing single-agent docetaxel to alternating docetaxel-gemcitabine as primary chemotherapy for MBC (NCT00191243). This study started in March 2002, and is still recruiting patients (target enrolment=240 patients). It is not clear when this study is due to complete.

4. A study (NCT00191438) of docetaxel/gemcitabine versus docetaxel/capecitabine in patients with MBC and funded by Eli Lilly is assumed to be the study by Chan and colleagues.[10] The study began in October 2002 and is expected to enrol 300 participants before completing in September 2006. Women were included if they had locally advanced and/or metastatic breast cancer and had been previously treated with an anthracycline.

5.1.4.2 Additional studies

The ERG searches did not identify any additional completed RCTs that are relevant to the gemcitabine/paclitaxel versus paclitaxel comparison. The ERG identified one abstract which reports pain and quality of life data from the JHQG trial yet was not mentioned in the review.[2]

Since the inclusion/exclusion criteria for the systematic review do not specify any particular outcome measures, it is not clear why this abstract was not included in table 1 on p. 29 of the MS ('list of publications based on the GT registration study, JHQG'). The ERG requested clarification of this point from the manufacturer, and they replied that this reference was not retrieved in their systematic search, although there was no reason for this considering the inclusion/exclusion criteria. The manufacturer confirmed that the quality of life and pain palliation data presented in the submission and the abstract were both based on the JHQG trial, so were not expected to differ. The ERG has obtained this abstract and data extracted it, and it does not contain any further relevant information to that provided in the MS.

5.1.5 Description and critique of manufacturer's approach to validity assessment

The manufacturer applied the quality assessment criteria recommended by NICE to the JHQG study, but it is not clear whether this was done by a single reviewer or consensus of multiple reviewers. The manufacturer did not apply any quality assessment criteria to the comparator studies which were included for indirect comparison, and did not quality assess the studies included to provide data for the economic model.

Since the JHQG trial has only been published in abstract format, it was not possible for the ERG to check the validity of the manufacturer's quality assessment. On the basis of information presented in the MS, the quality assessment criteria appear to have been applied adequately for questions relating to randomisation and follow-up. The trial was open-label, so observers were not kept fully blinded to treatment assignment. The text in the MS (p. 47) does not score the question on blinding, although the MS text suggests that a mixture of A and B should apply. Whilst the primary outcome (survival) is clearly free from observer bias, outcomes involving tumour response could be affected by bias. Although standard oncology criteria are stated to have been used, there is still a difference between investigator-assessed response and independently assessed response, so there is a degree of subjective interpretation in these outcome measures. As such, the ERG considers that the quality assessment question on blinding should be scored as 'A' (Table 2).

Table 6 in the MS (p.38) details reasons for discontinuation of treatment. The table lists 'death' (academic or commercial information removed in paclitaxel arm, academic or commercial information removed in gemcitabine/paclitaxel arm) and 'death from study disease' (academic or commercial information removed in paclitaxel arm, academic or commercial information removed in gemcitabine/paclitaxel arm). The manufacturer clarified

that the first category is for patients who died due to causes other than MBC. No patients were recorded as having died due to drug toxicity.

Table 2 Quality assessment of JHQG trial (key criteria)

Quality criteria	Description	MS Score	ERG Score
Randomisation	A) No details of randomisation are available, or the method used was inadequate (e.g. randomisation according to the day of the week, even/odd medical record numbers). B) An insecure randomisation method was used, where clinical staff could possibly learn of the treatment assignment (e.g. randomisation sequence kept in the clinical area and open/unblinded trial; treatment assignment kept in consecutive 'sealed' envelopes and open/unblinded trial). C) A secure randomisation method was used, where the randomisation sequence was kept away from the clinical area and administered by staff not directly involved in patient care.	C	C
Follow-up	A) There were significant numbers of drop-outs with no assessment of trial outcome(s) in the subjects who dropped out, and drop-out rates differed between treated and control groups. B) There were some drop-outs with no assessment of trial outcome(s) in the subjects who dropped out, and drop-out rates were (approximately) equivalent in treated and control groups. C) Trial outcome(s) were assessed in all treated and control subjects.	C	C (for primary outcome; QoL outcomes were not completed by all patients)
Blinding of outcome assessment	A) There was an inadequate attempt (or no attempt) to blind observer(s), and the measurement technique was subject to observer bias (e.g. blood pressure measurement with standard sphygmomanometer; measurement of vertebral height on an X-ray). B) The observer(s) were kept fully blinded to treatment assignment, or the measurement technique was not subject to observer bias (e.g. measurement of bone mineral density or survival).	Not classified in MS, but text suggests elements of A and B	A

No formal assessment was made in the MS of the quality of reporting and methodology of the two RCTs of the alternative comparisons (docetaxel versus paclitaxel; capecitabine/docetaxel combination versus docetaxel). Using the NICE guideline for manufacturers, the ERG has assessed these two trials to be of reasonable methodological quality (Table 3).

Table 3 ERG quality assessment of key elements of additional comparator trials

Trial	Quality Criteria	Score and Description
Jones et al[4]	Randomisation	C) A secure randomisation method was used, where the randomisation sequence was kept away from the clinical area and administered by staff not directly involved in patient care
	Follow-up	C) Trial outcomes were assessed in all treated and control subjects
	Blinding of outcome assessment	A) There was an inadequate (or no) attempt to blind observer(s) and the measurement technique was subject to observer bias.

O'Shaughnessy et al[5]	Randomisation	C) A secure randomisation method was used, where the randomisation sequence was kept away from the clinical area and administered by staff not directly involved in patient care
	Follow-up	C) Trial outcomes were assessed in all treated and control subjects
	Blinding of outcome assessment	A) There was an inadequate (or no) attempt to blind observer(s) and the measurement technique was subject to observer bias.

5.1.6 Description and critique of manufacturer's outcome selection

As discussed in Section 4.4, the manufacturer identified appropriate outcomes in the decision problem, but did not specify any outcome measures in its inclusion/exclusion criteria for the systematic review. The outcome measures described in the decision problem reflect those in the JHQG study. The primary outcome was overall survival, and secondary outcomes were: time to disease progression; tumour response; health related quality of life and adverse effects.

5.1.7 Description and critique of the statistical approach used

The JHQG trial was designed to enrol 500 patients, but this was later increased to 526 to counteract missing data from some patients. The protocol required repeat bone scans at approximately eight week intervals, but investigators were found not to be performing repeat scans in patients who had positive scans at baseline. The JHQG trial was initially designed to have TtDPD as the primary outcome, and this was the end point for the interim analysis. The sample size provided a 75% chance of finding a significant difference in TtDPD between the arms at a 2.8% significance level. This assumed a hazard ratio of 0.75 with 20% censoring.

The Food and Drug Administration later requested analysis of survival to ensure that this was not adversely affected by the combination therapy (p.43). Interim survival analysis at 343 deaths was therefore performed, at the 0.0001 significance level. Final survival analysis was conducted at a significance level of 0.04993 (0.5 minus the alpha spent on interim analysis). Final survival analysis significance was therefore changed from 0.03 in the protocol to 0.049983, so the number of projected deaths changed from 440 to 380. Final survival analysis took place at 377 deaths.

The estimates of treatment effect are presented appropriately in the MS, with absolute values, 95% confidence intervals (CI) and *p* values being reported for key outcome measures. The MS contains a transparent description of the statistical approach used for the

JHQG trial, but this is largely marked as being commercial in confidence data. Consequently, the summary here is restricted. The Log-rank test was used for analysis of overall survival, and the Kaplan Meier method was used for survival rates at 12, 18, 24, 30 and 36 months. A Cox model was used to assess treatment effect on survival time. Anyone who discontinued treatment (other than being lost to follow up) was still assessed for TtDPD and overall survival. Data from patients lost to follow up were censored from the last contact date.

The MS suggests that appropriate statistical techniques have been used to handle censored data, however, little data is presented in the MS. There is also limited discussion of whether censoring was independent or not and, if not, what effect it had on the outcomes presented. It is evident that the level of censoring varied depending on the outcome assessed. Although table 6 (p38 of MS) indicates that only 2 patients were lost to follow-up, tables 9 and 10 show censoring ranging from 14.1% to 31.6% depending on the outcome measure. Academic or commercial information removed Without additional information, it is difficult to assess whether censoring was dependent and the effects of an ITT analysis that was last observation carried may have had on the outcomes. Academic or commercial information removed

529 patients were randomised, 266 to the gemcitabine/paclitaxel arm and 263 to the paclitaxel monotherapy arm. Two of the paclitaxel patients were inadvertently given gemcitabine with paclitaxel. For the ITT analysis, these patients were assessed as if they had received paclitaxel monotherapy. For analyses based on the locked database, one of these patients is in the gemcitabine/paclitaxel group. Analysis of overall survival (the primary end point of the trial) was stated to be on an intention to treat (ITT) basis. Quality of life outcomes depended on patients completing a questionnaire and consequently resulted in an incomplete data set, so analysis of this outcome was not ITT. Safety evaluations were only performed for patients who received study treatment, so these were not ITT analyses. Secondary outcomes are reported for the full patient groups, but are not labelled as being ITT. It seems likely that analysis of these outcomes was based on the locked database, which contained one patient in the wrong group, rather than the ITT population.

It is not clear whether potential cross-over of patients affected analysis of overall survival. "JHQG was a parallel-group design. Subsequent therapies were at investigator's discretion, so patients may have crossed over, but this was not a prospective cross-over study". P.47. The CIC flow chart (fig 4, p37) suggests post-therapy follow up took place, and this is also suggested by the conference abstract: "Second-line therapy was nearly identical between

arms, except for a 4-fold greater use of gemcitabine in the T arm"[1]. It is not clear whether this would have affected analysis of overall survival, and the MS does not give any information on the number of patients in the paclitaxel arm who subsequently crossed over and received gemcitabine. Overall survival is defined as being from the date of randomisation to the date of death from any cause. Survival time was censored at the date of the last post-therapy follow-up visit for patients still alive. If paclitaxel patients crossed over to the gemcitabine treatment arm but were included in the ITT data set, this could 'dilute' the survival benefit of gemcitabine. Average survival time for the gemcitabine group would be reduced if the data set includes a large proportion of people who only received the drug after crossing over later in the trial.

Section 2.6 of the MS (p.59) reports that identified trials did not directly compare all of the specific treatments of interest, making it necessary to use indirect comparisons. The MS states that indirect comparisons are subject to bias as the benefit of randomisation does not hold. The ERG would suggest that if the comparisons of the interventions of interest are adjusted by the results of a common comparator group, then the benefit of randomisation is partially held.[11] The MS states that the lack of a common comparator arm across the three studies made it necessary to use only the absolute outcomes to indirectly compare the results of the respective trials. Pooling absolute values in this way provides data which is at best equivalent to an observational study.[11] The ERG considers that in this case, it may have been more appropriate to use other observational studies, as long as they were adequately assessed for bias.

Although there was no common treatment arm across all studies, it might have been possible to use an adjusted indirect comparison for gemcitabine/paclitaxel and docetaxel, since both the JHQG trial and that by Jones and colleagues used paclitaxel monotherapy as a control arm. However, a formal indirect comparison of this kind might still have limited validity in this case, as the characteristics of the trials' patients are heterogeneous. As discussed in Section 5.1.3.1, lines of prior therapy and other characteristics differ between the included trials. These factors restrict the degree of similarity between the participant groups, and may influence any resulting analyses. Although no statistical tests of heterogeneity were available, the manufacturer supplied tables of key patient characteristics.

In Section 2.6 of the MS, there is some confusion over the studies included and how they relate to the section on cost-effectiveness. Tables 13 to 18 in the MS present the data for the JHQG trial and the two comparator trials[4;5] used to assess clinical effectiveness, but there is

no statistical analysis of this data from which to make a formal comparison. The MS does not include any discussion of the data in these tables.

The text on p.56 then discusses trials by Chan and colleagues (1999), Winer and colleagues (2004) [8] and Sledge (2003).[12] Neither the Chan and colleagues 1999 trial nor the Winer and colleagues trial[8] are listed in the bibliography, and copies of these were not provided on the CD of electronic references. The study by Winer and colleagues[8] was explicitly excluded from the systematic review (table 3, p31). The MS reports that patients in these trials were not exposed to anthracyclines in the adjuvant setting, so are not reflective of the UK population or the license for gemcitabine. The MS states that: "they were included to increase the survival estimates of docetaxel and paclitaxel as, in both of these therapies, trials had been based on mixed lines of therapy due to the fact that when the trials were conducted it was still UK clinical practice to give anthracyclines in the metastatic setting."(p.56 of MS). The methods section on indirect comparisons in the MS concludes with equations for a pooled mean and pooled variance. The pooling was used to generate parameters in the cost effectiveness sections but was not used in any analysis in the clinical effectiveness section of the MS.

5.1.8 Summary statement of manufacturer's approach

- A systematic search of the literature was carried out for this review, but the results of this were not clearly reported. There are inconsistencies in the number of studies reported to have been included at different stages of the review. The two 'included' abstracts only reported interim analyses of the JHQG trial, so did not strictly meet the systematic review's criteria. An additional abstract of the JHQG trial was not identified by the manufacturer's searches. The submission is based on JHQG trial data that has not yet been fully published, and as such the MS is not based on evidence identified in the systematic review. It may therefore be subject to bias.
- The MS appears to be complete with regard to relevant studies, including three studies in the review of clinical effectiveness, and the ERG did not identify any additional RCTs. Only one study involved a comparison of gemcitabine and paclitaxel with paclitaxel monotherapy; two further studies were used for comparison with docetaxel and docetaxel/capecitabine.
- The manufacturer only applied the quality assessment criteria to the JHQG study, and did not quality assess the two RCTs used in the indirect comparison. The manufacturer's quality assessment of the JHQG study was generally appropriate, although the ERG noted that the lack of blinding may have introduced observer bias for some outcomes. The JHQG trial is an RCT, and appears to be free from any obvious sources of bias.

- The submitted evidence generally reflects the decision problem defined in the MS.
- The indirect comparison in the MS is simply a tabulation of additional studies for comparison of outcome measures, and there is no attempt to perform a methodologically rigorous statistical comparison of the two trials which have a common comparator (JHQG and Jones and colleagues). [4]
- There was a high degree of censoring in the JHQG trial, and it is not clear how this affects the analyses. For example, the rate of censoring was 35.16% at the interim survival analysis and 28.73% at the final survival analysis (p.43 of MS).

5.2 Summary of submitted evidence

5.2.1 Summary of JHQG results

The manufacturer included one study which reported the effectiveness of gemcitabine with paclitaxel, and the key outcome measures for this trial are included in Table 4.

Table 4 Key outcomes of the JHQG trial

Outcome	**Gemcitabine+ paclitaxel (n=266)**	**Paclitaxel (n=263)**	**P value**
Median overall survival *months* (95% CI)	18.6 (16.6-20.7)	15.8 (14.4-17.4)	P=0.0489
Hazard ratio (95% CI)	0.817 (0.67-1.00)		P=0.0495
Patients censored, n (%)	84 (31.6)	68 (25.9)	
Median TtDPD *months* (95% CI)	5.4 (4.61-6.1)	3.5 (2.9-4.0)	P=0.0013
Hazard ratio (95% CI)	0.73(0.607-0.889)		
Patients censored, n (%)	60 (22.5)	45 (17.2)	P=0.0015
Response rate % (95% CI)	[assessable *n=198*]	[assessable *n=184*]	
Investigator assessed	39(34-45)	26 (20-31)	P=0.0007
Independently assessed*	46 (39-52)	26 (19-32)	P=0.00005
Median PFS *months* (95% CI)	5.3 (4.4-5.9)	3.5 (2.8-4.0)	P=0.0021
Hazard ratio (95% CI)	0.749 (0.621-0.903)		P=0.0024
Patients censored, n (%)	49 (18.4)	37 (14.1)	

*only patients who underwent imaging could have an independent assessment of response

The combination of gemcitabine with paclitaxel had a significantly greater benefit than paclitaxel monotherapy in terms of the primary outcome of overall survival. The confidence intervals for this outcome overlapped, although the *p* value was just within the limit of statistical significance (P=0.0489). This indicates that the 3 month survival benefit seen with gemcitabine is a real statistical difference. Although the survival difference is only marginal in

terms of statistical significance, it would represent a real clinical difference in terms of impact on patients' lives. Overall survival may also have been influenced by subsequent gemcitabine treatment offered to the paclitaxel group after the initial follow-up phase, which was reported by a conference abstract[1]. The combination of gemcitabine with paclitaxel also had a significantly greater benefit than paclitaxel monotherapy on the secondary outcomes of response rate, time to disease progression and progression free survival.

Out of the total population of 529 patients in the JHQG trial, the MS states that 350 completed a Rotterdam Symptom Checklist (RSCL), and 291 completed a Brief Pain Inventory (BPI) questionnaire. Both instruments were completed at baseline, immediately prior to each cycle and 30 days after completing treatment. The ERG requested further details of these assessments from the manufacturer. Further details indicated that 231 of the patients who did not complete a BPI questionnaire were not able to due to the lack of a validated translation. Only seven patients did not complete the questionnaire for other reasons. In sites where a validated translation was available, compliance rates were 84.9% for the gemcitabine/paclitaxel arm and 84.6% for the paclitaxel arm.

