Descriptive and Quality Analysis of Markov type Decision Analytic Modeling in Economic Evaluations of Metastatic Breast Cancer Treatment Therapies

by

Nikoloz Tsuleiskiri

Student Number: 345823

Supervisor: Jennifer Gaultney, MPH, MSc

Co-readers: Dr. Rian Rijnsburger, PhD and Annemieke Leunis, MSc

Master program: Health Economics Policy and Law (HEPL)

Table of Contents

Abstract

Background: Breast cancer is the most common type of cancer in women. The assessment of characteristics and quality of decision analytic models of economic evaluations (EE) in metastatic breast cancer treatments has primary importance because EEs provide important evidence for decision-making concerning health care. The decision analytic modeling (DAM) process is central to performing economic evaluations in chronic diseases like breast cancer, but may produce low quality evidence depending on the methods applied.

Objective: The objective was to summarize the cost-effectiveness results, methodology and quality of DAMs in MBC with the aim to provide guidance on the cost-effectiveness of current therapy option and to appraise and illuminate the flaws in the quality of DAM so as to provide future recommendations for generating robust evidence on cost-effectiveness of MBC therapy.

Methods: A literature review was conducted to identify the EEs of MBC treatment written in English language between 2002 and 2012. Methodological characteristics were observed by extracting data regarding methodology of DAM. A quality appraisal checklist designed specifically for DAMs was applied to appraise the quality of all EEs in MBC treatments utilizing a Markov models.

Results: In total thirteen EEs were identified which represented DAM-based cost effectiveness and cost utility analysis, and outcome measures were quality adjusted life years or life years. Studies were conducted in various countries with the health care perspective being prevalent. In most studies lifetime horizon was used, though this differed according to target patient group and was found to be influential on the results for similar drugs. Discounting was applied according to guidelines and decision maker requirements. Comparators in these studies were included hormonal therapies, targeted therapy and chemotherapy. In general, combination therapy including 2 or more types of therapy was more cost-effective than combination therapy. The overall quality score of DAM was 70%. The best performance of all reviewed studies was in structure 80.93% and the data dimension 67%, while the consistency dimension scored the lowest at 35.38%. Model type (S6), cycle length (S9), and parameter uncertainty (D4d) were scored high in all EEs. The sub-dimensions which scored lowest were comparators and strategies (S5), rationale of structure (S3), data identification, (D1) Pre-model data analysis (D2), baseline data (D2a), and treatment effects (D2b) and quality of life weights (D2d).

Conclusions: Development of a generic model for MBC is recommended to improve quality of DAM and reduce the variations in final results. A systematic review of the efficacy evidence of all therapies in MBC is needed to identify the heterogeneity which can be integrated into the model by means of a meta-analysis. Such a model could then be used to economically evaluate MBC treatment therapies and adapted depending on patient subgroup.

1 Introduction

Breast cancer is the most common type of cancer in women. The expansion of the heath care sector has particularly affected by oncology care. Cancer is a disease whose burden is directly proportional to the world's population. Specifically, breast cancer is responsible for 23% of new cancer cases and 14 % of total cancer deaths in 2008 [1, 2]. Breast cancer is defined as metastatic if characterized by metastatic (M) stage or stage IV, according to TNM cancer staging system, and breast cancer stage grouping. In other words, metastatic breast cancer (MBC) is cancer originating from breast but which has spread to distant sites. In this setting [3], the objective of treatment is limited to life prolongation and symptom management [4]. Due to the numerous options for MBC treatment [5], it is important to economically evaluate the value of treatment therapies when decide the reimbursement decision [6].

Despite private or social insurance, within the jurisdictions where health care is reimbursed centrally, the importance of Economic Evaluation (EE) has significantly increased [7]. This is explained by need to ensure affordable and equitable access to effective medicines in a sustainable manner. This objective is difficult to obtain, taking in account the expansion of the health care sector and scarcity of recourses in health care. The increasing number of interventions, even within one disease, results in variability of economic return [8]. Thus, it becomes inevitable to decide which intervention is giving the best value for money [9].

Since the late 80's, health technology assessment (HTA) has been used to appraise new technologies in health care. Both a cost-effectiveness analysis (CEA) and a budget impact analysis (BIA) typically generate economic evidence about new health technologies [10]. The main objective is to address issue of efficient allocation of resources along with maintenance of affordability of services [11]. Therefore, prospecting possibility of solving two problems: increasing health care expenditures and scares resources demanded by policy makers [12], brings HTA to the center of researchers' attention.

The assessment of health care interventions by means of an EE is conducted by synthesizing all available information about a particular healthcare process and its implications. Philips et al. [13] have stated that the decision analytic modeling (DAM) process is central to performing a CEA. Especially when evidence about long-term outcomes is not available, the objective, of decision modeling is to provide clear relationship of incremental costs and consequences between two interventions [14]. Sun et al. [15], have also stated that modelling is a complex technique which demands great experience of modellers in order to achieve sufficient quality. Consequently, the quality of decision modelling has a significant influence in decision making between interventions.

Despite complexity and potential susceptibility to limitations, DAM is a powerful tool to synthesize available data and information. The results of the DAM are always conditional on structural and data assumptions. The greatest advantage of modelling studies is that it gives

possibility to work with imperfect data. But on other hand it requires to make series of assumptions. Often modelling studies are used to evaluate newly developed technologies or treatments, which are not widely used in practice. For this reason it is important to address the uncertainty related with assumptions made [16].

The problems of DAM studies can be grouped into two general categories. The problems associated with structure and problems associated with data. Problems associated with structure are derived from choice of correct rationale of model structure, its assumptions, comparators, choice of model type, prolongation of time horizon, choice of disease states and cycle length [17]. These issues, on other hand, are strongly influenced by developers of the DAM, whereas different analysts are not expected to have same vision. Problems associated with the data are two sided. On one hand, they depend on the quality of data identified and analyzed, on other hand they depend on analysts who choose which data will be incorporated into the model. Finally, despite the great influence of problems associated with data collection, analysis and structural assumptions on DAM, additionally quality of DAM is influenced by medical proficiency of analysts.

The Markov type of model, in chronic diseases like breast cancer, is the preferred type of model [18] to represent stochastic processes [19] as the decision tree type model does not define an explicit time variable which is necessary when modelling long term prognosis [9]. However, the results of Markov type decision modelling are influenced by the model structure [9], validity of data [20], its synthesis [21], and resulting uncertainty [22] therefore this susceptibility can be considered as a shortcoming of the model.

Susceptibility of DAM threatens the robustness of generated evidence, and on other hand provokes the following questions to be answered: what are the methodological characteristics of EEs for MBC? And what is the level of quality of DAM in EEs for MBC? In order to provide the answers to these questions, a review of the methodological characteristics and quality appraisal are thought to be vital.

The identification of characteristics and quality appraisal of DAM in HTA/EE are important in an environment where evidence supports decisions [23] on reimbursement, developing new drugs, therapies, disease management programs and value-added services and clinical practice guidelines [24]. Therefore, the link between the primary data and decision [14] represents a valuable tool for decision makers, third party payers, manufacturers, treatment centers, clinicians, and researchers [25].

Numerous assessment tools have been devised to critically appraise the quality of decision models. In 2009, Zimovetz et al. [26] in review have stated that none of the DAM assessment tools accurately reflect the quality criteria specified by the NICE, and proposed the checklist which adopts similar basic domains as one by Philips et al. [13]. However, in the same review the checklist proposed by Philips et al. [13], was merited as most comprehensive tool for quality

appraisal. This checklist has been developed by consolidating existing guidelines and appeared to become consistent framework for quality assessment. The checklist provides a broad framework for quality assessment and best practice in decision-analytic modeling for cost-effectiveness analysis. Using this checklist specifically the quality in “structure”, “data” and “consistency” of DAM can be evaluated.

By collecting and summarizing the information about the cost-effectiveness results, methodology and quality of DAM in MBC the aim of this research are two folded. First, a descriptive analysis of the methodological attributes and results of existing EEs in MB was performed to provide guidance on the cost-effectiveness of current therapy option. Second, an evaluative analysis focused on determining the quality of DAM in MBC EEs, in order to appraise and illuminate the flaws in quality of DAM and provide future recommendations for generating robust evidence on cost-effectiveness of MBC therapy in sustainable and effective way.

A number of economic evaluations utilizing a Markov model have been performed in MBC, but little is known about the characteristics and quality among existing studies. Therefore, the objective of this research was to provide answers to above stated questions, by descriptive analysis of the methodological characteristics and assessment of quality in economic evaluations of MBC.

2 Methods

2.1 Search strategy

To identify studies utilizing decision analytic model (DAM) in MBC, a literature search of all EEs of all therapies for MBC was conducted. Initially search terms were developed: “late stage”, “metastatic”, “advanced”, “breast cancer”, economic evaluation”, “cost utility”, “cost effectiveness”, and “cost benefit”. On May 12 (2012), a literature search was conducted in National Health Service Economic evaluation Database via Cochrane library and center for reviews and dissemination (CRD), Pubmed, and Embase data bases. Search terms were categorized into nine separate searches. Table 1 illustrates how search terms were applied in search process.

2.2 Study selection

By taking in account the objectives of this research the inclusion criteria were developed in order to enable the selection of studies for the research. All studies were first reviewed according to titles and abstracts, and inclusion criteria were applied to select studies relevant to this research. Inclusion criteria consisted of the following: 1) study focused on treatment of metastatic or advanced breast cancer, 2) focus of study CEA or cost-utility analysis (CUA) 3) modeling study based on Markov type of decision analytic model, 4) full text economic evaluation, 5) written in English language, 6) published in the period 2002-2012.