The MS states that overall valuation of quality of life on the RSL was statistically significantly higher for patients treated with gemcitabine/paclitaxel compared with paclitaxel monotherapy, but that there were no other statistically significant differences between treatment groups on this scale. Academic or commercial information removed.

The manufacturer's response indicated that ITT analyses had been performed for the BPI, but these were not reported. Analyses were only reported for the small subset of the patients who were symptomatic for pain at baseline (n=81 for gemcitabine/paclitaxel and n=71 for paclitaxel monotherapy). Symptomatic patients in the gemcitabine arm had significantly better BPI scores at cycles 4 and 5 compared with patients in the paclitaxel monotherapy arm (p=0.018 and 0.009, respectively), although this analysis may not be statistically valid due to the small numbers involved.

Only 521 patients were included in the safety analyses of the JHQG trial, and data were missing for the remaining eight patients. The toxicity profile of gemcitabine/paclitaxel therapy is manageable, with similar grade 3 and 4 toxicities being experienced by patients in the two arms of the JHQG trial. Haematological serious adverse events were experienced by statistically significantly more people in the gemcitabine/paclitaxel combination therapy arm than by people treated with paclitaxel monotherapy. Almost half of the gemcitabine/paclitaxel arm experienced neutropenia, compared with just over 10% of the paclitaxel monotherapy

group, and this difference was statistically significant. The rate of leukopenia was 10.6% in the gemcitabine/paclitaxel arm and 1.50% in the paclitaxel monotherapy group.

Of the non-haematological serious adverse events, only fatigue was experienced by statistically significantly more patients in the gemcitabine/paclitaxel arm than in the paclitaxel monotherapy arm, and asthenia was experienced by significantly fewer patients in the gemcitabine/paclitaxel arm than in the paclitaxel monotherapy arm.

5.2.2 Summary of results for comparator trials

Although the MS does not include a methodologically rigorous indirect comparison, results are tabulated for all studies. The MS does not undertake any statistical analysis or narrative summary to compare outcomes between trials. The outcomes are summarised below. Where possible, the ERG has checked the data in the MS with data in the published trials. Unless otherwise stated, the data in the MS corresponds with that in the publications. It should be noted that the trials included in this section had different populations, and may not be directly comparable (see Sections 5.1.3.1 and 5.1.7).

5.2.2.1 Survival

The MS reports median survival (months) for the three included RCTs. In the individual trials, results suggest that median survival is statistically significantly better for:

- docetaxel versus paclitaxel[4] (15.4 versus 12.7 months respectively, p=0.03);
- capetcitabine/docetaxel versus docetaxel[5] (14.5 versus 11.5 months, respectively, p=0.0126); and
- gemcitabine/paclitaxel versus paclitaxel (18.5 versus 15.8 months respectively), as reported in Albain et al. abstract.[1]
- gemcitabine/paclitaxel versus paclitaxel (18.6 versus 15.8 months respectively, p=0.0489), as described in MS.

Differences in patient populations may limit the comparability of these results.

5.2.2.2 Disease progression

In the individual trials, the results suggest that median time to disease progression is statistically significantly better for:

- docetaxel versus paclitaxel[4] (5.6 versus 3.6 months respectively, p<0.0001);
- Capetcitabine/docetaxel versus docetaxel[5] (6.1 versus 4.2 months respectively, p=0.001; and

- gemcitabine/paclitaxel versus paclitaxel (5.4 versus 3.5 months respectively), as reported in Albain et al. abstract.[1]
- gemcitabine/paclitaxel versus paclitaxel (6.0 versus 4.0 months respectively, p=0.0007), as reported in MS.

The MS states that time to documented progression of disease for the JHQG was determined using investigator-assessed data, as this was the endpoint used in the other studies. It is assumed that this is the reason for the slight discrepancy between the data in table 15 (p.60) and the data presented in the conference abstract[3]. The MS data reports median TtDPD to be 6.0 months for gemcitabine/paclitaxel and 4.0 months for paclitaxel monotherapy, but the values given in the executive summary (shown in Table 4 of this report) and the conference abstract[3] are 5.4 months for the gemcitabine/paclitaxel group and 3.5 months for the paclitaxel monotherapy group.

5.2.2.3 Response rate

In the individual trials, the results suggest that response rates were statistically significantly better for:

- capetcitabine/docetaxel versus docetaxel[5] (32% versus 23% respectively, p=0.025;
- gemcitabine/paclitaxel versus paclitaxel (45.5% versus 25.5% respectively, p=0.00005);
- docetaxel versus paclitaxel[4] trial (32% versus 25% respectively, p=0.1).

5.2.2.4 Adverse events

The MS reports comparative safety data on haematological and non-haematological adverse events for the three included comparator RCTs. The manufacturer tabulated results and provided a brief narrative summary, but no statistical analyses were reported. Rates of neutropenia were considerably lower in the JHQG trial than in either of the two comparator trials. For example, the docetaxel monotherapy arm in the trial by Jones and colleagues had a rate of 93% for this adverse event, compared with 48.4% in the JHQG trial. However, there is also a wide variation between the paclitaxel monotherapy arms in the JHQG trial and the trial by Jones and colleagues (10.8% vs. 55%, respectively). Given this variation, it seems likely that differences in baseline characteristics will have skewed the results, and the apparent difference in neutropenia rates between the gemcitabine/paclitaxel group and the other trials' arms may not be as great as would seem at first reading. Rates of febrile neutropenia are 5% in the gemcitabine/paclitaxel arm. Although this is higher than the rate in

the paclitaxel monotherapy arms (2% for both the JHQG trial and that by Jones and colleagues), it is only about a third to a quarter of that reported for docetaxel arms in the comparator trials.

Among the listed non-haematological serious adverse events, rates of stomatis/mucositis were slightly higher in the gemcitabine/paclitaxel arm than in the paclitaxel monotherapy arm (1.5% vs. 0.8%). But these were considerably lower than the rates experienced by people receiving docetaxel monotherapy in the trial by Jones and colleagues[4] (11%) or by those receiving docetaxel monotherapy (5%) or docetaxel/capecitabine therapy (17.4%) in the study by O'Shaughnessy and colleagues. [5]

The ERG noted a small discrepancy between the adverse events reported in tables 16/17 of the MS (p. 61-2) and the published trials. The proportions of patients with neutropenia in the study by O'Shaughnessy and colleagues[5] are reported to be 63.4% (for grade 3) and 72.2% (for grade 4) in the MS. The ERG's review of this publication identified rates of 68% and 77% for grade 3 and 4 neutropenia, respectively. However, the relative difference in the data is similar, and all other proportions noted appeared to correspond with the data presented in the individual trials.

The MS states that a comparison of safety was made between results from Chan and colleagues[10] and results from O'Shaughnessy and colleagues[5] (p.66). The ERG assumes that this is a mistake, and that the comparison was between Jones and colleagues[4] and O'Shaughnessy and colleagues[5] (table 17, p.62), since the study by Chan and colleagues[10] was not included in the systematic review of clinical effectiveness. The MS states that results from the gemcitabine and paclitaxel combination were better than the results from the other two trials [in terms of adverse events] (p.66). This is based on observation of the data only, and the ERG was not able to identify any systematic or statistical approaches to support this.

5.2.3 Critique of submitted evidence syntheses

No meta-analysis was undertaken by the manufacturer due to the differences in the comparators in the included trials. The manufacturer tabulated results from comparator trials, but did not perform a full indirect comparison or narrative synthesis of key outcomes for these (see Section 5.1.7). No formal statistical assessment of heterogeneity was performed, possibly owing to the lack of a standard comparator arm across trials. The ERG requested further details of heterogeneity, and the manufacturer supplied a table of patient characteristics for the different trials.

5.2.4 Summary

- The JHQG trial data in the MS appear to represent an unbiased estimate of the treatment effect of gemcitabine with paclitaxel compared with paclitaxel for MBC. However, the MS is based on unpublished RCT data, and the ERG was not able to check the manufacturer's assessment of methodological quality against anything more substantial than three conference abstracts, so it is not possible to state this with any confidence.

- Results from the JHQG trial suggest that gemcitabine added to paclitaxel improves overall survival, tumour response and time to documented progression of disease, compared with paclitaxel monotherapy. The overall valuation of quality of life on the RSL was statistically significantly higher for patients treated with gemcitabine/paclitaxel compared with paclitaxel monotherapy, but there were no other statistically significant differences between treatment groups on this scale.

- Rates of haematological serious adverse events were significantly higher for the gemcitabine/paclitaxel group than for the paclitaxel monotherapy group. For example, the rate of neutropenia in the gemcitabine/paclitaxel group was more than four times that in the paclitaxel group (48.4% vs. 10.80%). Gemcitabine appears to compare well with docetaxel in terms of adverse events. However, patients in the other trials had received more lines of previous therapy and so rates of adverse events in the two groups may not be directly comparable.

- In the absence of any formal methods of indirect comparison, there does not appear to be sufficient robust evidence to compare the relative effectiveness of gemcitabine/paclitaxel with docetaxel monotherapy or docetaxel/capecitabine combination therapy.

6 ECONOMIC EVALUATION

6.1 Overview of manufacturer's economic evaluation

The manufacturer's submission to NICE includes:

(i) a review of published economic evaluations of chemotherapy regimes for women with metastatic breast cancer. Studies were included if they reported on the cost-effectiveness of gemcitabine-based combination chemotherapy or comparator regimes included in the

economic model. Eight studies are reviewed in descriptive tables outlining the study aims, methods, results and relevance to decision making in England and Wales (MS Section 3.1.2, pages 72 – 75).

None of the studies included in the review (Section 3.1.2 of the MS) evaluated gemcitabine/paclitaxel combination chemotherapy. All included studies were in fully published form – no conference abstracts or reports from proceedings were reviewed. Databases searched are listed in Section 3.1.1 (Q64), page 68-69 of the MS. Other than the ASCO conference proceedings it appears no searches were undertaken to identify conference abstracts. The MS states that there was a concern that specific economic filters would be likely to miss some relevant references – hence very broad (high sensitivity, low specificity) search strategies were adopted, which the MS acknowledges would be expected to identify a large volume of references, many of which would be excluded. The MS does not report the total number of references identified by the searches, the number of excluded studies or the reason for exclusion. This makes it difficult to assess the comprehensiveness of this review. There are some surprising omissions which do not appear to be explained by the inclusion/ exclusion criteria. For example, Hutton and colleagues[13] and Brown and Hutton[14] would appear to meet the criteria for "studies of interest" stated in MS Section 3.1.1 (Q63) page 68 and the inclusion criteria (Q69) page 71, as both are evaluations of paclitaxel and docetaxel monotherapy, but are not reported in the review.

(ii) a report on an economic evaluation undertaken for the NICE STA process. The cost-effectiveness of gemcitabine/paclitaxel combination chemotherapy is estimated compared to three alternative regimes (docetaxel monotherapy, paclitaxel monotherapy and docetaxel/capecitabine combination therapy) discussed in the clinical evidence section (Section 2, pages 26 – 68) and an additional combination chemotherapy (gemcitabine/ docetaxel combination therapy) which has not previously been discussed in the submission.

6.2 CEA Methods

The CEA uses a Markov state transition model to estimate the effect of treatment with each chemotherapy regime included in the evaluation. The model adopted a three year time horizon, assumed to be the life expectancy for this group of patients (see MS question 81, Section 3.2.5, Page 81).

The results from the economic evaluation are presented for the base case assumptions initially using docetaxel monotherapy as the comparator for all interventions, with all

incremental cost-effectiveness ratios calculated relative to this common base. The MS justifies the choice of docetaxel as reference case based on a survey of clinical experts who stated that it is the standard of care in current UK practice. Market research quoted in the MS reported that docetaxel is used in the majority of taxane-based chemotherapy regimes for metastatic breast cancer.

Additional analyses are presented (discussed in section 6.6.3 as scenario analyses) using alternative comparators and allowing for a 55% price reduction for paclitaxel expected when the patent expires.

6.2.1 Natural history

The model of disease progression is similar to that used in the published economic evaluations reviewed in Section 3.1.2 of the MS. Four general health states are defined:

- Stable – no change;
- Response – this is based on reduction in tumour size and is defined as complete for patients with disappearance of all signs of the tumour or partial where tumour size is reduced by more than 50%;
- Progressive – defined as increase in tumour size or spread to other sites;
- Death.

These correspond to standard definitions of disease progression and treatment response that are widely used in oncology practice and commonly adopted in reporting treatment outcomes in clinical trials[15].

Each health state (other than death) is sub-divided to allow for the experience of treatment-related toxicity, which is broken down further by whether the toxicity is life-threatening, requires hospitalisation or is chronic (as discussed in section 6.4.4.2). There are 19 health states in the model.

6.2.2 Treatment effectiveness

Treatment effects used in the model are derived from the overall survival duration, time to disease progression, overall response rate and toxicity data reported in 15 clinical trials. Absolute values for each of these parameters for each intervention were extracted from relevant trial reports and weighted averages were calculated – each value was weighted by the number of cases in the relevant trial arm. The data extracted from the clinical trials and the pooled estimates are reported in Tables 21 to 32 (Pages 85 to 98) of the MS. Cycle probabilities for use in the Markov model were derived using standard transformations[16;17], discussed in section 6.4.4.2.

6.2.3 Health related quality-of-life

None of the quality of life data reported from the JHQG trial (discussed in section 5.2.1) or from other clinical trials was used in the model. The health state valuations used in the model are derived from a survey of 100 members of the general public academic or commercial information removed who completed valuation tasks using visual analogue scales and the standard gamble (SG) technique.

A model developed using the SG valuations suggests that the greatest reduction in utility – against a reference case of a patient with stable disease, experiencing no treatment-related toxicity – is associated with disease progression. This is substantially greater than the utility gain associated with treatment response (-0.27 versus +0.07, see Narewska and colleagues[18]). However, large utility decrements are reported for all toxicities – in all cases a patient who responds to treatment, but also experiences toxicity has a lower utility than the reference case of a patient with stable disease, but no toxicity.

6.2.4 Resources and costs

Resources included and costed in the evaluation were:

- chemotherapy drugs and chemotherapy administration costs;
- supportive care, including management of adverse events;
- palliative care costs.

Other than the costing of chemotherapy drug use (which was based on licensed dosages), the majority of resource use estimates were based on protocols developed using expert clinical opinion (discussed in section 6.4.4.4). Unit costs were derived from a variety of sources with different base years (discussed in section 6.4.4.5).

The MS reports that the cost year for the model was 2005/06. Where costs for other years were used as inputs these were uprated using the Hospital Pay and Prices Index[19], discussed in section 6.4.4.5.

6.2.5 Discounting

An annual discount rate of 3.5% was applied to both costs and outcomes. Section 3.3.1, Q111, p.128 of the MS states that rates between 0% and 6% were applied in sensitivity analyses, but the results of these analyses do not seem to be reported.

6.2.6 Sensitivity analyses

Probabilistic sensitivity analysis is reported alongside the base case results in Section 3.4.1. Means or measures of variation of costs and outcomes are not reported in tables – cost-effectiveness plots for pairwise comparisons (gemcitabine/paclitaxel versus docetaxel) and cost-effectiveness acceptability curves (CEACs) for each intervention against docetaxel are shown. One-way sensitivity analyses for selected variables are reported in Section 3.4.2.

6.2.7 Model validation

Approaches to validating the model are described in MS Section 3.3.4, Q115, p.134. The principal validation technique appears to have been establishing the face validity of structural assumptions in the model and the selection of parameters (and ranges) for the sensitivity analysis through expert clinical opinion, supplemented by a technical review of the model.

The approach to establishing external consistency was to compare the model results with the published evaluations reviewed in Section 3.1.2 of the MS. In the absence of published studies of the cost-effectiveness of the gemcitabine/paclitaxel combination the validation focussed on the results for comparator treatments.

6.3 Results

Results from the economic model are presented as incremental cost per life year gained and incremental cost per QALY gained. Life expectancy, quality adjusted life expectancy and lifetime costs are also presented. The base case analysis, with docetaxel as the comparator reports an estimated incremental cost per QALY of £17,168 for gemcitabine in combination with paclitaxel. One way sensitivity analyses are reported for a limited number of variables related to the efficacy of treatment and for health state utility values. The majority of variables examined in the one-way sensitivity analyses are related to resource use and cost. In all the reported analyses the incremental cost effectiveness ratios for the gemcitabine/paclitaxel combination compared to docetaxel are within the range £13,000 to £21,000 per QALY gained. Table 5 summarises the results of the base case and main scenario analyses reported – for brevity only the results for the gemcitabine/paclitaxel combination compared to the reference case are shown in the table.

Table 5 Cost-effectiveness results presented in MS

Analysis	Difference in mean discounted outcomes	Difference in mean discounted total costs	Incremental cost effectiveness ratios
Base case – docetaxel as comparator			
Base case analysis	Life years 0.43 QALYs 0.23	£ 4,013	Life years £ 9,253 QALYs £ 17,168
Threshold analysis - overall survival with docetaxel increased			

Overall survival with docetaxel increased from 59.4 to 63 weeks	Life years 0.37 QALYs 0.20	£ 4,089	Life years £ 11,185 QALYs £ 20,073
Overall survival with docetaxel increased from 59.4 to 70 weeks	Life years 0.23 QALYs 0.14	£ 4,261	Life years £ 18,658 QALYs £ 29,742
Scenario analysis – post-patent expiration price reduction for paclitaxel			
Paclitaxel cost reduced by 55%	Life years 0.43 QALYs 0.23	£ 1,109	Life years £ 2,556 QALYs £ 4,742
Scenario analysis – alternative reference case			
Paclitaxel as comparator	Life years 0.25 QALYs 0.15	£ 4,498	Life years £ 17,924 QALYs £ 30,096
Docetaxel/capecitabine as comparator	Life years 0.31 QALYs 0.20	£ 4,521	Life years £ 14,484 QALYs £ 23,152

The MS summarises the results of the PSA stating that there is 70% probability of gemcitabine/paclitaxel being cost-effective, relative to docetaxel monotherapy, at a threshold willingness to pay of £35,000 per QALY.

6.4 Critical appraisal of the manufacturer's submitted economic evaluation

6.4.1 Critical appraisal of economic evaluation methods

The ERG has considered the methods applied in the economic evaluation in the context of the critical appraisal questions listed in

Table 6 below, drawn from common checklists for economic evaluation methods (e.g. Drummond and colleagues 1997).

Table 6 Critical appraisal checklist of economic evaluation

Item	Critical Appraisal	Reviewer Comment
Is there a well defined question?	Yes	See Q22, page 19
Is there a clear description of alternatives?	Yes	Docetaxel monotherapy is used as comparator for all interventions. Interventions are: • Gemcitabine/Pacxlitaxel • Paclitaxel monotherapy • Docetaxel/Capacetabine • Docetaxel/Gemcitabine. Drug dosages listed in Section 3.2.4, Q79, page 79-80. See also tables of assumptions Page 107 (dosages) Page 108 (duration of treatment)
Has the correct patient group / population of interest been clearly stated?	Yes	Women with MBC who have relapsed following adjuvant/ neoadjuvant chemotherapy, which should have included an anthracycline unless clinically contraindicated. Patients receiving comparator or interventions as first-line treatment for MBC. No sub-groups identified. Baseline age for patient cohort not clear in MS
Is the correct comparator used?	?	Use of docetaxel as comparator deemed clinically relevant –

		justified by docetaxel having majority of taxane treatment for MBC. Issues for discussion: • evaluation is for first line treatment of MBC, but evidence is not all for first-line treatment; • comparisons are made for all regimes relative to the reference case (docetaxel in the base case). Would a frontier analysis be more appropriate? • relevance of DG comparator in economic evaluation? Section 1.4, Q26, Page 21 does not agree with comparators in economic evaluation
Is the study type reasonable?	Yes	Cost-utility study appropriate – the evaluation needs to capture quality of life difference for response rather than stable health state, and quality of life impact (as well as cost impact) of adverse events
Is the perspective of the analysis clearly stated?	Yes	NHS and Personal Social Services. See Q80 – perspective required for NICE reference case
Is the perspective employed appropriate?	Yes	Costs from NHS and PSS perspective. Outcomes from patient perspective – life expectancy and quality-adjusted life expectancy using utility weights based on values from survey of general public.
Is effectiveness of the intervention established?	?	Data from clinical trials – fully published [except JHQG]. Only the JHQG trial was quality assessed. Overall survival and TTP advantage for gemcitabine/paclitaxel combination over paclitaxel monotherapy established by direct comparison in JHQG trial. All other comparisons indirect – with questionable validity of pooling method. No formal assessment of heterogeneity. Data from trials with anthracycline naïve patients included in base case. Inconsistent use of independent versus investigator assessed response.
Has a lifetime horizon been used for analysis (has a shorter horizon been justified)?	Yes	Assumed three year survival used as time horizon for model. Based on reports by Perez[20] and by Blum and colleagues[21], referenced in Section 1.3, Page 17-18. Also see Q81 on Page 81.
Are the costs and consequences consistent with the perspective employed? *	Yes	Costs consistent with NHS and PSS perspective. Consequences presented as life expectancy and quality-adjusted life expectancy using utility weights based on values from survey of general public.
Is differential timing considered?	Yes	Discount rates applied 3.5% for costs and outcomes. Applied as annual rates, rather than per cycle.
Is incremental analysis performed?	Yes	Reported in tables in Section 3.4.1, Q116, Pages 134-135,
Is sensitivity analysis undertaken and presented clearly?	?	Sensitivity analysis is reported in MS – probabilistic sensitivity analysis reported alongside the base case results in Section 3.4.1, Q116-117, Pages 134-140. One-way sensitivity analyses reported in Section 3.4.2, Q119-120, Pages 140-145. Maybe regarded as limited. Not all variables included in PSA (MS does not discuss why those particular variables were chosen). Clarification from manufacturer received and discussed in section 6.6.5 Variables included in one-way sensitivity analysis listed but rationale for choosing those variables not discussed – ranges not justified or related to CIs from data pooling. Clarification from manufacturer received and discussed in section 6.6.1

6.4.2 NICE reference case

Table 7 NICE reference case requirements

NICE reference case requirements (see detail in NICE report)**:**	Included in Submission
Decision problem: As per the scope developed by NICE	NA
Comparator: Alternative therapies routinely used in the UK NHS	Yes - ?
Perspective on costs: NHS and PSS	Yes
Perspective on outcomes: All health effects on individuals	Yes
Type of economic evaluation: Cost effectiveness analysis	Yes (CUA)
Synthesis of evidence on outcomes: Based on a systematic review	?*
Measure of health benefits: QALYs	Yes
Description of health states for QALY calculations: Use of a standardised and validated generic instrument	?†
Method of preference elicitation for health state values: Choice based method (e.g. TTO, SG, not rating scale)	Yes‡
Source of preference data: Representative sample of the public	?‡
Discount rate: 3.5% pa for costs and health effects	Yes

Notes:
N/A=not applicable
* A systematic search was undertaken and reported in MS, but there was inadequate quality assessment of included trials, inadequate assessment of comparability of trials and questionable method of data pooling.
† unclear from Narewska and colleagues[18] Academic or commercial information removed

‡ see sections 6.2.3 and 6.4.4.3

6.4.3 Modelling methods

An outline critical review of modelling methods has been undertaken. The review has used the framework for good practice in modelling presented by Philips and colleagues (2004) as a guide, addressing issues of model structure, structural assumptions, data inputs, consistency, and assessment of uncertainty.

6.4.3.1 Modelling approach / Model Structure

The basic structure of the model is presented in section 3.2.6.1 of the MS. While this section makes no reference to previously published economic models in this setting, the executive summary makes clear that the model structure is based on that used by Cooper and colleagues[23], itself a modification of a model originally developed by Hutton and colleagues[13]. The latter publication was not included in the review in section 3.1.2 (pages 72-75) of the MS. However, the study by Brown and colleagues[24], which also adapted this model, was included.

Due to the advanced state of disease, it is assumed that all patients (who enter the model in the stable health state) will enter the progressive health state during the three year time horizon of the model. They will not have any spontaneous remission of disease, nor would treatment response be likely to be permanent or result in disappearance of all disease. The

effect of treatment is to postpone progression, either with the patient remaining in the stable state or by providing a temporary response with an associated increase in quality of life and a longer duration of progression-free survival than for a patient whose tumour does not show a response. This model accords with clinical expectations for patients with MBC who have relapsed following treatment in the adjuvant/ neoadjuvant setting. However the ERG clinical advisor is concerned that patients' response to second and subsequent lines of treatment has not been properly represented in the model.

None of the treatment options modelled represents a true natural history of disease progression, with supportive care – all options are intended to extend progression-free survival and to improve quality of life. It is assumed in the model that all patients are eligible for, and take up, treatment.

The use of a Markov cohort model seems appropriate, given:

- the use of time-dependent probabilities in the model (for example risk of febrile neutropenia and probability of response vary by treatment cycle);
- the need to track disease progression and treatment response;
- the need to adjust life expectancy estimates for quality of life associated with disease progression and remission, as well as the development of treatment-related toxicity;
- the available data;
- that the chemotherapies are administered in defined cycles.

The model adopted a cycle length equivalent to the chemotherapy cycle length (three weeks) which is common across all the regimes being investigated. Given the comparatively short life expectancy being modelled (three years[20;21]) a short model cycle is appropriate. No half-cycle correction has been applied in the model, though this is unlikely to have an impact due to the short cycle length adopted.

Alternative modelling approaches are discussed in the MS, including the option of patient-level modelling (rejected due to the lack of patient-level data for model inputs for all interventions). The principal issue, related to model structure, discussed in the MS is the inclusion of toxicities and their impact on quality of life (utility). Alternative structures, which take no account of the quality of life impact of toxicity would, as the MS suggests, be inappropriate given the differences in toxicity profiles for the drugs being compared. At the other extreme, including all grades of toxicity would be expected to over-complicate the model for little gain in precision, and would be hampered by the convention of reporting only grade 3 and 4 toxicities in clinical trial reports. The decision to include toxicities experienced

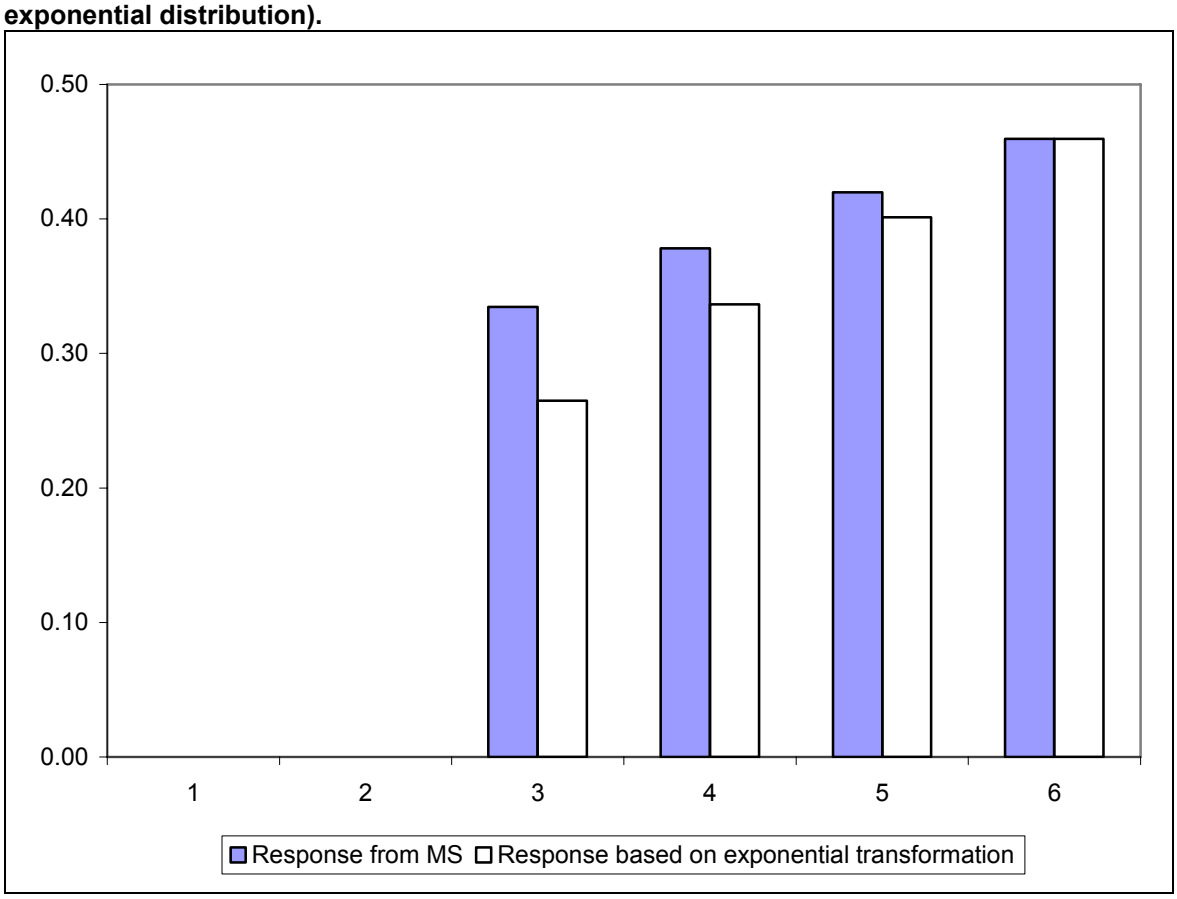
by greater than 6% of patients seems to be an appropriate pragmatic decision that does not exclude toxicities that would be of major clinical significance.

The sources of data used to develop and populate the model structure are clearly specified in tables in the MS (pages 85-116) – with the exception of some of the tables of assumptions on resource use, discussed in section 6.4.4.4.

6.4.3.2 Structural Assumptions

It is assumed that patients are not assessed for response until cycle 2 (for docetaxel and docetaxel/capecitabine) and cycle 3 (for gemcitabine/paclitaxel and paclitaxel). Data from two clinical trials were used to estimate the proportion of patients responding at these time points – for docetaxel and docetaxel/capecitabine this is derived from S273[10] (docetaxel monotherapy was not included in the trial, so the value for combination therapies containing docetaxel were used) and for gemcitabine/paclitaxel and paclitaxel this is derived from JHQG. Using these values it is assumed that 66-73% of responders will be in the response state in the first response cycle. The remaining 27-34% of response is equally distributed across the remaining treatment cycles using the standard transformations as per Miller and Homan[16]. This approach makes use of available data and ensures that response in the model fits the available data – use of a standard exponential transformation would have underestimated response in early cycles – see Figure 1.

Figure 1 Response by treatment cycle for gemcitabine/paclitaxel (independent assessment) showing values used in the model and values for standard transformation (assuming exponential distribution).



The ERG clinical advisor has questioned the clinical plausibility of this approach to scheduling response, suggesting that almost all responses will be apparent after the first three cycles of treatment. However they may not have achieved the threshold described in section 6.2.1 for defining partial or complete response – the impact of varying assumptions over scheduling of response is discussed in section 6.6.2.

The model follows a conventional approach to allocating patients to health states at the first response cycle. The probability of response and probability of early progression (patients whose disease has progressed at first assessment) at the first response cycle is determined – the probability of stable disease is the residual probability. Early progression has not been reported for the JHQG trial. Progression at the first response cycle (cycle 3 for paclitaxel and gemcitabine/paclitaxel) has been calculated in the model by applying the risk of disease progression for non-responding patients from the first treatment cycle. For gemcitabine/paclitaxel the risk of progression per cycle is 10.22%, which results in early progression in 21.1% of the cohort – the proportion with early progression for docetaxel and paclitaxel monotherapy are 27.5% and 23.6% respectively. These latter values compare to

13%-23.8% and 20%-22% used for docetaxel and paclitaxel monotherapy in previous evaluations.

Clinical trial reports do not generally distinguish between patients in the stable and response states when reporting the occurrence of toxicities. As a result the same cycle probabilities for developing toxicity are applied in the stable and response health states. The transformations applied to derive cycle probabilities from the toxicity data reported in the clinical trials mean that the model assumes a constant risk for each toxicity, in each cycle. These assumptions are appropriate given the available data. The exception to the assumption of constant risk is febrile neutropenia, where the risk is "front-loaded" into the first three cycles. The degree of front-loading is dependent on data observed in JHQG trial. This is set at 100% for paclitaxel (observed in the JHQG trial) and docetaxel (assumed to be the same as paclitaxel) and 58% for gemcitabine/ paclitaxel (observed in the JHQG trial). Previous evaluations[13;14;23;25] have assumed that febrile neutropenia and infections (as life-threatening events which may require hospitalisation) only occur in the first two treatment cycles. However these evaluations did not include gemcitabine/paclitaxel. This assumption over the scheduling of febrile neutropenia is unlikely to have a substantial impact on the results of the evaluation – see section 6.6.2.

The model assumes that patients' risk of developing toxicity is independent of their previous experience of toxicity and that patients can develop only one toxicity in any cycle – though those who have already developed a chronic toxicity (alopecia or peripheral neuropathy) may also experience acute toxicity in subsequent treatment cycles. These are pragmatic assumptions, required since data on patients experience of repeated or multiple toxicity are not typically reported – the MS reports (Page 112, table titled Exclusivity of experiencing adverse events) that clinical trial data indicate that very few patients experienced multiple toxicities during a single treatment cycle.

The model is structured so that patients need to enter the progressive state prior to death – no mortality risks, either all-cause or disease-specific, are applied in the stable or response health states. The assumption that breast cancer mortality results from disease progression, and is minimal in patients whose disease is stable or who respond to chemotherapy, agrees with clinical experience. Failure to include all-cause mortality is unlikely to have an impact in the model since breast cancer mortality will be the major force of mortality in this patient group. All-cause mortality for women aged 53 (median age in JHQG trial[3]) from UK life tables[26] is 0.34% per annum whereas the annual mortality for patients with progressive disease in the model is 45%-63%. The mortality risk for patients is based on an estimate of

the median life expectancy in the progressive state (derived by subtracting the median time to disease progression from the median overall survival time for each intervention).

The model assumed a constant risk for disease progression and for mortality (p. 146). The ERG estimated the survival probabilities and risk of disease progression for patients in the paclitaxel arm of the trial from survival plots reported in the conference presentation by Albain and colleagues,[27] and fitted a parametric survival function to these data using the outputs from an ordinary least squares regression on a log-cumulative hazard[28]. These suggest that the survival functions have non-constant risk over time – the implications of this on the results of the analysis are discussed in section 6.5.2.