2.3 Descriptive analysis

Data regarding type of study, effectiveness outcomes, perspective, indirect costs, target population, time horizon, discounting rate, comparators and incremental cost effectiveness ratio (ICER), were extracted for the descriptive analysis of methodological characteristics of the selected EEs.

2.4 Quality of decision-analytic modeling

The study by Zimovetz et al. [26] suggests that the quality assessment checklist by Philips et al.[13] is the most comprehensive and specifically aims to appraise DAM studies. Hence it follow that, the decision to apply the checklist by Philips et al. [13] was made. This checklist incorporates evidence from a systematic review of best practice guidelines and focuses on three dimensions of quality: structure (S), data (D) and consistency (C). It was also found that operationalized version of the checklist by Philips et al. [27] includes the additional attribute "cost" (D2c), whereas Philips et al. [13] did not include this attribute. For the purpose of better comprehension in this research it was decided to include additional D2c attribute to checklist by Philips et al. [13]. Methods of DAM applied in each EE were appraised by analyzing each quality dimensions using this operationalized checklist [13].

In most cases, each dimension of quality contains several attributes of good practice. Hence it follows that, each attribute of the checklist was numbered by one alphabetic letter. There are 63 attributes of good practice, in total.

The previously operationalized checklist was applied using "yes" or "no" answers. In this research it was assumed that each attribute had the same quality value resulting in a total of 63 points that could be awarded for quality. Therefore, as attribute-associated criterion with the answer "yes" was awarded one point whereas that with a "no" was awarded no point. In cases where the selected study did not provide explicit and clear information about an attribute, the criterion was scored with "no". In the case of a clear and explicit statement of attributes, the criteria were scored "yes". The dimensions of quality, attributes of good practice and questions applied in the critical appraisal of good practice guidelines for DAM as taken from Philips et al. [13], are provided in Table 2.

3 Results

3.1 Search and study selection

In total, the search resulted in identification of 1703 studies in four online databases. Figure 1 represents the selection process of the economic evaluations for this research. A total of 937 duplicates were removed. All 766 unique studies were reviewed by title and abstract, where 697 were excluded as they did not focus on treatment of MBC (n=575), and were not based on Markov type of decision modeling (n=122). Furthermore, studies which were not full text economic evaluations (n=13), were not published in period of 2002-2012 (n=13), and focused on early breast cancer (n=15)[1], and those which were systematic reviews (n=3) were excluded from pre-final set of studies. The pre-final set of (n=26) studies were reviewed by its content and 13 studies were excluded. The reasons of exclusion were; not CEA or CUA (n=1) [29] not focused on MBC (n=7) [30-36], not written in English language (n=1) [37], systematic literature review (n=1) [38], not full text EEs (n=2) [39, 40] and duplicate (n=1) [41]. After exclusion of duplicates and applying six exclusion criteria, final set of studies for this research comprised of 13 full text economic evaluations.

3.2 Description of selected economic evaluations

Implemented data extraction of cost-effectiveness results and methodological characteristics in selected EEs for MBC, is represented in Table 3.

3.2.1 Type of study and effectiveness outcomes

All selected studies used quality adjusted life years (QALY) or life years gained (LYG). Only 53.84 % of evaluations used QALY as well as LYG as a measure of health outcomes [28, 42-47]. In 46.15 % studies [48-53], only QALY was used. Moreover, in two EE quality adjusted life month (QALM) was used [48, 52].

By analyzing outcome measures, in 13 EEs, it was found that 53% of EEs are CEA and CUA type of studies [28, 42-47]. The remaining 46% are strictly CUAs [48-53].

3.2.2 Perspective

Perspective was explicitly mentioned in 92.30% of reviewed (12) EEs. Only in one EE perspective was not stated [48]. Health care or payer's perspective [28, 42-47, 49, 51, 53] were stated in 76.92%. Other 15.38% assumed societal perspective [50, 52]. EEs stated societal perspective, were conducted for the Swedish setting and for the United States of America (USA).

[1] The study by Karnon et. al. (2008) was considered to be sufficient for inclusion criteria, therefore was included in final set of studies for this review.

28. Karnon, J., T. Delea, and V. Barghout, *Cost utility analysis of early adjuvant letrozole or anastrozole versus tamoxifen in postmenopausal women with early invasive breast cancer: the UK perspective.* Eur J Health Econ, 2008. **9**(2): p. 171-83.

3.2.3 Indirect costs

EEs, which included indirect costs, comprised 15.38% of total share, which all stated the societal perspective. The study conducted in Sweden, included average annual inpatient costs and average annual informal care costs [50]. Studies conducted in USA [52], patient time costs and patients travel costs were included. The remaining 84.61 % studies did not include indirect costs, since the perspective was limited to the health care or payer's perspective [28, 42-49, 51, 53].

3.2.4 Target population

The target population, in all EEs, was patients with MBC. In 61.5% [42-46, 49, 51, 53] EEs the age of the target population was not stated. In studies where it was mentioned, age varied from 50 to 65 years. In the study by Marchetti et al. [49], a lifetime time horizon of 8.3 years was used for post-menopausal women with estrogen positive MBC. Whereas, Karnon et al. [28] used a 5-year time horizon for women of 61 years old with early invasive breast cancer.

3.2.5 Time horizon

Time horizon was not stated (7.69%) explicitly by Matter-Walstra et al. [53]. In the remaining 92.30 % of economic evaluations the time horizon varied from 5 to 28 years, and was considered to be lifetime.

3.2.6 Discounting rate

Discounting rate was not applied in 23.08 % of studies [50, 51, 53]. Throughout the remainder of studies the discounting rate ranged from 3% to 6%. Only in one study were resources and effects discounted with different rates, with resources at 6% and effects at 1.5% [42]. In six EEs [45-49, 52], a 3% discounting rate was used and three studies applied a 3.5% discounting rate [28, 43, 44].

3.2.7 Comparators and incremental cost effectiveness ratio

Throughout the reviewed studies, five EEs have focused on hormone therapy (HT). One EE specifically demonstrated comparison of aromatase inhibitors (AI) to selective estrogen receptor modulators (SERMs) [42]. Two EEs compared AI agents [28, 49]. Remaining two EEs compared estrogen receptor down regulator (ERD) agents [43, 45].

Only one EE focused on comparison of targeted therapy (TT) and usual care [47]. Elkin et al. [48] compared the Hercept test and fluorescence in situ hybridization (FISH) to FISH only, followed by TT and chemotherapy (CT).

CT and TT regimens were evaluated in six studies, four EEs focusing on combination of CT and TT agents [50-53], while two EEs compared only CT agents [44, 46].

3.2.7.1 Hormonal therapy

3.2.7.1.1 Aromatase inhibitors versus selective estrogen receptor modulators

EEs by Karnon et al. [42], aimed to compare AI to SERMs. This EE focused on comparison of letrozole (femara) to tamoxifen (nolvadex) as a 1st line therapy followed by 2nd, 3rd line therapy,

1st and 2nd chemotherapy. In this EE, letrozole and tamoxifen were the first drugs in treatment sequences. This EE demonstrated an ICER of 6068£ per LYG and that treatment sequence starting with letrozole was cost effective compared to one with tamoxifen at a cost of 2927-3969£ per QALY.

3.2.7.1.2 Aromatase inhibitors

A crossover design comparison by Marchetti et al. [49] compared two AI drugs, letrozole and anastrozole (arimidex), to tamoxifen (SERM). The first comparison represented two hormone therapy sequences[2], (1) starting with anastrazol followed by tamoxifen then megestrol (megace) compared to therapy sequence, (3) starting with tamoxifene followed by anastrazole, then megestrol. This comparison resulted in ICER €10795/QALY. The second therapy sequences (2) starting with letrozole followed by tamoxifen, then megestrol compared to therapy sequence (3) starting with tamoxifene followed by anastrazole, then megestrol resulted in ICER €16886/QALY.

Comparators of the study by Marchetti et al [49] and Karnon et al. [28] are same. The comparison [28] of anastrazole and lestrozole were accomplished by comparing these drugs to tamoxifen. The study revealed for letrozole compared to tamoxifene, the incremental cost of £10,502/LY and incremental cost per QALY of £10,379. Results also showed incremental cost of 11,703£ per LY and incremental cost per QALY of 11,428£ respectively for anastrazole compared to tamoxifen. The reference case analysis assumed no carry over efficacy for AIs beyond the 5-year treatment period, and showed similar levels of CE for both letrozole and anastrazole compared to tamoxifen of between £10,502 and £11,703. In this EE crossover design was used however, comparison was not implemented by comparing treatment sequences as it was done in previous two studies. The DAM used in this study was developed by Karnon et al. [54] in EE of tamoxifen plus chemotherapy and tamoxifen alone in postmenopausal women with early breast cancer.

3.2.7.1.3 Estrogen receptor down regulator

The EE of the HT estrogen receptor down regulator (ERD) drug by Cameron et al. [43] was carried out by comparing two treatment sequences in two cohort groups. Firstly, fulvestrant (faslodex) used as a second-line HT for MBC in Cohort A compared to cohort B without fulvestrant. When fulvestrant was used as a second-line HT, it resulted in £6500 per LYG, and £7300 per QALY. The second comparison compared fulvestrant, as a third-line HT in cohort A to cohort B without fulvestrant, which showed both an increase in health benefits and cost saving of £430 per patient.