6.4.4 Data Inputs

6.4.4.1 Patient Group

Input data used in the economic model were extracted from a number of trials that were not included in the systematic review of clinical effectiveness (see Section 5.1.2). These have therefore not been quality assessed, nor have the trial populations been formally assessed for comparability. The patient populations vary in terms of baseline characteristics such as prior therapy, metastatic setting, and line of chemotherapy. For paclitaxel monotherapy, for example, the percentage of patients who had anthracycline as prior therapy ranges from 68% to 98.2% (p. 201-202, this range excludes trials in anthracycline-naïve patients which are included in the base case), the percentage of patients who had neo/adjuvant therapy ranges from 31% to 99% (p. 59, 201-202) and the percentage of patients who had more than 3 metastatic sites ranges from 25% to 72% (p.205-206). While this assessment aims to review the use of the gemcitabine/paclitaxel combination as first line chemotherapy in metastatic breast cancer patients with prior anthracycline therapy, the pooled estimates may not reflect this indication. In particular, the inclusion of data from trials where all patients were anthracycline-naïve (one for docetaxel and two for paclitaxel monotherapy) is of questionable validity. While the analysis is run with and without these trials the report would have been clearer if a decision were made to include or exclude them from the input data and to have justified that decision.

Trials reported as abstracts have typically provided incomplete datasets – for example both Extra and colleagues[9] and Mouridsen and colleagues[29] lack data on median time to disease progression. Extra and colleagues[9] also lacks data on the majority of adverse events included in the economic model. Inclusion of other trials is open to question due to design of the trial (Extra and colleagues[9] includes HER-2 patients who crossed over to trastuzumab, with overall survival greater than the upper 95% confidence interval for the pooled estimate

for docetaxel) and methodological inadequacies (as noted in the MS, Icli and colleagues[30] used treatment received not randomised, thereby undermining the intention to treat principle).

6.4.4.2 Clinical Effectiveness

The pooled median overall survival durations for each intervention were converted to cycle probabilities for use in the Markov model using standard transformations (deriving rates from the medians, as described by Beck and colleagues[17], and transition probabilities from the rates as described by Miller and Homan[16]). Time to progression was estimated separately for the responders and non-responders using the pooled median time to progression, pooled overall response rates and additional trial data[5].

Toxicities included in the economic model were grouped under four headings:

- Life-threatening (febrile neutropenia, for which patients may be hospitalised. The proportion hospitalised was assumed not to vary by chemotherapy regime and was estimated as the number of febrile neutropenia cases hospitalised divided by the total number of cases in two clinical trials[1;5]);
- Hospitalised (diarrhoea & vomiting and stomatitis & mucositis);
- Non-hospitalised (fatigue, hand & foot syndrome and neutropenia);
- Chronic (alopecia and peripheral neuropathy).

The proportions of patients experiencing toxicity were converted to cycle probabilities for use in the Markov model using standard transformations (as described by Miller and Homan[16]) - assuming that patients receive a maximum of six cycles of chemotherapy for each intervention. The probability of developing toxicity for each of the above groups was calculated as the sum of the cycle probabilities for each toxicity included in the group.

The data extraction has not been consistent, with both arms from some trials being included but only single arms from others (see Appendix 2 – Additional tables for ERG report, Table 1). For example the economic evaluation has been expanded to include gemcitabine/docetaxel (as it is one arm of the trail reported by Chan and colleagues[10]), but gemcitabine is not currently licensed in this combination and there is no indication of its current use in first-line treatment in the UK.

The MS has been inconsistent in the approach to missing values when pooling data – for example, where independent assessment of response was not reported the investigator-assessed value was used, whereas for adverse events and median time to disease

progression data missing from trial reports are indicated as “NR” (not reported) or “missing” and are not included in pooled estimates.

In addition, the absolute estimates of the relevant arms in the included trials, *not* the relative estimates, were pooled. That means the benefits of randomisation of the trials is lost. Thus bias introduced in the pooled estimates would affect the analyses when comparisons are made against gemcitabine plus paclitaxel combination therapy, where estimates are based on a single trial (JHQG). For example, the median overall survival duration for the paclitaxel arm in the JHQG trial (reported as 68.5 weeks) is higher than the pooled estimate of 56.31 weeks (MS p.85) derived from trials that included anthracycline pre-treated patients. For the base case analysis this means that the model inputs provide a greater survival advantage for the gemcitabine/paclitaxel combination, over paclitaxel, than was observed in the head-to-head comparison. The effect of this is discussed in section 6.6.2. The pooled estimates for median time to disease progression (MS p.86) and overall response rate (p.87-88) were also lower than for the paclitaxel arm in the JHQG trial though the difference was less marked than for overall survival.

The pooled estimates of the proportion of patients experiencing adverse events – for example, febrile neutropenia (p. 89), fatigue (p.90-91), diarrhoea/vomiting (p.94-95), neutropenia (p.96-97) and sensory/ peripheral neuropathy (p.97-98) - for paclitaxel monotherapy are higher than those reported in the JHQG trial. This would make the gemcitabine/paclitaxel combination therapy seem to offer a better toxicity profile than would be the case if estimates from the JHQG trial were used in the model.

6.4.4.3 Patient outcomes

The development and validation of the health state descriptions and an outline of the valuation study that produced the health state descriptions used in the model have been published as a conference poster[18]. A paper describing the valuation survey, giving more detail and including the regression model, has been submitted for publication and was made available to the ERG[22].

The full set of health state descriptions are not presented in the MS or in the accompanying material. Based on details presented by Narewska and colleagues,[18] it appears that the health state descriptions cover broadly the same dimensions as previous studies aiming to derive utilities for patients with MBC.[13;14;23;25] The previous studies based their dimensions on items in the Health Utility Index[31] (ambulation, dexterity, emotion, cognition and pain/activities) and HUI-2[32] (personal care), supplemented by six items more specific to

chronic disease and health effects of cancer chemotherapy (fear/anxiety, depression, energy, hair loss, pain relief and nausea). Two additional items investigated by Narewska and colleagues[18] Academic or commercial information removed were appetite-loss (as a symptom of disease separate from toxicity inducing diarrhoea and vomiting) and sexual function. Cognitive function was removed at the validation stage as being unlikely to have a substantial impact[18].

The utility values applied in the model are broadly consistent with those adopted for previously published economic evaluations, which used the SG technique, but derived their valuations from samples of oncology nurses[13;14;23;25] – see Table 8.

Table 8 Health state utilities reported in MS and other published economic evaluations

Utility weights	Submission	Cooper and colleagues[23]	Launois and colleagues[25]	Hutton and colleagues[13]	Brown and Hutton[14]	Brown and colleagues[24]
Response	0.80	0.81	0.81	0.81	0.81	0.84
Stable	0.72	0.65	0.75	0.62	0.65	0.62
Progression	0.46	0.45	0.65§	0.41†	0.39†	0.33*
Response with toxicity						
Febrile Neutropenia	0.67	0.60 - 0.44‡	0.66 - 0.47‡		0.3‡	0.24‡
Diarrhoea/Vomiting	0.71	0.67				
Stomatitis	0.67					
Fatigue	0.70					
Hand/foot syndrome	0.70					
Neutropenia	0.80					
Hair loss	0.70					
Neuropathy	0.70		0.57	0.53	0.56	0.62
Oedema			0.74	0.75	0.76	0.78
Stable with toxicity						
Febrile Neutropenia	0.58	0.54	0.66 - 0.47‡		0.3‡	0.24‡
Diarrhoea/Vomiting	0.62					
Stomatitis	0.58					
Fatigue	0.61					
Hand/foot syndrome	0.61					
Neutropenia	0.72					
Hair loss	0.61					
Neuropathy	0.61					
Oedema			0.73			
Progression with toxicity						
Neuropathy	0.33	0.45	0.50			
Oedema			0.53			

Notes:
§ utility value of 0.16 for terminal disease (cycle immediately prior to death in progressive state)
† utility value of 0.16 for terminal disease (cycle immediately prior to death in progressive state)
* utility value of 0.13 for terminal disease (cycle immediately prior to death in progressive state)
‡ febrile neutropenia requiring hospitalisation

The values reported in the MS are shown in Table 8. Academic or commercial information removed. The average ages for patients in the clinical trials reviewed in section 2.7 of the MS are higher than this – median age ranges from 51 (for docetaxel monotherapy in O'Shaughnessy and colleagues[5]) to 56 (for docetaxel monotherapy in Jones and colleagues[4]). Academic or commercial information removedThe effect of using health state utilities calculated for an older cohort is examined in the ERG sensitivity analyses, section 6.6.2.

The MS asserts in a number of places that previous economic models have not addressed the quality of life impact of adverse events and treatment-related toxicity, justifying this with a quote from a technology assessment report[33] (see Q84 – p.117, Q87 – p.118, Q88 – p.119, Q124 – p.147). The quoted passage refers to a particular industry model submitted to NICE for that particular appraisal (see Section 5.2.1.3, p65 of the assessment report[33]) and does not apply to all models published prior to the appraisal. The five published evaluations listed in Table 8 (three of which were used in the MS when discussing the external validity of the values adopted in the model) all include health effects as well as cost impacts of treatment-related toxicity.

While each toxicity was identified separately in the health state valuation exercise and in the regression modelling, utilities have been applied in the cost-effectiveness model by adverse event category – defined as chronic AE, non-hospitalised AE and hospitalised AE. The average of all values within the category was taken as the utility value for the category. The health state utilities, as reported in Table 41 in MS, are applied to the relevant modelled life expectancy in each health state for each intervention, for each cycle and discounted as described in section 6.2.5.

6.4.4.4 Resource use

The MS does not report whether a systematic search for data on resource use for patients with MBC receiving chemotherapy was undertaken, nor are the resource assumptions in the UK studies included in section 3.1.2 of the MS discussed. The MS reports three sources for resource use estimates – clinical trial data, expert opinion and treatment protocols. It is not always clear from the tables of assumptions which sources have been used, though the majority of assumptions appear to be based on opinion (adverse event data from study S273, published as Chan and colleagues[10] and the JHQG trial were used to determine which adverse events required hospitalisation, the duration of hospitalisation for adverse events and duration of acute adverse events – these appear to be the only trial-based resource use assumptions referenced in the tables).

Treatment costs have been calculated using the licensed dosages for each chemotherapy regime (without adjustment for dose reductions or omitted doses observed in clinical trials) assuming a patient body surface area of $1.8m^2$ and a maximum of six chemotherapy cycles administered for first-line treatment of MBC. Data on dose reductions (gemcitabine 8% and paclitaxel 5% for the gemcitabine/paclitaxel arm versus 2% for the paclitaxel only arm) and omitted doses (gemcitabine 7% and paclitaxel less than 1% for the gemcitabine/paclitaxel arm and less than 1% for the paclitaxel only arm) have been reported in a conference presentation.[27] While the median dose delivered for paclitaxel in both arms of the JHQG trial was the same as the licensed dosage, the median dosage for gemcitabine was around $100mg/m^2$ lower than the dosage costed in the economic model. This indicates that costing at licensed dosages will over-estimate chemotherapy costs for some, but not all interventions. Similar data are available for other interventions included in the economic model. However the MS has only considered dose reduction for docetaxel.

The ERG considers the estimates of cost for chemotherapy regimes included in the economic model to be reasonable. The difference in the cost per cycle for gemcitabine/paclitaxel between using the licensed dosage and the median observed in the JHQG trial is unlikely to have a significant impact on the results of the evaluation, amounting to an approximately £400 difference over 6 treatment cycles.

The estimated resource use for each chemotherapy regime includes pre-medication, which have been quantified for the MS using clinical opinion – no reference is made in the MS to the statement of product characteristics for each drug. These broadly agree with the pre-medications and dosages assumed in the MS. A patient receiving 6 cycles of chemotherapy with gemcitabine/paclitaxel will have a treatment cost of £11,848, plus additional administration costs.

Chemotherapy administration costs are based on assumptions over the duration of drug administration, time taken for blood tests and consultation with the doctor, as well as time for the pharmacist to prepare the drugs and administer pre-medication. The source for these assumptions is not clearly stated; however the assumed durations appear reasonable. The robustness of the cost effectiveness estimates to these assumptions is tested in the one way sensitivity analyses, where the time taken to administer chemotherapy is increased by two-and-a-half hours (p.143 and p.146 of MS).

While the body surface area (BSA) of 1.8m^2 assumed in the model is taken from a study of patients with advanced non-small cell lung cancer (not included in reference list for MS), this value is similar to assumptions for BSA in previous evaluations[13;14;24] (range of assumed BSA for MBC patients is 1.66m^2 to 1.75m^2).

All resource use assumptions for the management of adverse events were based on clinical opinion - except the range applied in the one way sensitivity analysis for the number of inpatient days for adverse events requiring hospitalisation, which was based on data from S273 trial.

6.4.4.5 Costs

Unit costs for all chemotherapy drugs and for pre-medication are taken from the British National Formulary (BNF 50)[34] and are still current at June 2006. Other unit cost data are taken from a variety of sources, including 2004 NHS Reference Costs (for laboratory and radiological tests), NHS TFR Returns (for inpatient stays) and the National Blood Bank (for blood products). The sources of unit costs are appropriate and an attempt has been made to adjust costs from varying financial years to a consistent base of 2005/06 (see Section 6.1.4). However the method used for this may be open to question since the current published NHS Pay and Prices Index only runs to the financial year 2003/04 – with an estimated value for 2004/05. To uprate costs to 2005/06 prices, the 2004/05 (estimated) value has been applied for both 2004/05 and 2005/06 financial years.

6.5 Consistency

6.5.1 Internal consistency

The ERG has examined the submitted Excel model for internal and external consistency and accuracy. Random checking has been done for some of the key equations of the model although this has not been a comprehensive 'checking' process of all cells in the model. The model is fully executable and is run by clicking on the button 'Run CEA Analysis' in the *Model Set Up* Excel sheet. Inputs changed in the Input parameter Excel sheets produce appropriate changes when the model is run.

The parameter inputs of the model were checked for the relevant formula or against the estimates in the written report and a number of discrepancies were identified (see Appendix 2 – Additional tables for ERG report, Table 2). However these did not have a substantial impact on the results of the model.

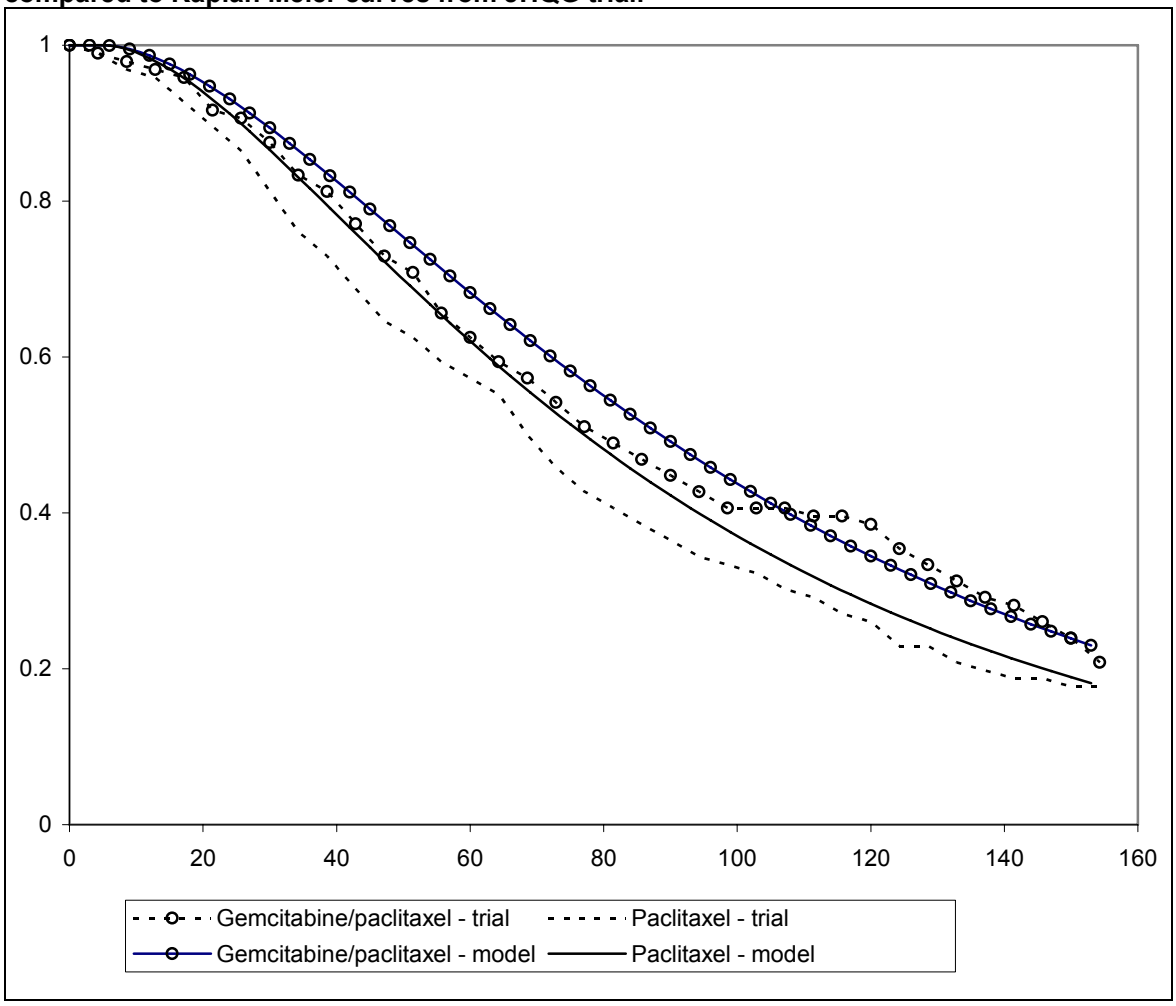
6.5.2 External consistency

Approaches to validating the model are described in MS Section 3.3.4, Q115, p.134. The principal validation technique appears to have been expert clinical opinion - by senior oncologists in UK clinical practice, UK National Oncology Breast Advisor and by the manufacturers own clinical research physicians. The role of validation here appears to be most concerned with model structure, identification of the decision problem and in specifying variables and ranges for the sensitivity analyses. The MS states that a technical review was conducted by two researchers working independently, but gives no detail of methods adopted, criteria/ checklists adopted, nor of the results of this validation.

The approach to establishing external consistency was to compare the model results with the published evaluations reviewed in Section 3.1.2. For the outcome of this external validity check the MS refers to the response to Q122 in Section 3.4.3. This offers limited evidence, as the MS acknowledges that no studies of gemcitabine/paclitaxel were found in their review, and that many of the publications could offer little insight on the validity of the model or results as they included comparators that were not relevant to UK practice or were conducted in currencies other than UK sterling.

The ERG estimated the external validity of the model by comparing it with survival estimates from the JHQG trial. The model was run with median survival times for gemcitabine/paclitaxel and paclitaxel as shown in the JHQG. Figure 2 shows the predicted survival from the model compared to the Kaplan Meier curves reported in a conference presentation[27].

Figure 2 Estimated survival for gemcitabine/paclitaxel and paclitaxel predicted by the model compared to Kaplan Meier curves from JHQG trial.



In the Figure, the solid lines are the survival curves predicted by the model and the dotted lines are those from the trial. In each case the benefit of gemcitabine/paclitaxel treatment over paclitaxel monotherapy is shown by the area between the survival curves. The model shows a reasonable fit with the Kaplan Meier curves, with a slightly increased survival for the model compared to the trial data for gemcitabine/paclitaxel and paclitaxel. The model and trial estimated an average survival benefit of 7.6 and 11.2 weeks respectively for gemcitabine/paclitaxel versus paclitaxel. Thus the model underestimates the treatment effect of gemcitabine/paclitaxel versus paclitaxel by about 30% compared to the JHQG trial.

6.6 Assessment of Uncertainty

6.6.1 One-way sensitivity analyses

The MS presents one way sensitivity analyses for a limited number of variables. The majority of variables examined are related to resource use and cost - for example reduction in the

number of women using a wig, number of days in hospital, time spent in hospital receiving chemotherapy. Other variables included in the one way sensitivity analyses were efficacy of treatment and health state utility values. Most of the analyses have little impact on the results reported. The sensitivity analysis does not show the effects of varying the key parameters such as overall survival or time to disease progression in the model. No rationale is reported for the selection of variables included in the one way sensitivity analysis and there is limited justification in the MS for the ranges applied. The manufacturer's response to a request for clarification stated that the variables included in the one way sensitivity analysis were those where ranges around mean or median values, used in the base case, could not be calculated. They stated that these analysis were intended to demonstrate the direction and magnitude of any shift in results with alternative assumptions.

6.6.2 ERG sensitivity analysis

The ERG has conducted one way sensitivity analyses for key parameters in the model (Table 9 reports the values used in the sensitivity analyses and the incremental cost-effectiveness ratios for gemcitabine/paclitaxel with docetaxel as the reference case). The ERG used the 95% confidence intervals for the gemcitabine/paclitaxel parameters, where available. Sensitivity analyses on costs were conducted by varying non-drug costs by plus or minus 25% and reducing the cost of paclitaxel by 50%. Based on these analyses, the most influential variables were overall survival, cost of paclitaxel and health state utilities.

Table 9 ERG one-way sensitivity analyses

Variable	Basecase	Inputs		CE ratios £		Range £
		Lower	Upper	Lower input	Upper input	
Response rates, %	46	39.0	52.9	17,199	17,052	147
Time to progression, weeks	26	21.5	30.5	16,601	17,406	805
Overall survival, weeks	80.60	66.65	94.55	30,446	12,310	18,136
AE discontinuation rate, %	6.7	3.7	9.7	16,335	17,994	1,659
Health state Utilities Stable	0.80	0.65	0.92			
Response	0.72	0.60	0.83	23,656	13,546	10,110
Progression	0.46	0.29	0.63			
Adverse event Utility rates Eg Stable neuropathy	0.70	0.55	0.83	17,396	16,972	424
Non drug costs		-25%	+25%	17,988	16,348	1,640
Post patent paclitaxel cost (50% reduction) Cost / course, £ gemcitabine/ paclitaxel	2,442	1,862	2,442	5,872	17,168	11,296
paclitaxel	1,462	862	1,462			

Changing the assumption over scheduling of febrile neutropenia for gemcitabine/paclitaxel so that all occurrences are in the first two treatment cycles (as for other chemotherapy

regimes) has very little impact on the ICER – which is £17,214 per QALY gained (£9,273 per life year gained) compared to docetaxel..

Changing the assumption over scheduling of response to chemotherapy so that all responses are apparent at the first response cycle also has very little impact on the ICER. For gemcitabine/paclitaxel the ICER is £17,131 per QALY gained (£9,234 per life year gained) when compared to docetaxel and £29,175 per QALY gained (£17,671 per life year gained) when compared to paclitaxel. Assuming that all responses are apparent at the third treatment cycle slightly reduces the ICER for gemcitabine/paclitaxel compared to docetaxel (to £16,231 per QALY gained), but has no effect on the comparison to paclitaxel as the third cycle was assumed to be the first response cycle for both these regimes.

Using utilities values calculated from the regression model for a respondent age of 53 (the median age in the JHQG trial) gives a slightly greater QALY gain for gemcitabine/paclitaxel than in the base case reported in Table 5 – QALY gain of 0.26 and 0.16 when compared with docetaxel and paclitaxel monotherapy respectively. This produces lower ICERs for gemcitabine/paclitaxel against both monotherapies, £15,379 per QALY gained when compared with docetaxel and £27,624 when compared with paclitaxel.

Two further sensitivity analyses were conducted to test the robustness of the estimated cost effectiveness ratios to changes in values of key input data by replacing the pooled estimates with plausible alternative values.

As discussed in section 6.4.4.2, the pooled estimates for overall survival, time to progression and tumour response for paclitaxel are lower than the values observed in the JHQG trial. Replacing the pooled estimates with the values observed in the JHQG trial (reported in the MS) increases the ICER for gemcitabine/paclitaxel against paclitaxel to £42,830 per QALY gained. See Table 10 for the effect of each parameter individually.

Table 10 Incremental outcomes, incremental costs and ICERs for comparison based on JHQG trial data

Analysis	Difference in mean discounted outcomes	Difference in mean discounted total costs	Incremental cost effectiveness ratios
Overall response rate = 26%	Life years 0.25 QALYs 0.15	£ 4,498	Life years £ 17,923 QALYs £ 30,099
Time to disease progression = 17.3 weeks	Life years 0.25 QALYs 0.16	£ 4,649	Life years £ 18,496 QALYs £ 29,560
Overall survival = 68.5 weeks	Life years 0.10 QALYs 0.15	£ 4,641	Life years £ 31,747 QALYs £ 44,899

All above changes simultaneously	Life years QALYs	0.11 0.15	£ 4,790	Life years QALYs	£ 32,339 £ 42,830

The pooled estimate for overall survival with docetaxel monotherapy is lower than the pooled estimate for overall survival with paclitaxel (see column 4, Table 11) – these are the values used to derive the transition probabilities in the base case analysis. However the only clinical trial included in the MS that reports a head-to-head comparison of paclitaxel monotherapy against docetaxel monotherapy (Jones and colleagues[4]) shows a longer survival duration for docetaxel (see column 2, Table 11). The survival benefit for gemcitabine/paclitaxel over paclitaxel in the JHQG trial is similar to that reported for docetaxel over paclitaxel by Jones and colleagues[4]. The ERG undertook an illustrative analysis to examine the possible impact of using the relative (rather than absolute) effects observed in the two clinical trials, adopting a method similar to the classical method for indirect comparisons[11]. This analysis is presented for illustrative purposes and is not a recommendation for adopting this method for conducting indirect comparisons of this type. Using relative effects produces an estimate for overall survival with docetaxel monotherapy that is closer to that observed for gemcitabine/paclitaxel in the JHQG trial (column 5, Table 11). Using this value in the model generates an ICER for gemcitabine/paclitaxel compared to docetaxel monotherapy of £45,811 per QALY gained.

Table 11 Overall survival from studies and values used in model

Parameters	Jones and colleagues[4]	JHQG	Parameters used in model	Estimated values from indirect comparison
Gemcitabine/ Paclitaxel	N/A	80.60	80.60	80.60
Docetaxel	66.7	N/A	59.44	78.2
Paclitaxel	55.0	68.50	61.96	68.50

6.6.3 Scenario Analysis

A set of alternative scenarios reported in the MS include cost-effectiveness estimates using alternative comparators (paclitaxel monotherapy and docexaxel/capecitabine combination therapy rather than docetaxel monotherapy) and additional estimates using a price reduction expected once the patent for paclitaxel expires (the current price of branded Taxol was used in the base case analysis).

The incremental costs, incremental effects and incremental cost effectiveness ratios for gemcitabine/paclitaxel against the alternative comparators (paclitaxel monotherapy and docetaxel/capecitabine combination therapy) are reported in Table 5 (see section 6.3). The

ICERs for gemcitabine/paclitaxel are less favourable with these alternative comparators than for the base case comparison against docetaxel monotherapy.

The post-patent expiration price reduction improves the cost effectiveness estimates for gemcitabine/paclitaxel against regimes that do not include paclitaxel (i.e. docetaxel monotherapy and docetaxel/capecitabine combination therapy). The ICER for gemcitabine/paclitaxel with docetaxel as the reference case, with the post-patent expiration price reduction is reported in Table 5 (section 6.3). The size of price reduction assumed in the scenario analysis is supported by CIC data included as an appendix to the MS, but is substantially greater than the difference between Taxol and generic paclitaxel listed in the current BNF (No. 51).

The MS reports a threshold analysis which assesses the likely 'breakpoint' for the cost effectiveness of gemcitabine/paclitaxel by ranging the survival estimates for the reference case. The threshold analysis shows that the cost effectiveness estimate remains below £30,000 if overall survival with docetaxel is increased to 70 weeks (overall survival with docetaxel is 59.44 weeks in the base case). However, there are many uncertainties attached to the estimates for pooled overall survival and with regard to indirect comparisons between gemcitabine/paclitaxel and docetaxel, - these are discussed in sections 6.6.2 and 6.8.

A further analysis is reported in the tables of one-way sensitivity analyses, related to dose reduction for docetaxel, which may be better described as a scenario analysis, since the overall survival and response rates used in the base case analysis have been replaced with values observed in a trial evaluating the lower dosage, as well as a reduction in chemotherapy drug cost. The ICER for gemcitabine/paclitaxel compared to docetaxel at this reduced dosage is lower than for the base case (£15,918 per QALY gained and £9,323 per life year gained).

6.6.4 ERG scenario analysis

A scenario analysis was conducted using effectiveness data from the JHQG trial for both gemcitabine/paclitaxel and paclitaxel, and the pooled estimates from trials including anthracycline pre-treated patients for other chemotherapy regimes. Investigator-assessed (rather than independently assessed) overall response rates were used and the assumed hospitalisation rate for febrile neutropenia was increased to 75%. This resulted in an ICER for gemcitabine/paclitaxel compared to paclitaxel of £49,842 per QALY gained. The ICER for gemcitabine/paclitaxel compared to docetaxel was £15,692 per QALY gained.

6.6.5 Probabilistic Sensitivity Analysis

The Industry model has a probabilistic sensitivity analysis on the *Model Set up* Excel spreadsheet. The PSA can be run by clicking on the 'Run PSA Analysis' button and takes about an hour and a half to run 1000 iterations (on a computer with a 2.8 GHz processor). Results are updated on the *PSA Results* Excel sheet.

The PSA reported in the MS was performed for selected input variables - these were utility weights, tumour response, time to progression, overall survival duration and chemotherapy discontinuation. No rationale is included in the MS explaining why these variables were selected for inclusion in the PSA. These are key variables likely to have a direct impact on the effectiveness of treatment and while the choice of parameters seems reasonable there is no accounting for variation in cost. Given that chemotherapy costs constitute the majority (76%-80%) of the estimated treatment costs in the model, it is surprising that no variables related to chemotherapy costs (for example, body surface area) were included in the probabilistic sensitivity analysis. It was noted earlier (section 6.4.4.4) that the BSA estimate used in the model is not derived from clinical trial populations (or indeed for patients with metastatic breast cancer) – it seems likely that the manufacturer would have access to sufficient data on BSA for the MBC clinical trial populations to estimate measures of average and variation, as well as to make a judgement on an appropriate sampling distribution for these data. Also, given that all resource assumptions for the management of adverse events were based on clinical opinion, it might have been informative to include these in the probabilistic sensitivity analysis to characterise further the uncertainty around the costs in the model. For example, the number of inpatient days required for adverse events requiring hospitalisation could have been sampled with a mean of 5 days, lower limit of 2 days and upper limit of 14 days (as in the one way sensitivity analysis).

The manufacturer provided clarification on the choice of variables included in the PSA, stating that these were the variables with "pre-described" or calculable ranges to represent uncertainty around the averages used in the base case (see Appendix 1 – ERG's questions and manufacturer's response). A further PSA was submitted with the clarification, which includes some additional variables (use of wig by alopecia sufferers, non-drug costs and hospital stay for treatment of serious adverse events) using ranges adopted for the one way sensitivity analyses.

The results of the PSA are presented as a scatterplot of the cost effectiveness results (for gemcitabine/paclitaxel with docetaxel as the comparator) and as acceptability curves, with each intervention (gemcitabine/pacxlitaxel, paclitaxel monotherapy, docetaxel/capecitabine

and docetaxel/gemcitabine) against docetaxel - MS pages 136-138. The interpretation of the CEACs in the MS is limited to a consideration of the probability that an intervention is cost-effective (relative to docetaxel monotherapy) at an arbitrary threshold of £35,000 – which is 70% for gemcitabine/paclitaxel. This presentation assumes that docetaxel is the only current treatment option against which alternatives should be evaluated. However the MS describes three currently licensed taxane-based treatments available in the UK (two monotherapy and one combination therapy), which gemcitabine/paclitaxel might be expected to displace and against which it should be evaluated. In this situation a presentation with multiple CEACs and a frontier analysis may be more appropriate (see Figure 4).

The ranges and distributions used for the PSA are reasonable although the ERG suggests that, rather than normal distributions, the gamma distribution would be more appropriate for time related inputs (such as overall survival) and the beta distribution for sampling proportions (such as response and chemotherapy discontinuation, which are both described in MS as rates). While the ranges on these variables mean they are unlikely to produce values beyond their logical bounds (i.e. durations should be sampled between zero and infinity and probabilities between zero and one) more appropriate choices of distributions would have assured this.

Of particular concern in the design of the PSA is that the survival duration, median time to progression for responders and median time to progression for the whole treated cohort are sampled independently. Since the mortality probability in the model is derived from both median time to disease progression and median overall survival, there is an implicit assumption that median survival exceeds time to disease progression. While the ranges on these variables make it unlikely that a sampled value for median time to disease progression would exceed a sampled value for median survival, there is nothing in the design of the PSA to avoid this happening – nor do there appear to be traps to ensure that unfeasible values will be discarded. The assumption that median time to disease progression and median overall survival duration are entirely independent does not seem valid. It seems likely that longer disease-free survival for responding patients would be a principal factor driving increased survival duration. Similarly longer median time to disease progression for responders would be likely to be reflected in a longer median time to disease progression for the whole treated cohort.