Lux et al. [45] compared fulvestrant, as a second line treatment sequence in cohort A to no fulvestrant in cohort B. This revealed fulvestrant as second line treatment to result in a 0.021

[2] The numbers in parenthesis refer to treatment sequences compared by Marchetti et. al. (2004).

49. Marchetti, M., M. Caruggi, and G. Colombo, *Cost utility and budget impact of third-generation aromatase inhibitors for advanced breast cancer: a literature-based model analysis of costs in the Italian National Health Service.* Clin Ther, 2004. **26**(9): p. 1546-61.

QALY gain at a cost saving of 564 euro per patient, dominating the treatment sequence without fulvestrant.

A number of similarities such as the type of model were observed among these two studies. In the EE by Lux et al. [45], patients were assumed to have previously received tamoxifen or AI as adjuvant anti-hormonal therapies, whereas the EE by Cameron et al. [43] assumed patients had previously received only adjuvant tamoxifen. Although, both authors found that inclusion of fulvestrant as second line agent in treatment sequence is cost effective.

3.2.7.2 Targeted Therapy

Hedden et al. [47] demonstrated the comparison of 12-month adjuvant protocol of trastuzumab (Herceptin) to usual care. The findings show that a 12-month adjuvant protocol of trastuzumab resulted in a gain of 1.38 QALY or 1.17LYG at a cost of $18,133 per patient. Thus, the ICER was estimated to be $13,095/QALY and 15,492/LY.

3.2.7.2.1 Targeted therapy and Chemotherapy

Elkin et al. [48] compared the Hercept Test and FISH to FISH only, followed by trastuzumab and chemotherapy agent, paclitaxel (taxol). Findings yielded an ICER of $125,000/QALY when HercepTest with FISH, and ICER of $145,000/QALY when FISH alone.

Lidgren et al. [50] compared HER2 testing and trastuzumab (TT) in combination[3] with CT agent docetaxel (taxotere) to CT alone, which revealed that the least costly and least effective strategy was CT alone. However, the most effective strategy (5) was FISH testing for all patients, with trastuzumab and CT for FISH positive patients. When compared with strategy (4)[4] it resulted in ICER of 561,207 SEK/QALY. This is only study which included HER2 testing and trastuzumab together in comparison of TT and CT.

Dedes et al. [51] have found that paclitaxel (CT) and bevacizumab (avastin) (TT) together led to a gain of 0.21 QALYs per patient at an additional cost of €40,369 versus paclitaxel (taxol) alone at an ICER equal to €189.427/QALY.

Le et al. [52] compared Capecitabine (CT) mono-therapy to capecitabine (xeloda) and lapatinib (tykerb) (TT) combination, and found that combination therapy cost an additional $19,630 with an expected gain of 0.12 QALY at an ICER equal to $166,113.

The final EE which compared TT to CT was conducted by Matter-Walstra et al. [53]. trastuzumab (TT) and capecitabine (CT) versus capecitabine alone were compared. It was

[3] The comparison is initiated by comparing the treatment strategies, the numbers in parenthesis indicate the treatment strategy, these numbers are provided in original study (EE).

[4] Strategy 4: IHC testing for all patients with FISH confirmation of 2+ and 3+ and Trastuzumab and chemotherapy for FISH Positive patients (P.1022). 55. Lidgren, M., et al., *Cost-effectiveness of HER2 testing and 1-year adjuvant trastuzumab therapy for early breast cancer.* Annals of Oncology, 2008. **19**(3): p. 487-495.

demonstrated that trastuzumab and capecitabine combination led to additional cost of €33,980 and yielded a gain of 0.35 QALYs at an ICER of €98,329/QALY.

3.2.7.3 Chemotherapy

Comparison of chemotherapy regimens was done by Benedict et al. [44]. A direct comparison was conducted for 3-weekly regimen of docetaxel and paclitaxel based on patient data from the TAX311 trial. The resulting ICERs for docetaxel were £12 032/QALY versus Pac3w, £4583/QALY versus Pac1w and £14 694 per QALY versus Nab-P.

Frias et al. [46] assessed the cost-effectiveness of docetaxel administered every 3 weeks and weekly paclitaxel regimen. The comparison was indirect and based on efficacy and toxicity data from the literature and individual patient level. The results demonstrated docetaxel to yield higher benefits as well as costs. Therefore the ICER for docetaxel versus paclitaxel was €190/LYG and €295/QALY.

3.3 Quality assessment

The quality appraisal checklist was applied to analyze decision analytic modeling of selected economic evaluations. Table 4 presents the quality appraisal of all reviewed studies. Table 5 summarizes the quality of DAM in appraised EEs.

3.3.1 Structure

The dimension of decision problem and objective statement is comprised of three attributes. In 76.92%, the decision problem was stated, moreover objective and model of evaluation were consistent with aforementioned decision problem. In 23.07%, the primary decision maker was not declared. With regards to overall quality of the S1 dimension, 23.07% were insufficient to score all attributes.

In EE by Elkin et al. [48] perspective of the model was not stated, besides consistency of model inputs and outcomes with model perspective was not tested. Model outcomes and overall objective of evaluation were tested. In this quality dimension, there were two EEs [43, 44] where consistency of model outcomes with perspective, scope or overall objective were not tested. Finally, it was concluded that 23.07% (n=3) did not fully score "statement of perspective and scope" dimension (S2).

The evidence regarding model structure as well as casual relationships were described and justified in all EEs under evaluation. The sources of data, used to develop model structure were also provided. The structure of the model was not consistent with coherent theory of the health condition under evaluation in 30.76% (n=4). Consideration of competing theories of model structure was carried out in none of reviewed EEs. Finally, the quality of "rationale for the structure" dimension was not scored by any of the EEs (S3).

All EEs assured the transparency of structural assumptions. Moreover, assumptions were reasonable given overall objective, perspective and scope of model in all EEs except one (7.69%). Only one EE failed to test consistency of structural assumptions with perspective of model [48]. The quality dimension "structural assumptions" was scored in 92.30% of EEs (S4).

Comparators were appropriately defined in all EEs. However, 92.30% of EEs failed to evaluate and justify exclusion of practical options. This resulted in a failure to score "strategies/comparators" quality dimension by 12 EEs. Only one study scored all attributes of S5 quality dimension was carried out by Elkin et al. [48].

The type of the decision analytic model in all economic evaluations was Markov type decision model. The choice of this specific type of model was justified (S6) in all studies.

In 92.30% of reviewed EEs applied lifetime time horizon was sufficient to assure reflection of all differences between comparators, besides time horizon, duration of treatment and treatment effects were described and justified. Only one EE failed to score in "time horizon" dimension (S7), because none of attributes were declared.

In 69.23%, the Markov state transition model included disease states that imitated underlying biology of disease rather than current patterns of service provision. Finally, four EEs did not score in "disease states" quality dimension (S8).

Cycle length was explicitly defined in all economic evaluations. However, the justification for cycle length was not provided in 61.53% of EEs. Additionally, it was found that 15.38% EEs gave justification for cycle lengths that represented time between doctor visits, patient re-examinations, time between drug administrations and natural history of disease (7.69%). Finally, this quality dimension (S9) for "cycle length" was found to be sufficient in 100% EEs.

3.3.2 Data

The methods of data identification for the model were transparent and coincided with the stated objective in all EEs. All EEs justified the choice made about data source used in the model (more than one data source were available) and paid particular attention to selecting crucial parameters. The process of parameter selection was also explained for all EEs by providing systematic methods for identification of crucial data. The quality of data was not analyzed in 76.92% of EEs. Expert opinion was used in 38.46% of EEs and methods were explained and justified. In total, 15.38% of EEs have scored in all attributes of "data identification" quality dimension (D1).

The methods of pre-model data analysis was explained and justified by statistical and epidemiological methods by all reviewed EEs. This dimension (D2) of quality involves four sub dimensions. However, none of the EEs have scored in this dimension because sub-dimensions were not fully scored.

The choice of baseline data as well as its description and appropriate calculation of transition probabilities were positively scored by all reviewed EEs. The application of half cycle correction on all transitions in the model was found only in two (15.38%) EEs [28, 42]. For these EEs, there was no need for justification of half cycle omission. In 84.61% EEs, half cycle correction was not applied and omission was not explained. Finally, "baseline data" dimension (D2a) was sufficiently scored by only two EEs.

In all reviewed EEs, the synthesis of relative treatment effects was performed by appropriate techniques. Justification of short term to final results extrapolation has been documented. However, the assumptions made in the extrapolation process were not tested using sensitivity analysis by one EE [45]. Assumptions regarding continuing effects of treatment have been documented in 61.53% of EEs. Testing of these assumptions through sensitivity analysis was not conducted by 25% out of the 61.53% (2 out of 8) of EEs. Finally, the D2b dimension "treatment effects" was fully scored by six (46.15%) of reviewed EEs.

Among all EEs, a description and justification of all included costs were documented. The sources of the cost data were provided. Discounting rates were explicitly stated and justified according to the decision maker in 76.92% of EEs. Finally, the D2c dimension "costs" was sufficient in 76.92% of EEs.

In all EEs, the incorporation of utilities was performed appropriately and the sources of utility weights were provided. However, in seven EEs (53.84%), utility derivation methods such as direct or indirect methods, were not provided or justified. Finally, the D2d, "quality of life weights" dimension was fully scored by 46.15% of EEs.

The dimension of "data incorporation" was fully scored by all but one EE (D3). In the EE conducted by Cameron et al. [43], data was not incorporated by distributions and therefore the distribution for each parameter was not described. Moreover the second order uncertainty was not reflected.

In all reviewed EEs, at least one principal type of uncertainty was addressed. The justification for omission of particular forms of uncertainty was not declared in any reviewed EEs. The overall dimension of "assessment of uncertainty" (D4) was sufficiently scored by 15.38% of EEs.

Methodological uncertainty (D4a) was addressed in 23.07% of EEs. The structural uncertainty (D4b) was addressed in 53.84% economic evaluations. The heterogeneity (D4c) of different patient subgroups was not addressed in 46.15% of studies. In all EEs the methods of addressing the parameter uncertainty (D4d) were transparent and appropriate. Also the probabilistic sensitivity analysis was performed. However, it is important to mention that in only one EE data was incorporated with point estimates.

3.3.3 Consistency

Only in one (7.69%) EE [44] was it stated that mathematical logic was tested. However, the process was not provided in the text. Dimension (C1) of "internal consistency" was not fully scored in 92.30% of EEs.

With regards to the "external consistency" dimension (C2), contra intuitive results were summarized and described only in 15.38% of studies [44, 46]. The model was validated through calibration against independent data in 30.76% of EEs. Comparison of models with previously existing models was positively assessed by 84.61% of economic evaluations. None of the EEs have fully scored this quality dimension.

4 Discussion

4.1 Summary of findings for cost-effectiveness and methodology in Markov type decision analytic model economic evaluations

To summarize the descriptive analysis in this review, in this section a discussion of the results and methodological characteristics is provided. Comparators in reviewed EEs comprised of three types of breast cancer treatment therapies. Reviewed EEs presented comparisons within and across MBC therapy types. In most cases the comparison was carried out by comparing treatment sequences and combinations. Also it was found that crossover design or indirect comparison methods were used. Within the drug class of HT, it can be concluded that AIs are more cost-effective then SERMs. ERD therapy (fulvestrant) is more cost-effective as second line agent in treatment sequence.

A TT drug, in most cases, was compared with CT agents. The findings of these EEs showed that combination of TT and CT are more cost effective treatment regimen than mono-therapy of TT and CT.

Comparison of two CT showed different estimates of ICER results. This can be explained by the type of (indirect/direct) comparison, different source of data and health care settings (UK/Spain). Additionally, the difference of ICERs were caused by incorporating different treatment regimens, 3 week regimen for both CT drugs, while another EE incorporated 1 week and 3 week regimens for both CT drugs. It is crucial to take in account the risks of conducting indirect comparisons where target populations are different across trials [56] and question still remains if risks of bias in results, when conducting indirect comparisons are foreseen and if measures are taken to reduce this risk in future.

EEs were conducted from different contexts, although discrepancies in results were minor. These minor differences in results suggest that the finding may be consistent across different health care systems. However, reason for discrepancies includes the sensitivity of model to utility parameters. Another example of sensitivity to utility parameters was found in EEs comparing CT and TT.

Among the reviewed studies, the greater portion assumed the 3rd-party payers and health care system perspective. However, the societal perspective is desired when conducting economic evaluations [57, 58]. It is the decision of the modeler which perspective to take in EE taking in account requirement of primary decision maker. Furthermore, the perspective of EE, influences cost effectiveness as concluded by previous researches [59, 60].

Despite limiting the target populations of all reviewed EEs to patients with MBC, it is important to take in account biological subtypes of breast cancer, and age. As a consequence the start of modeling and duration of time horizon differed across the EEs. The choice of time horizon and its impact on ICER warrant careful consideration [61]. It is important to assure that duration will

enable reflection of all important differences between comparators [13]. The NICE guidelines suggest that lifetime horizon is appropriate to most of the models [27, 62]. The discounting rates in reviewed EEs were based on decision makers' requirements. But besides the rates, it is important to state at which year of time horizon (time frame), costs and benefits are discounted. Whereas, at which year discounting rate is applied it have effect on the present value of costs or benefits. Despite the importance of this issue, in this research, this type of approach, towards discounting rate and time horizon, from modelers was not identified.

The comparison of treatment sequences, use of cross over design of studies and the use of indirect comparisons in DAMs of selected the EEs were identified as trends. The combinations of MBC treatments are more cost-effective than single therapies (mono-therapy). Also difference in health care system settings influence cost-effectiveness, specifically this happens when different utility parameters or discounting rates are used in DAM. Time horizon, which was most often lifetime, was observed in all EEs. However, the age of target populations was different. This indicated that duration of time horizon was different, and this has substantial influence on cost-effectiveness of the same therapies assessed in different patient groups. For example, the time horizon influences cost-effectiveness through the discounting rate, because the longer the time horizon is more discounting rate influences the final results as well as differences in disease severity. The fact that cost-effectiveness estimates were highly sensitive to the estimated duration of treatment benefit as found in this study also was observed by Hall et al. [63].

4.2 Summary of quality and its implications

To summarize the evaluative assessment in this review, in this section a discussion of the quality assessment findings are provided. When quality was appraised for each attribute, within all three dimensions the average quality was 70%. Structure dimension appeared to have highest average quality of 80.93%. In data dimension average quality was 67%, and in consistency dimension estimated quality was 35.38%.

In the "structure" dimension, the decision problem, objective, scope and perspective were stated in 76.92% of EEs. It is crucial to state decision problem and objective of DAM [15, 57-60]. The consequences may lead to bias for critical readers, especially if EE took societal perspective. It may be difficult to determine whether primary decision maker is patient, hospital or other entities.

Defining the perspective of the model is important [57, 58]. Omitting the statement of perspective does not permit testing consistency of outcomes and perspective. Testing of structural assumption on consistency with perspective scope and objective is not possible either. In order to achieve good quality of decision modeling [64], assumptions regarding the model structures should be stated.

Rationale for structure was not completely fulfilled by any EEs. Structural assumptions were stated by all EEs (100%) and the inclusion of disease states that reflected underlying biological process was seen in 69.23% of EEs.

The rationale of the model is suggested to resemble theory of disease, and health states have to reflect underlying biological processes [64-66]. Consideration of competing structure of the model does not affect the quality of DAM as such, if DAM is assumed to be perfect. In real world, this is not a case.

The comparators were evaluated by only one EE (7.69%). The strategies under evaluation and assessment of feasible options are important [64, 66, 67]. If above mentioned will not be fulfilled, it may result in omission of relevant comparator, and will reflect on consistency of decision, made by decision makers'.

Model type was stated explicitly in 100% of EEs. The choice of type of the model is important step in structuring the decision model. Disadvantages of decision tree type model drive choice of Markov type of model [9, 22]. Markov type DAM in selected EEs was dictated by decision problem and choices made regarding the casual relationships within the model [13].

Time horizon was stated in 92.30% of EEs, and cycle length was defined in all EEs. The cycle length should be defined according the natural history of disease [57, 60]. However, according to the literature cycle length should have same interval as pathology of symptom alters.

Data identification is one of the central steps of decision modeling. Data used in decision modeling should be derived from well-designed or high quality studies [13, 64]. In reality, evidence is scarce. Thus the methods of obtaining data, its quality, and sources should be reflected in all studies to guarantee the sufficient data for the model.

The analysis of baseline probability, treatment effect, utility score and cost, as well as its transformation to form which can be incorporated into DAM are steps in pre-model data analysis [68]. Disclosure of the methods is important, whereas it ensures clarity of decision makers. The incorporation of this data into model is delicate step and quality of this step influence final results of DAM [14, 64, 65, 69]. All types of uncertainty, related to the data, structure methods and parameters, should be addressed [14], in order to ensure the legitimacy of decisions made about reimbursement.

Model validations through testing mathematical logic [68] is also an important step, which involves using extreme values for input parameters to examine the results [67]. External consistency by comparison models with existing models and calibration of models, are important attributes of good practice [64-67, 70, 71]. Despite its importance this dimension scored the least quality.

For the concluding remarks of this section, it is important to emphasis the sub-dimensions for which quality was observed to be less than 50%. In the structure dimension, the evaluation of comparators and strategies (S5) was scored by only one EE and rationale of structure (S3) was scored by none. With regards to data dimension, data identification (D1) was completely scored by only two EEs. Pre-model data analysis (D2) was not scored by any EEs; however this sub-dimension consisted of four general attributes out of which baseline data (D2a), treatment effects (D2b) and quality of life weights (D2d) were scored by less than 50% of EEs. The assessment of uncertainty was completely scored by only two EEs, and it is important to mention that methodological, structural and heterogeneity uncertainty were addressed by less the 50% EEs. Concerning the dimension with the least quality, consistency, internal consistency was scored by only study and external consistency was not completely scored by any EEs. These weaknesses in quality highlight areas for needs for improvement in future EEs using DAM in MBC.

4.3 Summary of weaknesses of the methods

One weakness of this study was that only full text EEs published in English language, in 2002-2012 was included. This could have caused omission of earlier EEs in MBC treatment. Another weakness of this study is that it was conducted by one person, whereas better approach is to perform a systematic review where subjectivity is lower. Furthermore, the influence of inflation and purchasing power was not taken in account and currencies were not converted into one unified currency for better transferability of the study.

Weaknesses of the quality appraisal methods were revealed when assessing quality of the DAM of MBC EEs. When applying the checklist in this study, a full score of the dimension could be achieved only under condition of fulfilling all attributes. However, in the process if all attributes were sufficient and one attribute was not scored, it resulted in failure of meeting the full dimension criteria.