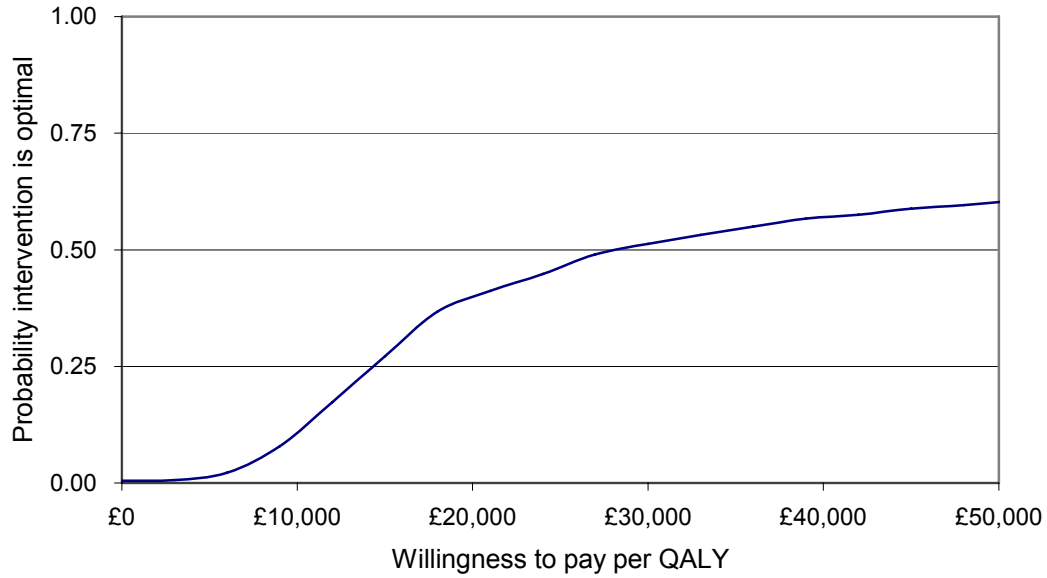
6.6.5.1 ERG probabilistic sensitivity analysis

The PSA was re-run with the pooled estimates for overall survival, time to disease progression and overall response rate for paclitaxel monotherapy replaced by values (both

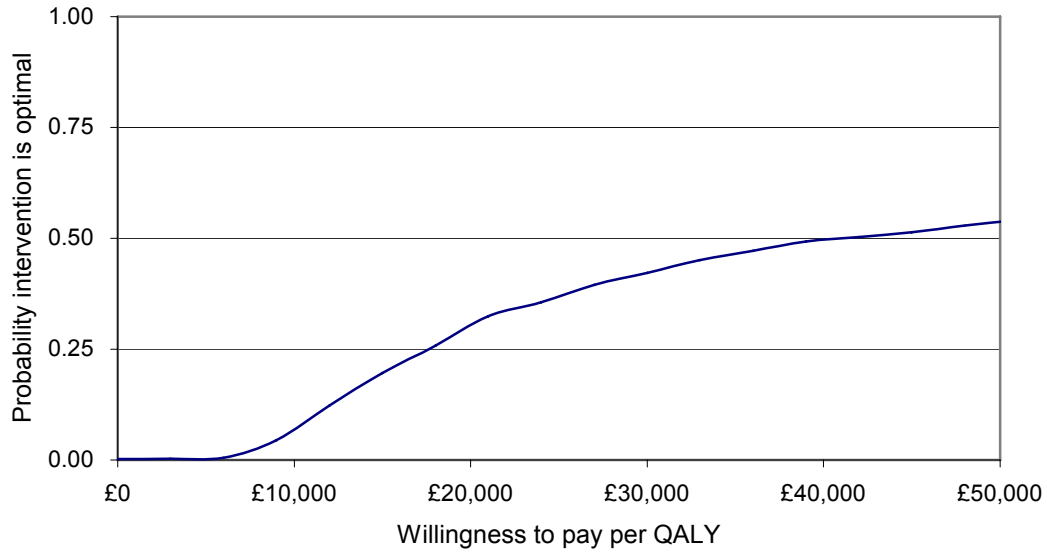
point estimates and standard errors) from the JHQG trial. This gives a less favourable CEAC than for the base case comparison using paclitaxel as comparator (Figure 3). The gemcitabine/paclitaxel combination has a lower probability of being cost-effective at all willingness to pay values. At the threshold willingness to pay of £35,000 per QALY adopted in the MS, the probability of gemcitabine/paclitaxel being cost effective, compared to paclitaxel monotherapy, is 46% whereas the probability was 55% using the pooled estimates. The corresponding values at a threshold of £30,000 per QALY are 51% for a comparison based on pooled estimates and 42% for a comparison using inputs based on the JHQG trial.

Figure 3 CEACs for gemcitabine/paclitaxel against paclitaxel using (a) pooled estimates used in base case analysis and (b) values from JHQG trial

(a)



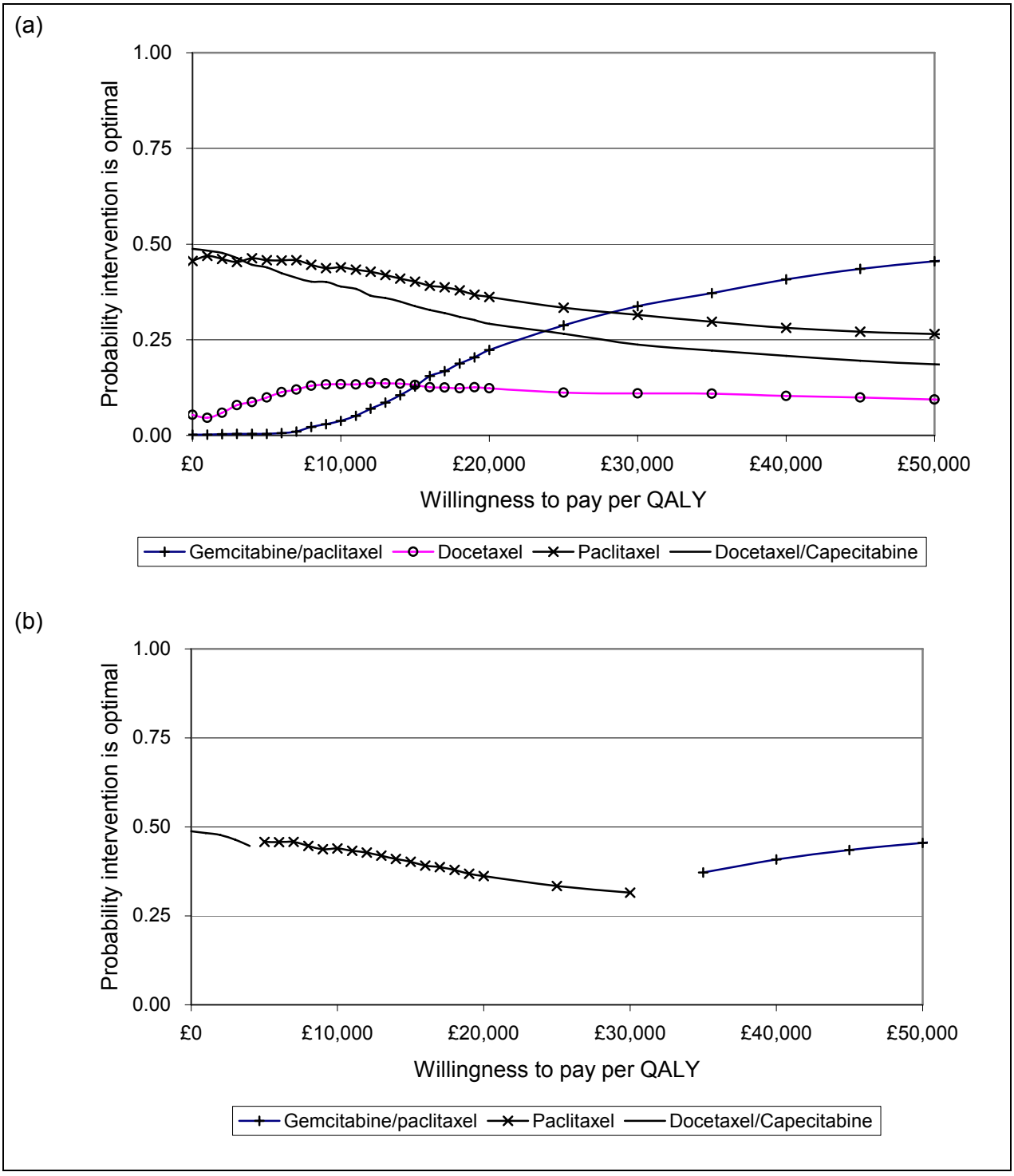
(b)



This indicates that the results of this analysis are sensitive to choice of included trials and it would appear from this and the preceding analyses that the most significant factor affecting the cost-effectiveness estimates is the overall survival benefit assumed for gemcitabine/paclitaxel.

Figure 4 shows the CEACs comparing each of the four taxane-based chemotherapy regimes against each other, rather than the presentation in the MS where all regimes were compared against docetaxel. The analysis presented here is based on the pooled estimates of clinical effectiveness for all chemotherapy regimes. Gemcitabine/paclitaxel becomes the optimal choice at higher willingness to pay thresholds – around £30,000/QALY. The closest alternative regime, which is optimal at lower willingness to pay values, is paclitaxel monotherapy. There is a large degree of uncertainty over which regime is most likely to be cost effective, with none of the probabilities exceeding 50%. Figure 4 should be interpreted with caution, given the reservations outlined in sections 6.6.2 and 6.8 regarding the validity of the pooled estimates used as inputs to the model. For example, docetaxel appears to be dominated by other options over the range of willingness to pay shown. However it was noted in section 6.6.2 that, while the pooled estimate for overall survival with docetaxel is lower than for paclitaxel monotherapy, the only included trial that directly compared docetaxel and paclitaxel showed a significant survival benefit with docetaxel (15.4 months with docetaxel versus 12.7 months with paclitaxel, hazard ratio of 1.41, p=0.03).

Figure 4 (a) Multiple CEACs comparing four taxane-based chemotherapy regimes (b) CEA frontier



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

Overall, the model structure adopted for the MS seems reasonable and is based on that adopted in previous economic evaluations of chemotherapy for women with metastatic

breast cancer. Key assumptions that have been adopted, in common with previous studies are that:

- Survival functions (overall survival and time to progression) can be fully inferred from the median, assuming an exponential function;
- Response at each cycle can be inferred from the total response observed in trials, assuming an exponential function. However this was modified to ensure the modelled response at the first response cycle matched values observed in clinical trials. It should also be noted that the MS assumes 6 cycles of chemotherapy are given whereas in the design of the JHQG trial was that chemotherapy should continue until disease progression occurred. Analysis presented on page 78 of the MS showed that the majority of responding patients had shown response by the sixth chemotherapy cycle;
- No mortality risk is applied until entry to progressive state. Breast cancer is likely to be the greatest cause of mortality for women with MBC and clinical opinion suggests that this would occur when disease progresses rather than with stable disease or response to chemotherapy;
- Toxicity over the trial period can be converted to cycle probabilities by conventional assumptions. Includes typical assumption of developing only one toxicity in a given cycle, with risk in subsequent cycles independent of current experience. Assumption over scheduling of febrile neutropenia had little impact when tested in ERG sensitivity analysis;

The principal source of concern is over the method adopted for pooling results and performing indirect comparisons. The MS states on page 147 that

> "the model is based on pooled weighted absolute outcomes – which is not as satisfactory in terms of data quality. This therefore places a high weight on larger studies and also has implications in terms of **ensuring patient cohorts are as comparable as possible**." (ERG emphasis)

While the MS has discussed the comparability of patient populations in three trial reports,[1;4;5] there is no discussion of the comparability of patient populations in the remaining 12 trials that contribute data included in the economic model. Nor is there any quality assessment of the methodology or reporting of these clinical trials, other than the presentation of data extracted from these reports in tables in Appendix 12.

6.8 Summary of uncertainties and issues

In general, the approach taken to model disease progression and cost effectiveness in this patient group seems reasonable. However concerns over the comparability of studies

contributing input data to the model and the validity of the indirect comparison method adopted have lead the ERG to raise the following issues:

- Clinical trials used to derive inputs for economic model have not been reviewed or quality assessed in clinical effectiveness section. One trial was excluded from the clinical effectiveness review, but has contributed data inputs for the economic model;
- There is an inadequate assessment of comparability of trials (both methodologically and in terms of patient populations) included within the comparison – no formal assessment of heterogeneity was reported;
- The method of pooling data inputs has been called into question – absolute values for median survival duration, median time to progression, overall response rates and occurrence of adverse events have been extracted and pooled (breaking randomisation). The method of indirect comparison may be best described as naive according to an HTA methodological review[11].

Additional issues, not related to concerns over the pooling of data for model inputs, raised during the review of the MS are:

- The ERG has identified a number of mistakes in the Excel spreadsheet submitted. However these did not have any substantial impact on results;
- The sensitivity analyses undertaken are limited, and there may be a greater variability in the cost effectiveness of treatment than presented in the manufacturer's submission.

Overall survival appears to be the key variable affecting cost-effectiveness – relative effectiveness between paclitaxel monotherapy and gemcitabine/paclitaxel combination therapy has been established by a head-to-head comparison in the JHQG trial. The relative effectiveness of gemcitabine/paclitaxel combination therapy against docetaxel monotherapy – the comparator for UK practice – has not been established. An illustrative analysis suggests that the survival benefit for gemcitabine/paclitaxel over docetaxel in the economic model may be over-estimated.

7 DISCUSSION

7.1 Summary of clinical effectiveness issues

The manufacturer suggests that gemcitabine should be considered as one option for first line therapy for MBC in some patients, but does not appear to advocate that it should replace any of the current taxane treatments. The manufacturer's submission to NICE includes a systematic review of the clinical effectiveness literature and tabulated/narrative reporting of the (unpublished) JHQG trial data. Only interim analyses from the JHQG trial are available in the literature, reported in conference abstracts. Results from two other published trials are included in the submission to provide a tabulated comparison with docetaxel monotherapy and docetaxel/capecitabine combined therapy. However, there are underlying differences in the patient characteristics in the comparator trials, notably the number of lines of prior therapy received, and the trials have not been subject to a formal indirect comparison.

In the absence of any formal methods of indirect comparison, there is insufficient robust evidence to compare the relative effectiveness of gemcitabine/paclitaxel with docetaxel monotherapy or docetaxel/capecitabine combination therapy. A head-to-head RCT of gemcitabine/paclitaxel compared with docetaxel would provide more relevant evidence of the effectiveness of gemcitabine with respect to the UK setting.

7.2 Summary of cost effectiveness issues

The general approach to model disease progression and cost effectiveness in this patient group presented in the MS and accompanying electronic model seems reasonable. However, data inputs come from trials which have been inadequately assessed, both in terms of study quality and patient heterogeneity. Pooling of absolute values from these trials would have broken the effects of randomisation.

The evidence for gemcitabine's clinical effectiveness comes from an RCT comparing gemcitabine/paclitaxel with paclitaxel. However, the economic evaluation uses docetaxel as the comparator in the reference case, with additional scenario analyses using alternative comparators (including paclitaxel). The relative effectiveness of gemcitabine/paclitaxel combination and docetaxel – the comparator for UK practice – has not been established, and an illustrative analysis suggests that the survival benefit for gemcitabine/paclitaxel combination over docetaxel in the economic model may be over-estimated. The sensitivity analyses undertaken are limited, and there may be a greater variability in the cost effectiveness of treatment than presented in the manufacturer's submission.

8 REFERENCES

(1) Albain KS, Nag S, Calderillo-Ruiz G, Jordaan JP, Llombart A, Pluzanska A et al. Global phase III study of gemcitabine plus paclitaxel (GT) vs. paclitaxel (T) as frontline therapy for metastatic breast cancer (MBC): First report of overall survival. J Clin Oncol 2004; 22(14):5S.

(2) Moinpour C, Wu J, Donaldson G, Liepa A, Melemed A, Oshaughnessy J et al. Gemcitabine plus paclitaxel (GT) versus paclitaxel (T) as first-line treatment for anthracycline pre-treated metastatic breast cancer (MBC): Quality of life (QoL) and pain palliation results from the global phase III study. J Clin Oncol 2004; 22(14):32S.

(3) O'Shaughnessy J, Nag S, Calderillo-Ruiz G, Jordaan J, Llombart A, Pluzanska A et al. Gemcitabine plus paclitaxel (GT) versus paclitaxel (T) as first-line treatment for anthracycline pre-treated metastatic breast cancer (MBC): interim results of a global phase III study [abstract 25]. Proceedings of the American Society of Clinical Oncology 2003; 22:7.

(4) Jones SE, Erban J, Overmoyer B, Budd G, Hutchins L, Lower E et al. Randomized Phase III Study of Docetaxel Compared With Paclitaxel in Metastatic Breast Cancer. J Clin Oncol 2005; 23(24):5542-5551.

(5) O'Shaughnessy J, Miles D, Vukelja S, Moiseyenko V, Ayoub JP, Cervantes G et al. Superior survival with capecitabine plus docetaxel combination therapy in anthracycline-pretreated patients with advanced breast cancer: phase III trial results. Journal of Clinical Oncology 2002;(12):2812-2823.

(6) National Institute for Health and Clinical Excellence. Technology Appraisal No. 54. Guidance on the use of vinorelbine for the treatment of advanced breast cancer. 2002.

(7) National Institute for Health and Clinical Excellence. NICE Technology Appraisal 62. Guidance on the use of capecitabine for the treatment of locally advanced or metastatic breast cancer. 2003.

(8) Winer EP, Berry DA, Woolf S, Duggan D, Kornblith A, Harris LN et al. Failure of Higher-Dose Paclitaxel to Improve Outcome in Patients With Metastatic Breast Cancer: Cancer and Leukemia Group B Trial 9342. J Clin Oncol 2004; 22(11):2061-2068.

(9) Long-term survival demonstrated with trastuzumab plus docetaxel: 24-month data from a randomised trial (M77001) in HER2-positive metastatic breast cancer. 2005.

(10) Gemcitabine plus docetaxel (GD) versus capecitabine plus docetaxel (CD) for anthracycline-pretreated metastatic breast cancer (MBC) patients: results of a European Phase III study. 2005 ASCO Annual Meeting; 2005.

(11) Glenny AM, Altman DG, Song F, Sakarovitch C, Deeks JJ, D'Amico R et al. Indirect comparisons of competing interventions. Health Technol Assess 9[26]. 2005.

(12) Sledge GW. Gemzar, paclitaxel and trastuzumab in metastatic breast cancer. Oncology 2003; 17(12):33-35.

(13) Hutton J, Brown R, Borowitz M, Abrams K, Rothman M, Shakespeare A. A new decision model for cost-utility comparisons of chemotherapy in recurrent metastatic breast cancer. Pharmacoeconomics 1996; 9(Suppl 2):8-22.

(14) Brown RE, Hutton J. Cost-utility model comparing docetaxel and paclitaxel in advanced breast cancer patients. Anti-Cancer Drugs 1998; 9:899-907.

(15) Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. Cancer 1981; 47(1):207-214.

(16) Miller DK, Homan SM. Determining Transition-Probabilities - Confusion and Suggestions. Medical Decision Making 1994; 14(1):52-58.

(17) Beck RJ, Pauker SG, Gottlieb JE, Klein K, Kassirer JP. A convenient approximation of life expectancy (the "DEALE"). II. Use in Medical Decision Making. American Journal of Medicine 1982; 73:889-897.

(18) Health state descriptions for metastatic breast cancer: a qualitative study. International Society for Pharmacoeconomics and Outcomes Research, 8th European Congress, Florence, Italy; 2005.

(19) Curtis L, Netten A. Unit Costs of Health and Social Care. Canterbury: Personal Social Services Research Unit, University of Kent; 2005.

(20) Perez EA. Current management of metastatic breast cancer. Seminars in Oncology 1999; 26(4:Suppl 12):1-10.

(21) Blum JL, Dieras V, Lo Russo PM, Horton J, Rutman O, Buzdar A et al. Multicenter, Phase II study of capecitabine in taxane-pretreated metastatic breast carcinoma patients. Cancer 2001; 92(7):1759-1768.

(22) Academic or commercial information removed

(23) Cooper NJ, Abrams KR, Sutton AJ, Turner D, Lambert PC. A Bayesian approach to Markov modelling in cost-effectiveness analyses: application to taxane use in advanced breast cancer. Journal of the Royal Statistical Society Series A-Statistics in Society 2003; 166:389-405.

(24) Brown RE, Hutton J, Burrell A. Cost effectiveness of treatment options in advanced breast cancer in the UK. Pharmacoeconomics 2001; 19(11):1091-1102.

(25) Launois R, Reboul-Marty J, Henry B, Bonneterre J. A cost-utility analysis of second line chemotherapy in metastatic breast cancer: Docetaxel versus Paclitaxel versus Vinorelbine. Pharmacoeconomics 1996; 10(5):504-521.

(26) GAD - Government Actuaries Department. Current Interim Life Tables (2002-2004). http://www.gad.gov.uk/Life_Tables/Interim_life_tables.htm . 2006.

(27) Albain KS, Nag S, Calderillo-Ruiz G, Jordaan JP, Llombart A, Pluzanska A et al. Global phase III study of gemcitabine plus paclitaxel (GT) vs. paclitaxel (T) as frontline therapy for metastatic breast cancer (MBC): First report of overall survival. Internet [2004 [cited 2006 Apr. 1]; Available from: URL:http://www.asco.org/portal/site/ASCO/menuitem.34d60f5624ba07fd506fe310ee37a01d/?vgnextoid=76f8201eb61a7010VgnVCM100000ed730ad1RCRD&vmview=abst_detail_view&confID=26&abstractID=2708

(28) Collett D. Parametric proportional hazards models. Modelling survival data in medical research. 2nd ed. ed. Boca Raton, Fla.: Chapman & Hall/CRC; 2003. 151-193.

(29) Mouridsen H, Harvey V, Semiglazov V, et al. Phase III study of docetaxel 100 versus 75 versus 60 mg/m2 as second line chemotherapy in advanced breast cancer. Abstract 327. Breast Cancer Research and Treatment 2002; 76(suppl 1):S88.

(30) Icli F, Akbulut H, Uner A, Yalcin B, Baltali E, Altinbas M et al. Cisplatin plus oral etoposide (EoP) combination is more effective than paclitaxel in patients with advanced breast cancer pretreated with anthracyclines: a randomised phase III trial of Turkish Oncology Group. British Journal of Cancer 2005; 92(4):639-644.

(31) Feeny D, Furlong W, Boyle M, Torrance GW. Multi-attribute health status classification systems. Health Utilities Index. Pharmacoeconomics 1995; 7(6):490-502.

(32) Feeny D, Furlong W, Barr RD, Torrance GW, Rosenbaum P, Weitzman S. A comprehensive multiattribute system for classifying the health status of survivors of childhood cancer. J Clin Oncol 1992; 10(6):923-928.

(33) Jones L, Westwood M, Wright K, Riemsma R, Hawkins N, Richardson G. A rapid and systematic review of the clinical effectiveness and cost-effectiveness of capecitabine (Xeloda®) for locally advanced and/or metastatic breast cancer. Report commissioned by NHS R&D HTA Programme on behalf of The National Institute for Clinical Excellence. http://www.nice.org.uk/download.aspx?o=52452 . 2002.

(34) Joint Formulary Committee. British National Formulary. Volume 50. London: British Medical Association and Royal Pharmaceutical Society of Great Britain; 2005.

(35) Sjostrom J, Blomqvist C, Mouridsen H, Pluzanska A, Ottosson-Lonn S, Bengtsson NO et al. Docetaxel compared with sequential methotrexate and 5-fluorouracil in patients with advanced breast cancer after anthracycline failure: a randomised phase III study with crossover on progression by the Scandinavian Breast Group. European Journal of Cancer 1999; 35(8):1194-1201.

(36) Nabholtz JM, Senn HJ, Bezwoda WR, Melnychuk D, Deschenes L, Douma J et al. Prospective randomized trial of docetaxel versus mitomycin plus vinblastine in patients with metastatic breast cancer progressing despite previous anthracycline-containing chemotherapy. 304 Study Group. J Clin Oncol 1999; 17(5):1413-1424.

(37) Bonneterre J, Roche H, Monnier A, Guastalla JP, Namer M, Fargeot P et al. Docetaxel vs 5-fluorouracil plus vinorelbine in metastatic breast cancer after anthracycline therapy failure. British Journal of Cancer 2002; 87(11):1210-1215.

(38) Chan S, Friedrichs K, Noel D, Pinter T, Van BS, Vorobiof D et al. Prospective randomized trial of docetaxel versus doxorubicin in patients with metastatic breast cancer. J Clin Oncol 1999; 17(8):2341-2354.

(39) Nabholtz JM, Gelmon K, Bontenbal M, Spielmann M, Catimel G, Conte P et al. Multicenter, randomized comparative study of two doses of paclitaxel in patients with metastatic breast cancer. J Clin Oncol 1996; 14(6):1858-1867.

(40) Gradishar WJ, Tjulandin S, Davidson N, Shaw H, Desai N, Bhar P et al. Phase III trial of nanoparticle albumin-bound paclitaxel compared with polyethylated castor oil-based paclitaxel in women with breast cancer. J Clin Oncol 2005; 23(31):7794-7803.

(41) Sledge GW, Neuberg D, Bernardo P, Ingle JN, Martino S, Rowinsky EK et al. Phase III trial of doxorubicin, paclitaxel, and the combination of doxorubicin and paclitaxel as front-line chemotherapy for metastatic breast cancer: an intergroup trial (E1193). J Clin Oncol 2003; 21(4):588-592.

Appendix 1 – ERG's questions and manufacturer's response

Section A: Clinical evidence

A1. Please provide statistical evaluations of heterogeneity for the studies from which absolute efficacy estimates were pooled (section 2.7, question 59, page 55). We specifically request that homogeneity in patients' characteristics and degree of metastatic setting is evaluated using a method such as the graphic approach and Q statistic.

Information extracted from the phase III RCTs on patient characteristics was used to assess the homogeneity of patients in a non-quantitative manner. This was due to differences in the reporting of key characteristics such as the site(s) of metastases and the number of metastatic sites. Tables A1-A3 (shown in Appendix 1.1 below) illustrates the variation between studies in how these data were presented. There was not one single variable upon which a formal (quantitative) comparison of homogeneity could be reliably performed, with the exception of data on patients' age (which are of limited value in terms of demonstrating the comparability of patient populations in this context)

A2. Please provide justification for the exclusion of a third abstract: *Moinpour, C. et al. (2004)* from Table 1 given that the inclusion/exclusion criteria for the systematic review do not specify particular outcomes. Two abstracts are cited for the JHQG study, but the submission does not include this third abstract: "Gemcitabine plus paclitaxel (GT) versus paclitaxel (T) as first-line treatment for anthracycline pre-treated metastatic breast cancer (MBC): Quality of life (QoL) and pain palliation results from the global phase III study"; Journal of Clinical Oncology 22 (14) 32(S).

The abstract by Moinpour (2004) was not retrieved in the systematic search conducted by Lilly, although there was no reason for this considering the inclusion/exclusion criteria.

The quality of life and pain palliation data presented in the submission and in this abstract were based on JHQG clinical trial, therefore the information is not expected to differ. Clinical outcome data were consistent with data presented by Albain K et al, Proc ASCO 2004.

A3. Please provide further details relating to the quality of life data presented in response to question 54 (pages 51 – 54). Specifically, we request:

- **The absolute quality of life scores underlying the % change from baseline depicted in Figure 5.**

The table below provides a summary of the RSCL overall valuation of life scale scores for randomized patients with RSCL data. A high score represents better quality of life on the overall valuation of life item.

Table1. Academic or commercial information removed

In Figure 5 on the submission the percentage change was estimated from the above table - i.e. 6.7 on a scale of 0 - 100 represents 6.7% change. The mean change were plotted instead of the absolute scores

- **Further details regarding the study of pain alleviation (page 53); in particular a definition of what is meant by 'Improved' in Table 12 and a brief assessment of the strengths and weaknesses of this study.**

Health outcomes were measured through both objective and subjective endpoints. The corroboration of endpoints would increase the confidence in the clinical significance of the results. Subjective endpoints were obtained from patients by completion of the Brief Pain Inventory (BPI) and Rotterdam Symptom Checklist (RSCL). Objective measures, prospectively determined as important in patients, included class of pain medications consumed (analgesic level).

One hundred forty-one patients on the GT Arm and 150 patients on the T Arm completed at least one BPI. A total of 231 patients did not complete the BPI because of the lack of validated translations, and 7 patients did not complete the BPI for other reasons. For investigational sites with validated translations, on-study compliance rates (defined as the number of questionnaires completed divided by the total number of expected questionnaires based on cycles administered) were 84.9% for the GT Arm and 84.6% for the T Arm.

For patients who completed the BPI, analyses were performed on both an ITT population and a subset population that included only those patients who were symptomatic at baseline. The symptomatic subset included all patients with a baseline analgesic level ≥ 1 (that is, use of any analgesics). Eighty-one patients on the GT Arm and 71 patients on the T Arm were considered symptomatic at baseline. Analgesic level was well balanced between the treatment arms for all randomized and symptomatic patients. The distribution of analgesic level for the patients with BPI data was similar to that of all randomized patients; this suggests that results from patients with BPI data could be extrapolated to the entire study population.

The analyses of the BPI data included the analysis of a single question, referred to as "worst pain," and the analysis of the mean of seven questions that addressed the impact of pain on various aspects of life, referred to as "mean BPI interference items". Scores for the BPI are reported on a scale of 0 to 10, with 0 representing no pain or the pain does not interfere with daily living.

When the data were summarized at the individual patient level, more patients on the GT Arm reported at least two consecutive improvements from baseline in analgesic level during the course of therapy. These improvements were noted primarily in patients with baseline analgesic levels of 1 and 2. Table 12 in the submission summarizes the results of improved analgesic level.

'Improvement' was defined as a score better than baseline (i.e., lower or decreased) over >=2 consecutive cycles.

Strengths and weaknesses of pain analyses (BPI and analgesic level)

Strengths

- BPI is a reliable and valid tool for pain assessment; only used translations that had been validated
- objective (analgesic level) and subjective assessment

- analyses and subgroup of symptomatic patients were defined in a priori in protocol
- on-study compliance was relatively good (85%)
- subgroups appear to be balanced (although baseline pain or analgesic level were not part of stratification, other factors such as KPS and visceral/non visceral disease may have helped control for any imbalance)
- improvement in analgesic level needed to be sustained (much like tumour response criteria)

Weaknesses

- only subgroup of patients participated in BPI analysis due to lack of validated translations
- small proportion of patients with pain at baseline and even those were managed with weak analgesics (NSAIDs); however, this is consistent with the performance status of these patients; little opportunity for numerical or statistical improvement
- no quantitative data on analgesic use (analgesic diary which may note changes in analgesic consumption within an analgesic level)
- no long-term data are available

A4. Please clarify the difference between 'death' (0 in T arm, 2 in GT arm) and 'death from study disease' (2 in T arm, 8 in GT arm) as presented in Table 6, Summary of Patient Disposition by Reason for Discontinuation (page 38).

Academic or commercial information removed

A5. Please provide justification for the inclusion of 'ovarian neoplasms' in the search terms: pages 175-8, appendix 1.6.

'Ovarian neoplasms' in the search strategy because we used a pre-existing strategy suggested by NICE: http://www.nice.org.uk/pdf/taxanesreviewhtareport.pdf (pages 63 & 64).

This search term is unlikely to have resulted in exclusion of relevant paper but may have only contributed to volume rather than compromise the accuracy of the search.

A6. Please clarify the treatment pathway for patients diagnosed with Stage IV breast cancer and explain why these patients are ineligible for GT as indicated in the flow chart given in Appendix 1, page 155. What happens to those patients?

Patients who are diagnosed with metastatic disease and who have not received prior chemotherapy would not be eligible to receive GT. This is in line with the gemcitabine licence for breast cancer which states that patients are required to have received one anthracycline-based chemotherapy regimen in the adjuvant/neoadjuvant setting. A non-anthracycline-based regimen in the adjuvant/neoadjuvant setting is required if use of an anthracycline was clinically contraindicated.

Almost all the patients in JHQG received prior chemotherapy (GT =100%, T = 99.2%) in the adjuvant/neo-adjuvant setting.

Section B: Cost Effectiveness

B1. Please give a brief explanation for the choice of the variables used for probabilistic sensitivity analysis (PSA) and clarify how many iterations were performed. Although the scatter plots in the submission indicate that a larger number of iterations were performed, there only appear to be ten in the Excel spreadsheet submitted. The report is clear as to the variables included in the PSA (page 129), but there is no discussion as to why those particular variables were chosen and others were excluded (for example, assumptions over the scheduling of response rates which are included in the one-way sensitivity analysis; probability of developing toxicity, or treatment costs).

The PSA variables were selected as the key model parameters, deemed most likely to have impact on the cost effectiveness results, for which we had credible ranges pre-described (or calculable) to represent the level of uncertainty in the mean or median parameter values.
These variables covered

- utility weights
- tumour response rates
- time to progression (All)
- time to progression (responders)
- overall survival duration
- AE discontinuation rate

All the remaining sensitivity one-way analyses were based on scenarios where we considered the impact on the results from setting the model to a range of alternative assumptions or treatment scenarios – for example using a different assumption on body surface area (impacting on treatment costs).
These one-way analyses are therefore intended to demonstrate the scale of shift in results that alternative assumptions would have – in fact these had little impact overall and confirmed that the main area of variability lay in the key PSA parameters.
The standard PSA results and CEAC curves were based on a data set of incremental costs and benefits based on 5000 iterations of the model (the model version provided was set at 10 iterations purely to limit the file size for electronic transfer – and no analyses were based on this)

B2. Please perform a full PSA across a wider range of parameters, including as a minimum all of those which are varied in the one- and multi-way sensitivity analyses. Please state the number of iterations performed.

A sub-set of the one way analyses did vary specific parameters across a value range
Use of wig for alopecia sufferers (50-100%)
Non drug unit cost variation (85% to 115%)
Duration of hospitalisation for treatment of serious AEs (2 to 14 days)
We have therefore run a set of PSA analyses which also include these variables – based on the min max range assumed and also adopting a uniform distribution across the range (maximising the variability).
Please see accompanying CD for updated PSA analyses.

B3. Please clarify what is shown in the cost effectiveness acceptability curves presented on pages 137 to 139. Is each intervention being independently compared against a common comparator?

The CEA curves on pages 137 and 139 show the probability of reaching cost effectiveness for a range of cost effectiveness threshold values (cost per QALY and cost per life year values).
The treatments have been independently compared against a common comparator based on monotherapy docetaxel 100mg/m2 (therefore docetaxel does not have a line on the charts)

B4. Please present separate cost effectiveness acceptability curves which show the incremental cost effectiveness of each treatment option versus the comparator treatment.

We have re-run and provided combined and separate CEACs for the treatments – based on the new PSA including the 3 additional parameters identified above in B2.
Please refer to accompanying CD for separate CEAC curves for each treatment for QALY and LY - as well as the multiple CEAC charts we originally provided.