It is important to mention that review appraised quality of EEs by observing the presented information about quality dimensions. The process under each the quality dimension was not assessed, this is considered to be a weakness of this study. For example, in C1 the evidence about mathematical logic was stated in one EE, however in this thesis, the mathematical logic under statement was not tested. In other words only representation of given attributes were evaluated.

After applying the checklist [13] in this review, one important flaw was identified. The Philips checklist was thought to be too long and instead is suited best for appraisal of DAM in a single EE. In other words, it includes many attributes and in turn is impractical when systematically evaluating a number of EEs. A separate search of literature was conducted in order to identify other studies performing the quality assessment of DAM in MBC using the checklist by Philips et al. [13].

The search incorporating following key words, "good practice guidelines decision analytic model", were conducted. As a result three studies were identified, those used Philips checklist

[13]. One study appraised the quality of DAM in diabetes [72], another appraised DAM in Parkinson's disease [73] and third evaluated screening for abdominal aortic aneurysm [74].

By reviewing these studies it was found that Becker et al. have used Philips checklist in same manner as it was use in this study. However, study by Shearer et al. [73] used simplified approach of assessing quality, specifically only 15 quality criteria were used to assess quality. It was stated that simplified approach has been useful in review for drawing attention to the areas of strength and weaknesses across the modeling methods.

In these studies, findings similar to findings of this review were found. For example, Becker et al. [72] concluded that there is need of additional data on costs and quality of life. Furthermore, most of good practice requirements for diabetes modeling have been incorporated into the different model structures. Campbell et al. [74] have found no improvement in the standard of reporting, in studies, over time. Shearer et al. [73] have found that modeling studies did not provide sufficient argument on choices and decisions about model structure. Furthermore, pre-model analyses were not sufficiently conducted and internal and external consistency of modeling studies were not explored sufficiently. In conclusion, it is most likely that this review is one of the few which have used the checklist by Philips et al. [13] to appraise the quality of all DAMs for a particular disease.

4.4 Summary of general findings and future recommendations

Throughout the review it was found that costs and cost effectiveness of MBC treatments are rising with technological advancement over the period of reviewed EEs. This brings the importance of providing budgetary impact to decision makers instead provision only ICERs.

By analyzing the characteristics of EEs for MBC, it was found that inclusion of indirect cost alter the final results of EEs. The durations of time horizon may vary despite application of lifetime horizon, which is resulted by the varying age of the target populations considered by the reviewed EEs. Moreover, the choice of time horizon in EEs affects the discounting rate, and if time horizon is not mentioned, discounting rate cannot be applied.

In methodological characteristics of EEs for MBC, several trends were found. EEs have compared treatment sequences, and crossover design of comparison was used. A particular order of comparing of treatment sequences indicated the trend in EEs. Also it can be concluded that ICERs are sensitive to sequence of treatments. Additionally, it was found that ICERs are also sensitive to utility scores, type of comparison, sources of data used, patient age and country settings. Combinations of therapies are more cost-effective.

As a concluding remark about results of quality assessment, it is important to mention that only three sub-dimensions were fully scored by all EEs, including model type (S6), cycle length (S9), and parameter uncertainty (D4d). However, weaknesses in the quality of DAM were observed in all other sub-/dimensions. Based on this prevailing evidence, it is recommended that the

development of a disease specific DAM in MBC treatment therapies be conducted. The need for generic model, in early breast cancer, was first stated by Karnon et al.[56] and Annemans et al. [75], It is believed that a disease specific model will increase the quality in structure and data dimensions, especially in sub-dimensions were currently quality is less than 50%. Quality improvements in the variation in final results will be reduced, as a consequence of standardization of utility scores, type of comparison, sources of data used, health states, and patient age to which DAM results are sensitive.

Taking into account the biology of MBC development, methods of implementation, and quality of DAM for MBC therapies, the disease specific DAM should represent the combination and integrations of existing DAMs into one generic DAM. This will explain heterogeneity between the results of individual EEs, thus the first step is to review MBC disease progression and second, review of all DAMs conducted to evaluate MBC treatment therapies. Such a model could then be used to economically evaluate MBC treatment therapies and adapted depending on patient subgroup. A disease specific model will increase quality of DAMs in MBC by improving the quality in structure and data only under the condition that it also includes concrete information about the rationale of structure, data identified, pre-model data analysis, baseline data, treatment effects, and quality of life weights.

According to our results, rationale for structure and description of comparators were missing in most of the EEs. However, structural assumptions, time horizon, model type, disease pathways and cycle length were considered and chosen properly in most of the EEs. Regarding the data identification and pre-model data analysis quality was less satisfactory according to our appraisal. All these factors influence the DAM however, it is important to distinguish which of those have greater impact on results of DAM. Considering this review, it is thought that greater impact on results of DAM have data identification its analysis and incorporation into the model, which partially depends on analyst but greatly depends on availability of data. The rationale of structure is important and fully depends on analyst, which means that analyst is responsible for structuring model which should be consistent with coherent theory of health condition and underlying biological processes. In conclusion, it is thought that greater value should be given to reduction of subjectivity of individual analysts. Generic model will guarantee major improvements in construction and reporting of health economic decision models, and this on other hand will increase credibility of DAM for health policy decisions [74].

Based on the weaknesses and most importantly the causes, the disease specific generic model may contribute to quality improvements. The generic model in this case will contribute to narrow the discrepancies in visions of analysts when considering the DAM for same disease. The generic model will reduce subjectivity of individual analysts because the structure and assumptions will be considered beforehand with collaborations with medical representatives practicing treatment of MBC.

The disease specific generic model should incorporate the structure, its assumptions, and data which are considered to be a scarce. Firstly, it is thought that generic model should be developed through meta-analysis of literature about MBC disease, to guarantee the generic structure and most common assumptions regarding structure. Moreover, the existing data for input parameters should be incorporated in order to guarantee existence of "baseline" input parameters in generic model. On other hand there should be possibility for analyst to add additional data on his/her consideration. This may encompass values of treatment effects, costs, and quality of life weights. Furthermore, it is important to consider country settings, currency and methods of derivation of utility weights, whereas generic model should be country specific. Country specific MBC generic model will eliminate the discrepancies in results caused by usage of different currencies, different methods of utility derivation and costs of service provision.

Despite the findings of this review, it is still difficult to suggest the structure of disease specific generic model. Construction of generic model requires collaborative effort among variety of leading professionals in decision analytic modeling and medicine. It is important to mention that disease specific generic model will contribute to systematization and quality improvement of input data, model structure and will reduce subjectivity of individual modelers. However, it will not contribute to improvement of follow up periods of constructed models.

Despite the limitations of Markov models [18], Markov type decision analytic models do generate solid economic evidence and are suitable for long term modeling like in MBC therapies. The ability of the Markov model to represent repetitive events and the time dependence of both probabilities and utilities allows for more accurate representation of clinical settings [76]. The comprehensive economic appraisal model is a powerful tool for decision making over a range of economic environments. Taking into account the characteristics and quality of EEs for MBC, it was concluded that it is valuable to standardize the process by which DAM for MBC therapies is developed and to develop a generic model for MBC treatment therapy comparisons. It is thought that a unified generic model will contribute to improvements in quality of DAM and consistency of results in EEs for MBC.

5 Conclusions

This review was emerged from fact that economic evidence supports decisions, and it is crucial to assure the fine quality of this evidence. The trends in EEs of MBC were identified, also it was found that therapies in combination are more cost effective and duration of modeling (time horizon) influences final results. The flaws in quality of DAM were also observed. After reviewing the quality and trends in EEs, the idea of development disease specific generic DAM was provoked. The aim of this idea was strengthened by manifested heterogeneity of DAM in MBC and disease progression of MBC. Meta-analysis of all EEs based on DAM will give possibility to develop generic model. Main advantage of generic model is that it will improve the quality of DAM and reduce the variation of final results of DAM and hence improve the quality of economic evidence used in decision making.

6 References

1. Jemal, A., et al., *Global cancer statistics.* CA: a cancer journal for clinicians, 2011. **61**(2): p. 69-90.
2. Coughlin, S.S. and D.U. Ekwueme, *Breast cancer as a global health concern.* Cancer epidemiology, 2009. **33**(5): p. 315-318.
3. Greenberg, E.J., *The treatment of metastatic breast cancer.* CA: a cancer journal for clinicians, 1991. **41**(4): p. 242-256.
4. Blank, P.R., K.J. Dedes, and T.D. Szucs, *Cost effectiveness of cytotoxic and targeted therapy for metastatic breast cancer: a critical and systematic review.* PharmacoEconomics, 2010. **28**(8): p. 629.
5. Gralow, J.R., *Optimizing the treatment of metastatic breast cancer.* Breast Cancer Research and Treatment, 2005. **89**(1): p. S9-S15.
6. Earnshaw, J. and G. Lewis, *NICE Guide to the Methods of Technology Appraisal: Pharmaceutical Industry Perspective.* PharmacoEconomics, 2008. **26**(9): p. 725-725.
7. Coast, J., *Is Economic Evaluation In Touch With Society's Health Values?* BMJ: British Medical Journal, 2004. **329**(7476): p. 1233-1236.
8. O'Brien, B.J. and M.J. Sculpher, *Building Uncertainty Into Cost-Effectiveness Rankings.* Medical Care, 2000. **38**(5): p. 460-468.
9. Drummond, M.F., *Methods for the economic evaluation of health care programmes: Michael F. Drummond ... [et al.]*2005, Oxford [etc.]: Oxford University Press.
10. Mauskopf, J.A., et al., *Principles of good practice for budget impact analysis: report of the ISPOR Task Force on good research practices--budget impact analysis.* Value in Health, 2007. **10**(5): p. 336-347.
11. Trueman, P., M. Drummond, and J. Hutton, *Developing guidance for budget impact analysis.* PharmacoEconomics, 2001. **19**(6): p. 609-609.
12. Gulácsi, L., E. Orlewska, and M. Péntek, *Health economics and health technology assessment in Central and Eastern Europe: a dose of reality.* The European Journal of Health Economics, 2012. **13**(5): p. 525-531.
13. Philips, Z., et al., *Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment*, 2006, Adis International: New Zealand. p. 355-355.
14. *Decision Analytic Modelling in the Economic Evaluation of Health Technologies.* PharmacoEconomics, 2000. **17**(5): p. 443-444.
15. Sun, X. and T. Faunce, *Decision-analytical modelling in health-care economic evaluations.* The European Journal of Health Economics, 2008. **9**(4): p. 313-323.

16. Drummond, M. and M. Sculpher, *Common methodological flaws in economic evaluations.* Medical Care, 2005. **43**(7 Suppl): p. 5-II14.
17. Drummond, M. and M. Sculpher, *Common Methodological Flaws in Economic Evaluations.* Medical Care, 2005. **43**(Supplement): p. II-5-II-14.
18. Caro, J.J., J. Möller, and D. Getsios, *Discrete event simulation: the preferred technique for health economic evaluations?* Value in Health, 2010. **13**(8): p. 1056-1060.
19. Briggs, A. and M. Sculpher, *An Introduction to Markov Modelling for Economic Evaluation.* PharmacoEconomics, 1998. **13**(4): p. 397-409.
20. Sutton, A.J., *Methods for meta-analysis in medical research: Alex J. Sutton ... [et al.]*2000, Chichester [etc.]: Wiley.
21. Prevost, T.C., K.R. Abrams, and D.R. Jones, *Hierarchical models in generalized synthesis of evidence: an example based on studies of breast cancer screening.* Statistics in Medicine, 2000. **19**(24): p. 3359-3376.
22. Briggs, A., K. Claxton, and M.J. Sculpher, *Decision modelling for health economic evaluation: Andrew Briggs, Karl Claxton, Mark Sculpher*2006, Oxford [etc.]: Oxford University Press.
23. Griffin, S.C., et al., *Dangerous omissions: the consequences of ignoring decision uncertainty.* Health economics, 2011. **20**(2): p. 212-224.
24. Siebert, U., *When Should Decision-Analytic Modeling Be Used in the Economic Evaluation of Health Care?* The European Journal of Health Economics, 2003. **4**(3): p. 143-150.
25. Khan, Z.M. and D.W. Miller, *Modeling economic evaluations of pharmaceuticals: manipulation or valuable tool?* Clinical Therapeutics, 1999. **21**(5): p. 896-908.
26. Zimovetz, E. and S.E. Wolowacz, *PMC45 REVIEWER'S CHECKLIST FOR ASSESSING THE QUALITY OF DECISION MODELS.* Value in Health, 2009. **12**(7): p. A395-A395.
27. Philips, Z., et al., *Review of guidelines for good practice in decision-analytic modelling in health technology assessment.* Health technology assessment (Winchester, England), 2004. **8**(36): p. iii.
28. Karnon, J., T. Delea, and V. Barghout, *Cost utility analysis of early adjuvant letrozole or anastrozole versus tamoxifen in postmenopausal women with early invasive breast cancer: the UK perspective.* Eur J Health Econ, 2008. **9**(2): p. 171-83.
29. Cooper, N.J., et al., *Comprehensive decision analytical modelling in economic evaluation: A Bayesian approach.* Health Economics, 2004. **13**(3): p. 203-206.

30. Mould-Quevedo, J.F. and I. Contreras-Hernandez, *Economic evaluation of adjuvant hormone therapy for postmenopausal women with hormone receptor positive early stage breast cancer.* Pharmacoeconomics - Spanish Research Articles, 2011. **8**(2): p. 59-70.
31. Carter, J.A., et al., *Cost-effectiveness of zoledronic acid (ZOL) versus denosumab (Dmab) in prevention of skeletal-related events (SREs) in metastatic breast cancer (mBC).* Journal of Clinical Oncology, 2011. **29**(15).
32. Lothgren, M., et al., *Cost-effectiveness of denosumab versus zoledronic acid (ZA) for the prevention of skeletal-related events (SRE) in patients with bone metastases from solid tumors in the Netherlands.* Value in Health, 2011. **14**(7): p. A455.
33. Annemans, L., et al., *Health economic evaluation of controlled and maintained physical exercise in the prevention of cardiovascular and other prosperity diseases.* Eur J Cardiovasc Prev Rehabil, 2007. **14**(6): p. 815-24.
34. Ladabaum, U. and K.A. Phillips, *Colorectal cancer screening: differential costs for younger versus older Americans.* American Journal of Preventive Medicine, 2006. **30**(5): p. 378-384.
35. Yang, M. and A. Issa, *Assessment of the cost-effectiveness of pharmacogenomic diagnostics for breast cancer.* Pharmacoepidemiology and Drug Safety, 2010. **19**: p. S107.
36. Rahman, M., D. Malone, and E. Armstrong, *An economic evaluation of circulating tumor cells detection in metastatic breast cancer.* Value in Health, 2009. **12**(3): p. A43.
37. Okubo, I., et al., *Cost-effectiveness of letrozole versus tamoxifen as first-line hormonal therapy in treating postmenopausal women with advanced breast cancer in Japan.* Gan To Kagaku Ryoho, 2005. **32**(3): p. 351-63.
38. Prunieras, F., et al., *Indirect comparison and uk cost-effectiveness analysis of paclitaxel albumin vs docetaxel for pretreated metastatic breast cancer.* Annals of Oncology, 2010. **21**: p. viii107-viii108.
39. Tenorio, C., J. Vargas, and J. Martinez, *Cost-effectiveness evaluation of the use of capecitabine+docetaxel vs gemcitabine+docetaxel in patients with recurrent breast cancer who previously failed to anthracycline chemotherapy and/or with metastatic disease.* Value in Health, 2009. **12**(7): p. A271.
40. Jones, J., et al., *Gemcitabine for the treatment of metastatic breast cancer.* Health Technol Assess, 2009. **13 Suppl 2**: p. 1-7.

41. Dedes, K.J., et al., *Bevacizumab in combination with paclitaxel for HER-2 negative metastatic breast cancer: An economic evaluation.* Cancer Research, 2009. **69**(2).
42. Karnon, J. and T. Jones, *A stochastic economic evaluation of letrozole versus tamoxifen as a first-line hormonal therapy: for advanced breast cancer in postmenopausal patients.* Pharmacoeconomics, 2003. **21**(7): p. 513-25.
43. Cameron, D.A., et al., *Economic evaluation of fulvestrant as an extra step in the treatment sequence for ER-positive advanced breast cancer.* Br J Cancer, 2008. **99**(12): p. 1984-90.
44. Benedict, A., et al., *An economic evaluation of docetaxel and paclitaxel regimens in metastatic breast cancer in the UK.* Pharmacoeconomics, 2009. **27**(10): p. 847-59.
45. Lux, M.P., et al., *Cost-utility analysis for advanced breast cancer therapy in Germany: Results of the fulvestrant sequencing model.* Breast Cancer Research and Treatment, 2009. **117**(2): p. 305-317.
46. Frias, C., et al., *Cost-effectiveness analyses of docetaxel versus paclitaxel once weekly in patients with metastatic breast cancer in progression following anthracycline chemotherapy, in Spain.* Clin Transl Oncol, 2010. **12**(10): p. 692-700.
47. Hedden, L., et al., *Assessing the real-world cost-effectiveness of adjuvant trastuzumab in HER-2/neu positive breast cancer.* Oncologist, 2012. **17**(2): p. 164-171.
48. Elkin, E.B., et al., *HER-2 testing and trastuzumab therapy for metastatic breast cancer: a cost-effectiveness analysis.* Journal of Clinical Oncology, 2004. **22**(5): p. 854-863.
49. Marchetti, M., M. Caruggi, and G. Colombo, *Cost utility and budget impact of third-generation aromatase inhibitors for advanced breast cancer: a literature-based model analysis of costs in the Italian National Health Service.* Clin Ther, 2004. **26**(9): p. 1546-61.
50. Lidgren, M., et al., *Cost-effectiveness of HER2 testing and trastuzumab therapy for metastatic breast cancer.* Acta Oncologica, 2008. **47**(6): p. 1018-1028.
51. Dedes, K.J., et al., *Bevacizumab in combination with paclitaxel for HER-2 negative metastatic breast cancer: an economic evaluation.* Eur J Cancer, 2009. **45**(8): p. 1397-406.
52. Le, Q.A. and J.W. Hay, *Cost-effectiveness analysis of lapatinib in HER-2-positive advanced breast cancer.* Cancer, 2009. **115**(3): p. 489-498.