B5. Please confirm whether the expected further chemotherapy cost that is applied to each cycle only applies to those who have newly entered the progressive state in the corresponding cycle.

Yes, the further chemotherapy cost is only applied as a one off total cost (based on an expected number of cycles of 3rd line treatment) applied at the point of progression.

B6. Please confirm whether the treatment discontinuation rates listed in table 35 (page 100) are pooled estimates.

Yes, the discontinuation rates are treated in the same way as the efficacy variables and have been pooled for treatments that have multiple trial sources.

B7. Please provide the additional analyses of clinical trial data relating to the table of assumptions (on page 109) about scheduling of response rates.

See Appendix 2 for additional analyses based on S273 and JHQG

B8. Please provide the additional analyses of clinical trial data relating to the table of assumptions (page 110) about time to disease progression - differentiating time to disease progression for responders and non-responders.

See Appendix 3 for additional analyses based on S273 and JHQG

B9. Tumour response rates:

- **Please clarify why the submission states that investigator assessment will usually give higher response rates than independent assessment (page 121 of the submission) yet the proportion is higher for independent assessment in the GT arm of JHQG trial (proportion is identical for T arm).**

  In general one might anticipate investigator-assessed tumour response may be subject to a degree of observer bias favouring the new treatment, despite using objective criteria. With independent assessment of imaging this is eliminated and a more structured, accurate assessment of the extent of disease at all time points in the study is obtained. We cannot explain the reason for the higher proportion of response in the independently assessed tumour response in JHQG other than as a reflection of the subjectivity of investigator assessment.

- **Please explain why the number of cases assessed is lower for the independent assessment (198 vs. 267).**

  For the peer (independent) review of lesion data, patients with only lesions assessed by physical exam (with or without bone lesions) were NOT sent for peer review. Therefore, the number of patients with best response for investigator assessed and peer (independent) reviewed are different, i.e., independent review is lower than investigator review.

- **Please provide working Excel spreadsheet which describes how investigator-assessed response rates were pooled for use in the sensitivity analysis reported in table 23 (page 87).**

  The use of investigator response rates and the pooling of response data is contain in the Excel model within the Response section of the Default Data sheet – in row 289 (can also be accessed using the default button on the Input Efficacy sheet).

B10. Please provide a more detailed answer to question 114 (page 133). In particular, please provide a copy of time-to-event analyses for overall survival and time-to-disease progression in trials S273 and JHQG.

Transition probabilities for tumour response and AE events were obtained from detailed analysis of the clinical trials available to us, using the tables of data by cycle from JHQG and S273. The distribution of responders and febrile neutropenia events was 'front-loaded' in the first few cycles and the transition probabilities reflect this (e.g. see tables provided in answer

to question B7). Transition probabilities for TTDP and OS were assumed to have a constant risk following the exponential distribution of the 'survival' curves provided in the study report.

B11. Please state clearly and explicitly how the health states in the model (in Excel spreadsheet) were defined. How are S4AE1, S4AE2 and S4AE3 different from SAE4? The same applies to R4AE1, R4AE2, R4AE3 and RAE4? Also, please illustrate how the transition probability, expected utility score and expected cost for each of these states were estimated?

In the model the stable and responsive patients could also be experiencing one of a number of adverse events as represented in the model.

To apply the costs and disutilities of these events the individual AEs were clustered into like groups

These AE groups are defined below

Life Threatening AE (Group AE 1)
Febrile neutropenia - cycle 1
Febrile neutropenia - cycle 2
Febrile neutropenia - cycle 3
Febrile neutropenia - cycle 4+

Hospitalised AE (Group AE2)
Diarrhoea / Vomiting
Stomatitis

Non-Hospitalised AE (Group AE3)
Fatigue including asthenia
Hand-foot syndrome
Neutropenia

Chronic Long Term (Group AE4)
Alopecia (Hair loss)
Neuropathy

The costs and utility estimates for each AE group were based on the specific % and distribution of AEs for each specific 2nd line chemo treatment option – this generated a weighted average cost and utility for AE1-4

This process was repeated separately for Stable and Responsive health states.

Details of this pooling process can be found in the model on the Data Store sheet in rows 100-145.

B12. Please advise the source for the uplift to 2005/06 prices. Costs are reported as being inflated to 2006 prices using the Pay and Prices Index reported by PSSRU, however, there does not seem to be a reference that gives the Pay and Prices Index for the 2005/06 financial year - the 2005 Unit Costs of Health and Social Care (the most recent we can find) gives values for 1995/96 through 2003/04 and an estimated value for 2004/05.

We used an estimated inflation factor for 2005/2006 based on previous years ratios from the PPI (rather than leave the cost data in 2004-5 prices).

Year	Pay and Prices Index (1987/8=100)	% Increase
2003/04	225.6	
2004/05 (E)	234.2	1.038121
2005/06 (Estimated in the Model)	(234.2*1.038121) = 243.1	

Section C: Textual Clarifications

C1. Please confirm whether the figures in Table 17 (page 62) are percentages or absolute numbers.

The figures in Table 17 are percentages.

C2. Please clarify whether the 5.72% figure cited for the Chan et al study in table 32, page 97 is a typing error. Shouldn't a corresponding frequency be given or was the data "Not Registered'

The 5.72% figure is a typing error as it should be blank - what the model does do is assume the same level of neuropathy for GD as seen with GT – which is the 5.72% figure (as no other data were available). Please note that a similar error occurs in Table 30 but this did not make a difference in the model.

We have also noted an error on table 36. Instead of the costs representing a 55% reduction, the costs represent a 45% reduction. A revised copy of table 36 is provided below. The costs are correct in the model so the mistake does not impact the results.

Table 36: Unit Costs of Chemotherapy Treatments

Treatment	Administration	Pack Name	Pack Size	Pack Cost	Off-patent price
Gemcitabine	Injection, powder for reconstitution		1000	£162.76	
Gemcitabine	Injection, powder for reconstitution		200	£32.55	
Paclitaxel	Concentrate for intravenous infusion	6mg/mL	30	£116.05	£52.22
Paclitaxel	Concentrate for intravenous infusion	6mg/mL	100	£347.82	£156.52
Paclitaxel	Concentrate for intravenous infusion	6mg/mL	150	£521.73	£234.78
Paclitaxel	Concentrate for intravenous infusion	6mg/mL	300	£1043.46	£469.56
Docetaxel	Concentrate for intravenous infusion	40mg/mL	20	£162.75	
Docetaxel	Concentrate for intravenous infusion	40mg/mL	80	£534.75	
Capecitabine	Tablets, f/c, peach		150	£0.74 (per tablet)	
Capecitabine	Tablets, f/c, peach		500	£2.46 (per tablet)	

C3 Please provide a full answer for question 92. The answer in the submission refers to question 87 but this does not seem to contain sufficient detail on survival.

Overall survival is defined in 9 of the 14 RCTs (Albain et al., 2004; Chan et al., 1999; 2005; Jones et al., 2005; Nabholtz et al., 1996, 1999; Icli et al., 2005; Winer et al., 2004; Sledge et al., 2003).

Six of these studies adopted the same definitions, with the exception of Icli et al., (2005).

Here is the most commonly used definition:

- Overall survival was calculated from the date of randomisation to the date of death from any cause

Winer et al., (2004) and Sledge et al., (2003) defined survival as:

- 'calculated as the time from study entry to date last known alive or to date of death'.

Icli et al., (2005) on the other hand defined overall survival as:

- The time interval between the first day of treatment and date of death.

C4. Please provide the reference for the study by O'Shaughnessy et al. 2004 (page 116).There is no reference provided in the submission.

This is an error; the reference is O'Shaughnessy et al. 2003, which has been provided.

C5. Please clarify whether the reference to Lloyd 2005 is correct (page 124) or whether it should read 2006. No 2005 paper is given in the reference list for the document.

The reference should read Lloyd 2006. The manuscript has been submitted to the British Journal of Cancer.

C6. Please clarify whether the figure inserted on page 64 has been inserted in error. The figure doesn't seem to reflect the discussion in the text.

The figure has been inserted in error. The correct figure is provided below:

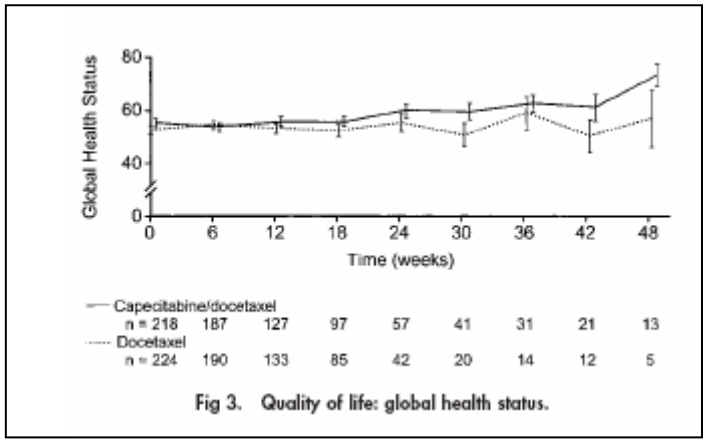


Fig 3. Quality of life: global health status.

C7. Please provide a key to the superscripts which appear in Table 43 (page 133) as no key was provided in the submission.

The key for Table 43 is as follows:

[1] Based on the difference between Overall Survival and Time to Progression
[2] Assumes that the majority of the treatment response is achieved early in the treatment and the remaining response is achieved at a constant rate per cycle over the remaining treatment cycles
[3] Calculated in the model as the patients who remain in the stable state after applying per cycle probabilities of progression, response or death
[4] Based on FN per cycle data from; JHQG trial (for GT), S273 trial (for GD), assumed as T (for D), JHQG trial (for T), S273 trial (for DC)
[5] Based on AE rates taken from the pooled trial data for each treatment option
[6] Transition rates for response, probability and death have been derived by assuming an exponential curve form for the time to event – i.e. an assumed constant risk per cycle for treatment response, progression and death (for response an initial higher response rate was included for the first cycle of response)
[7] Based on the pooled discontinuation rates for AE or patient request from the identified trial

Appendix 1.1. Supporting data for Question A1.

Table A1: Sites of metastases

Study	Brain	Peritoneum	Visceral metastases	Lymph nodes	Lung metastases	Other	Pleura	Bone	Skin	Liver metastases	Non-visceral only metastases	Soft tissue
Albain et al., (2004)			Y		Y	Y				Y	Y	
Chan et al., (2005)					Y					Y	Y	
Bonneterre et al., (2002)					Y	Y	Y	Y	Y	Y		
O'Shaughnessy et al., (2002)				Y	Y			Y	Y	Y		
Sjostrom et al., (1999)			Y					Y		Y		Y
Jones et al., (2005)												
Nabholtz et al., (1999)			Y					Y		Y		Y
Chan et al., (1999)			Y					Y		Y		Y
Extra et al., (2005)												
Mouridsen et al., (2002)			Y					Y				Y
Icli et al., (2005)	Y	Y		Y	Y			Y	Y	Y		
Gradishar et al., (2005)				Y	Y			Y		Y		Y
Winer et al., (2004)			Y					Y				Y
Sledge et al., (2003)			Y									Y

Table A2: Number of metastatic sites

Study	1	2	2-3	3	3-4	>3	4	>5	5-6	7-8
Albain et al (2004)	Y	Y		Y			Y	Y		
Chan et al (2005)	Y	Y				Y				
Bonneterre et al (2002)	Y	Y		Y		Y				
O'Shaughnessy et al., (2002)	Y	Y				Y				
Sjostrom et al., (1999)	Y	Y				Y				
Jones et al., (2005)		Y								
Nabholtz et al., (1999)	Y	Y				Y				
Chan et al., (1999)	Y	Y				Y				
Extra et al., (2005)										
Mouridsen et al., (2002)						Y				
Icli et al., (2005)	Y	Y				Y				
Gradishar et al., (2005)	Y		Y			Y				
Winer et al., (2004)	Y	Y			Y				Y	Y
Sledge et al., (2003)	Y	Y				Y				

Table A3. Performance status

Study	KPS≥90	KPS≥80	KPS, n (%) 70 = 80 = 90 = 100 =	WHO Performance Status 0 = 1 = 2 =	Median KPS, %	Median KPS, (Value given)	ECOG Median	ECOG Status 0 = 1 = 2-3 =	Day 1 PS 0 = 1 = 2 = 3 = Unknown =
Albain et al (2004)	Y	Y							
Chan et al (2005)			Y						
Bonneterre et al (2002)				Y					
O'Shaughnessy et al., (2002)					Y				
Sjostrom et al., (1999)				Y					
Jones et al., (2005)						Y			
Nabholtz et al., (1999)						Y			
Chan et al., (1999)						Y			
Extra et al., (2005)									
Mouridsen et al., (2002)							Y		
Icli et al., (2005)				Y					
Gradishar et al., (2005)								Y	
Winer et al., (2004)									
Sledge et al., (2003)									Y

Appendix 1.2: Academic or commercial information removed

Appendix 1.3 Academic or commercial information removed

Appendix 2 – Additional tables for ERG report

Table 1 – Key characteristics of trials included in data input for economic model.

Trial	Interventions	Reviewed in clinical effectiveness SR	Interventions included in model input data	Full publication/ conference abstract
O'Shaughnessy and colleagues[3]	gemcitabine/paclitaxel vs paclitaxel	Yes	Yes	conference abstract
Albain and colleagues[1]	gemcitabine/paclitaxel vs paclitaxel	Yes	Yes	conference abstract
Jones and colleagues[4]	docetaxel vs paclitaxel	Yes	Yes	Full publication
Sjostrom and colleagues[35]	docetaxel vs methotrexate and 5-fluorouracil	No	docetaxel arm only	Full publication
Nabholtz and colleagues[36]	docetaxel vs mitomycin plus vinblastine	No	docetaxel arm only	Full publication
Bonneterre and colleagues[37]	docetaxel and 5-fluoracil and vonorelbine	No	docetaxel arm only	Full publication
Chan and colleagues[38†]	docetaxel vs doxorubicin	Not referenced	docetaxel arm only	Full publication
Extra and colleagues[9]	docetaxel and docetaxel plus trastuzimab	Not referenced	docetaxel arm only	conference abstract
Mouridsen and colleagues[29]	docetaxel 100mg vs docetaxel 75mg (second line)	Not referenced	Yes	conference abstract
O'Shaughnessy and colleagues[5]	docetaxel plus capacetabine vs docetaxel	No	Yes	Full publication
Icli and colleagues[30]	paclitaxel and cisplatin+oral etoposide	No	paclitaxel arm only	Full publication
Nabholtz and colleagues[39]	paclitaxel (175mg/m^2) vs paclitaxel (135mg/m^2)	No	paclitaxel 175 mg/m^2 only	Full publication
Gradishar and colleagues[40]	Albumin-Bound paclitaxel vs standard paclitaxel	No	standard paclitaxel T only	Full publication
Winer and colleagues[8‡]	paclitaxel 175mg/m^2 vs paclitaxel 210mg/m^2 vs paclitaxel 250mg/m^2	Excluded	paclitaxel 175mg/m^2 arm only	Full publication
Sledge and colleagues[41]	paclitaxel vs doxorubicin vs doxorubicin plus paclitaxel	Not referenced	paclitaxel monotherapy arm only	Full publication
Chan and colleagues[10]	gemcitaine/docetaxel vs capecitabine/ docetaxel	No	Yes	Conference abstract

† no reference in MS – assume [38] is required reference

‡ submitted and referenced paper is of Phase II study of 45 patients receiving GT plus trastuzimab and has no paclitaxel monotherapy arm – assume [41] is required reference

Table 2 –Discrepancies between electronic model and model documentation.

Parameters	Spreadsheet	Cell	Entries in model	Values in report or formula in check
Median overall survival time	'Default Data'	N970 C1006 C1008 C1009	137 262 95 235	203 184 94 223
Median time to disease progression	'Default Data'	C636	28.40	=D649
Overall response rate (Independent assessment)	'Default Data'	C452	107	82
Frequency of febrile neutropenia	'Default Data'	V43 W43	18.2 20.2	7.00 7.00
Frequency alopecia	'Default Data'	U58 U59 U60 U61	6.48 40.64 6.48 40.64	56.00 60.51 56.00 60.51
Frequency diarrhoea and vomiting	'Default Data' 'Data Store'	V26 W26 E408 F408 H408	11.58 11.25 ='Input AE Risk'!$D42. ='Input AE Risk'!$D62 ='Input AE Risk'!$D102	4.00 4.00 ='Input AE Risk'!$D43 ='Input AE Risk'!$D63 ='Input AE Risk'!$D103
Frequency stomatitis or mucositis	'Default Data' 'Data Store'	V27 W27 E409 F409 H409	8.58 8.25 ='Input AE Risk'!$D43 ='Input AE Risk'!$D63 ='Input AE Risk'!$D104	2.00 2.00 ='Input AE Risk'!$D44 ='Input AE Risk'!$D64 ='Input AE Risk'!$D103
Frequency neutropenia	'Default Data'	V31 W31	93.62 91.88	84.00 84.00
Frequency sensory/ peripheral neuropathy	'Default Data'	V34 W34	5.17 5.20	2.00 2.00
Febrile neutropenia mortality	'Default Data'	C274	=IF(psa_switch=1,F274,E274)]	[=IF(psa_switch=1,E274,D274)