53. Matter-Walstra, K.W., et al., *Trastuzumab beyond progression: A cost-utility analysis.* Annals of Oncology, 2010. **21**(11): p. 2161-2168.
54. Karnon, J., et al., *Tamoxifen plus chemotherapy versus tamoxifen alone as adjuvant therapies for node-positive postmenopausal women with early breast cancer: a stochastic economic evaluation.* PharmacoEconomics, 2002. **20**(2): p. 119-119.
55. Lidgren, M., et al., *Cost-effectiveness of HER2 testing and 1-year adjuvant trastuzumab therapy for early breast cancer.* Annals of Oncology, 2008. **19**(3): p. 487-495.
56. Karnon, J., *Aromatase Inhibitors in Breast Cancer: A Review of Cost Considerations and Cost Effectiveness*, 2006, Adis International: New Zealand. p. 215-215.
57. Johannesson, M., *A note on the depreciation of the societal perspective in economic evaluation of health care.* Health policy (Amsterdam, Netherlands), 1995. **33**(1): p. 59-66.
58. Jönsson, B., *Ten Arguments for a Societal Perspective in the Economic Evaluation of Medical Innovations.* The European Journal of Health Economics, 2009. **10**(4): p. 357-359.
59. Bager, P. and J.F. Dahlerup, *The health care cost of intravenous iron treatment in IBD patients depends on the economic evaluation perspective.* Journal of Crohn's and Colitis, 2010. **4**(4): p. 427-430.
60. Nuijten, M. and T. Mittendorf, *A health-economic evaluation of disease-modifying drugs for the treatment of relapsing-remitting multiple sclerosis from the German societal perspective.* Clinical Therapeutics, 2010. **32**(4): p. 717-728.
61. Sander, B. and M.D. Krahn, *MC4 TIME HORIZON BIAS IN ECONOMIC EVALUATIONS.* Value in Health, 2007. **10**(3): p. A8-A8.
62. Williams I, M.S., Moore D, Bryan S. , *The use of economic evaluations in NHS decision-making: a review and empirical investigation.* Health Technol Assess, 2008. **12**(7).
63. Hall, P.S., et al., *Updated Cost-Effectiveness Analysis of Trastuzumab for Early Breast Cancer A UK Perspective Considering Duration of Benefit, Long-Term Toxicity and Pattern of Recurrence.* PharmacoEconomics, 2011. **29**(5): p. 415-432.
64. Soto, J., *Health economic evaluations using decision analytic modeling. Principles and practices--utilization of a checklist to their development and appraisal.* International Journal of Technology Assessment in Health Care, 2002. **18**(1): p. 94-111.

65. Nuijten, M.J., et al., *Reporting format for economic evaluation. Part II: Focus on modelling studies.* PharmacoEconomics, 1998. **14**(3): p. 259-259.
66. Sculpher, M., E. Fenwick, and K. Claxton, *Assessing Quality in Decision Analytic Cost-Effectiveness Models: A Suggested Framework and Example of Application*, 2000, Adis International: NEW ZEALAND. p. 461-461.
67. Sonnenberg, F.A., et al., *Toward a peer review process for medical decision analysis models.* Medical Care, 1994. **32**(7 Suppl): p. JS52-JS64.
68. Weinstein, M.C., et al., *Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies.* Value in Health, 2003. **6**(1): p. 9-17.
69. Soto, J., *HEALTH ECONOMIC EVALUATIONS USING DECISION ANALYTIC MODELING.* International Journal of Technology Assessment in Health Care, 2002. **18**(1): p. 94-111.
70. Philips, Z., et al., *Review of guidelines for good practice in decision-analytic modelling in health technology assessment.* Health Technology Assessment, 2004. **8**(36): p. 1-1.
71. Buxton, M.J., et al., *Modelling in economic evaluation: an unavoidable fact of life.* Health economics, 1997. **6**(3): p. 217-227.
72. Christian, B., L. Astrid, and L. Reiner, *The quality of three decision-analytic diabetes models: a systematic health economic assessment.* Expert Review of Pharmacoeconomics & Outcomes Research, 2011. **11**(6): p. 751.
73. Shearer, J., et al., *The use of decision-analytic models in Parkinson's disease: A systematic review and critical appraisal.* Applied Health Economics and Health Policy, 2011. **9**(4): p. 243-258.
74. Campbell, H., et al., *The credibility of health economic models for health policy decision-making: the case of population screening for abdominal aortic aneurysm.* Journal of health services research & policy, 2007. **12**(1): p. 11-17.
75. Annemans, L., *Methodological issues in evaluating cost effectiveness of adjuvant aromatase inhibitors in early breast cancer: a need for improved modelling to aid decision making*, 2008, Adis International: New Zealand. p. 409-409.
76. Sonnenberg, F.A. and J.R. Beck, *MARKOV-MODELS IN MEDICAL DECISION-MAKING - A PRACTICAL GUIDE.* Medical Decision Making, 1993. **13**(4): p. 322-338.

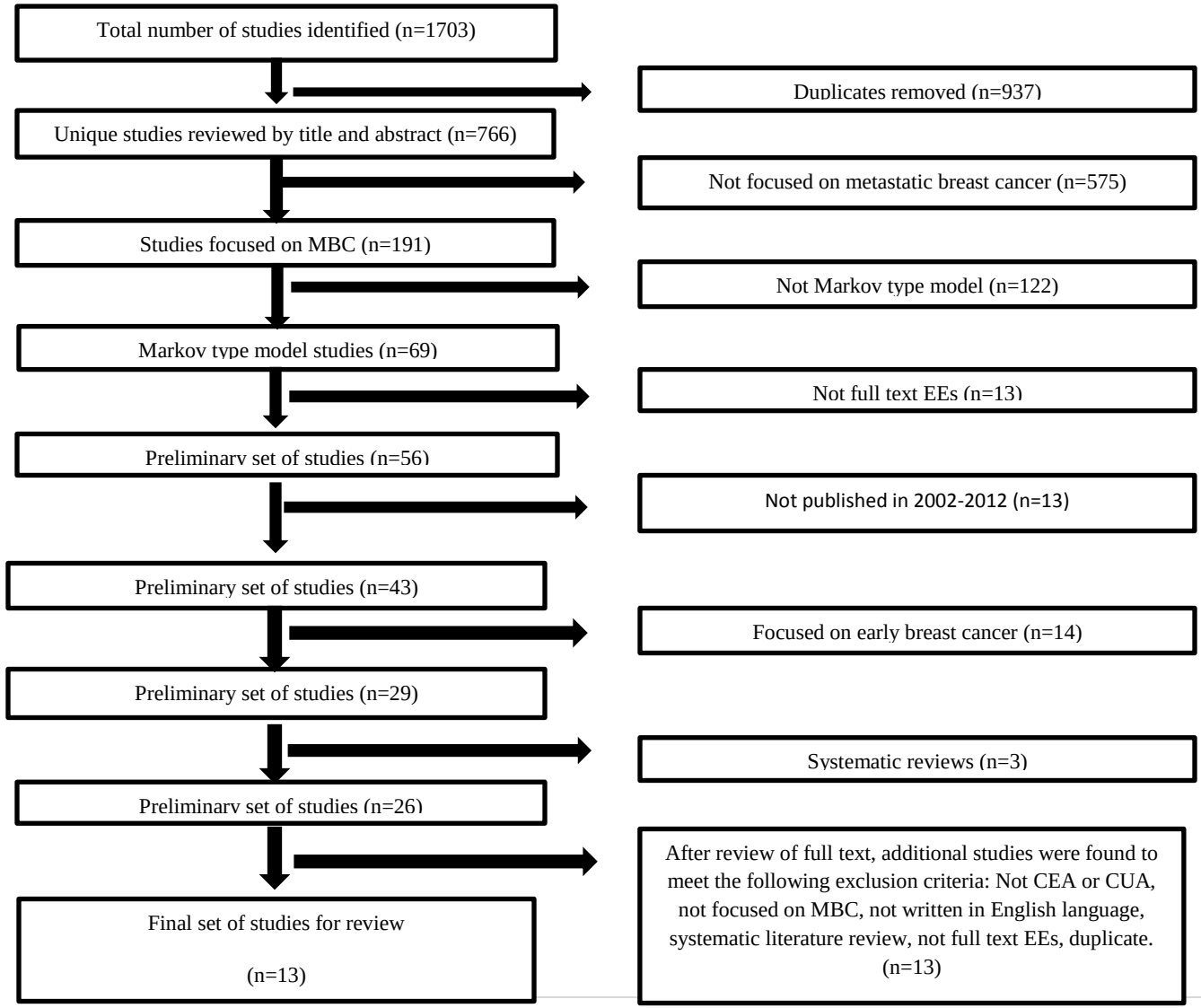
7 Appendix 1

Table – 1 Search Strategy

#	Query
#9	'late stage' AND 'breast cancer'/exp AND 'cost benefit'/exp AND 'economic evaluation'/exp
#8	'metastatic' AND 'breast cancer'/exp AND 'cost benefit'/exp AND 'economic evaluation'/exp
#7	'advanced' AND 'breast cancer'/exp AND 'cost benefit'/exp AND 'economic evaluation'/exp
#6	'late stage' AND 'breast cancer'/exp AND 'cost utility'/exp AND 'economic evaluation'/exp
#5	'late stage' AND 'breast cancer'/exp AND 'cost effectiveness'/exp AND 'economic evaluation'/exp
#4	'metastatic' AND 'breast cancer'/exp AND 'cost utility'/exp AND 'economic evaluation'/exp
#3	'advanced' AND 'breast cancer'/exp AND 'cost utility'/exp AND 'economic evaluation'/exp
#2	metastatic' AND 'breast cancer'/exp AND 'cost effectiveness'/exp AND 'economic evaluation'/exp
#1	'advanced' AND 'breast cancer'/exp AND 'cost effectiveness'/exp AND 'economic evaluation'/exp

8 Appendix 2

Figure – 1 Selection of Economic Evaluation Studies



9 Appendix 3

Table 2 - Proposed framework for quality assessment of decision-analytic models taken from Philips et al. [13]		
Dimension of Quality	Attribute for good practice	Question for critical appraisal
Structure (S)		
S1 statement of decision problem/objective	There should be clear statement of decision problem prompting analysis. The objective of the evaluation and of model should be defined. The primary decision maker should be stated clearly	a) Is there clear statement of decision problem?
		b) Is the objective of the evaluation and model specified and consistent with stated decision problem?
		c) Is the primary decision maker specified?
S2 statement of scope/perspective	Perspective of the model (relevant costs and consequences) should be stated clearly, and the model inputs should be consistent with stated perspective and overall objective of the model. The scope of the decision model should be specified and justified. The outcomes of the model should reflect the perspective and scope of the model and should be consistent with the objective of the evaluation.	a) Is the perspective of the model stated clearly?
		b) Are model inputs consistent with stated perspective?
		c) Has the scope of the model been stated and justified?
		d) Are the outcomes of the model consistent with perspective, scope and overall objective of the model?
S3 rationale for structure	The structure of the model should be consistent with a coherent theory of the health condition under evaluation. Treatment pathways (disease states or branches) should be chosen to reflect the underlying biological process of the disease in question and the impact of the intervention. The structure should not be dictated by current patterns of service provision. All sources of evidence used to develop and inform the structure of the model (i.e. the theory of disease) should be described. The structure should be consistent with this evidence	a) Has the evidence regarding the model structure been described?
		b) Is the structure of the model consistent with a coherent theory of the health condition under evaluation?
		c) Have any competing theories regarding model structure been considered?
		d) Are the sources of data used to develop the structure of the model specified?
		e) Are the casual relationships described by the model structure justified appropriately?
S4 structural assumptions	All structural assumptions should be transparent and justified. They should be reasonable in the light of the needs and purposes of the decision-maker.	a) Are the structural assumptions transparent and justified?
		b) Are the structural assumptions reasonable given the overall objective, perspective and scope of the model?
S5 strategies/comparators	There should be a clear definition of the options under evaluation. All feasible and practical options related to the	a) Is there a clear definition of the options under evaluation?

	stated decision problem should be evaluated. Options should not be constrained by the immediate concerns of the decision-maker or data availability, or limited to current clinical practice.	b) Have all feasible and practical options been evaluated?
		c) Is there justification for exclusion of feasible options?
S6 model type	The appropriate model type will be dictated by the stated decision problem and the choices made regarding the casual relationships within the model	a) Is the chosen model type appropriate given the decision problem and specified casual relationships within the model?
S7 time horizon	A model's time horizon should extend far enough into the future for it to reflect important differences between options. It is important to distinguish between the time horizon of the model, the duration of treatment and the duration of treatment effect.	a) Is the time horizon of the model sufficient to reflect all important differences between options?
		b) Are the time horizon of the model, the duration of treatment and the duration of treatment effect described and justified?
		c) Has a lifetime horizon been used? If not, has a shorter time horizon been justified?
S8 disease states/pathways	disease states/pathways should reflect the underlying biological process of the disease in question and the impact of interventions	a) Do the disease states (state transition model) or the pathways (decision tree model) reflect the underlying biological process of the disease in question and the impact of intervention?
S9 cycle length	For discrete time models, the cycle length should be dictated by the natural history of disease. It should be the minimum interval over which the pathology or symptoms are expected to alter	a) Is the cycle length defined and justified in terms of the natural history of disease?
Data (D)		
D1 data identification	Methods for identifying data should be transparent and it should be clear that the data identified are appropriate given the objective of the model. There should be justification of any choices that have been made about which specific data inputs are included in a model. It should be clear that particular attention has been paid to identifying data for those parameters to which the results of the model are particularly sensitive. Where expert opinion has been used to estimate particular parameters, sources and methods of elicitation	a) Are the data identification methods transparent and appropriate given the objectives of the model?
		b) Where choices have been made between data sources, are these justified appropriately?
		c) Has particular attention been paid to identifying data for the important parameters in the model?
		d) Has the process of selecting key parameters been

	should be described	justified and systematic methods used to identify the most appropriate data?
		e) Has the quality of the data been assessed appropriately?
		f) Where expert opinion has been used, are the methods described and justified?
D2 pre-model data analysis	All data modeling methodology should be described and based on justifiable statistical and epidemiological methods. Specific issues to consider include those listed under D2a-d, below	a) Is the pre-model data analysis methodology based on justifiable statistical and epidemiological techniques?
D2a Baseline data	Baseline probabilities may be based on natural history data derived from epidemiological/observational studies or related to the control group of an experimental study. Rates and interval probabilities should be transformed into transition probabilities appropriately. If there is evidence that time is an important factor in the calculation of transition probabilities in state transition models, this should be incorporated. if a half-cycle correction has not been used on all transitions in state transition model (costs and outcomes), this should be justified.	a) Is the choice of baseline data described and justified?
		b) Are the transition probabilities calculated appropriately?
		c) Has the half cycle correction been applied to both cost and outcome?
		d) If not, omission been justified?
D2b treatment effects	Relative treatment effects derived from trial data should be synthesized using recognized meta-analytic techniques. The methods and assumptions that are used to extrapolate short-term results to final outcomes should be documented and justified. This should include justification of the choice of survival function (e.g. exponential or Weibull forms). Alternative assumptions should be explored through sensitivity analysis. Assumptions regarding the continuing effect of treatment once treatment is complete should be documented and justified. If evidence regarding the long-term effect of treatments lacking, alternative assumptions should be explored through sensitivity analysis.	a) If relative treatment effects have been derived from trial data, have they been synthesized using appropriate techniques?
		b) Have the methods and assumptions used to extrapolate short-term results to final outcomes been documented and justified?
		c) Have the alternative extrapolation assumptions been explored though sensitivity analysis?
		d) Have the assumptions regarding the continuing effect of treatment once treatment is complete been documented and justified?
		e) Have the alternative assumptions regarding the continuing effect of treatment been explored thought sensitivity analysis?

D2c Costs	Costing and discounting methods should be accord with standard guidelines for Economic Evaluation	a) Are the costs incorporated into the model justified?
		b) Has the source for all costs been described?
		c) Have discount rates been described and justified given the target decision-maker?
D2d quality-of-life weights (utilities)	Utilities incorporated into the model should be appropriate for specified decision problem	a) Are utilities incorporated into the model appropriately?
		b) Is the source for the utility weights referenced?
		c) Are the methods of derivation for the utility weights justified?
D3 Data incorporation	All data incorporated into the model should be described and the sources of all data should be given and reported in sufficient detail to allow the reader to be aware of the type of data that have been incorporated. Where data are not mutually consistent in the model, the choice and assumptions that have been made should be explicit and justified. The process of data incorporation should be transparent. it should be clear whether data are incorporated as a point estimate or as a distribution. If data have been incorporated as distributions as part of probabilistic analysis, the choice of distribution and its parameters should be described and justified.	a) Have all data incorporated into the model been described and referenced in sufficient detail?
		b) Has the use of mutually inconsistent data been justified (I.e. are assumptions and choices appropriate)?
		c) Is the process of data incorporation transparent?
		d) If data have been incorporated as distributions, has the choice of distribution for each parameter been described and justified?
		e) If data have been incorporated as distributions, is it clear that second order uncertainty is reflected?
D4 assessment of uncertainty	In assessing uncertainty, modelers should distinguish between the four principle types of uncertainty	a) Have the four principal types of uncertainty been addressed?
		b) If not, has the omission of particular forms of uncertainty been justified?
D4a methodological	Methodological uncertainty relates to whether particular analytical steps taken in the analysis are the most appropriate	a) Have methodological uncertainties been addressed by running alternative versions of the model with different methodological assumptions?
D4b structural	There should be evidence that structural uncertainties have been evaluated using sensitivity analysis	a) Is there evidence that structural uncertainties have been addressed via sensitivity analysis?

D4c heterogeneity	It is important to distinguish between uncertainty resulting from the process of sampling from a population and variability due to heterogeneity (i.e. systematic differences between patient subgroups)	a) Has heterogeneity been dealt with by running the model separately for different subgroups? (treatment effect variation)
D4d parameter	Where data have been incorporated into model as a point estimates, the ranges used for sensitivity analysis should be stated and justified. Probabilistic analysis is the most appropriate method of handling parameter uncertainty because it facilitates assessment of joint effect of uncertainty over all parameters.	a) Are the methods of assessment of parameter uncertainty appropriate?
		b) Has probabilistic sensitivity analysis been done, if not has this been justified?
		c) If data are incorporated as a point estimates, are the ranges used for sensitivity analysis stated clearly and justified?
Consistency (C)		
C1 internal consistency	There should be evidence that the internal consistency of the model has been evaluated in terms of its mathematical logic	a) Is there evidence that the mathematical logic of the model has been tested thoroughly before use?
C2 external consistency	The results of a model should be explicable. Either result should make intuitive sense or contra intuitive results should be fully explained. All relevant available data should be incorporated into a model. Data should not be withheld for purposes of assessing external consistency. The results of a model should be compared with those of previous models and any differences should be explained.	a) Are the conclusions valid given the data presented?
		b) Are any contra intuitive results from model explained and justified?
		c) If the model has been calibrated against independent data, have any differences been explained and justified?
		d) Have the results of the model been compared with those of previous models and any differences in results explained?

10 Appendix 4

Table 3 – Cost-effectiveness results and methodological characteristics of published economic evaluations in metastatic breast cancer									
Study	**Study Design**	**Perspective/ country**	**Indirect costs**	**Comparators**	**ICER**	**effectiveness outcomes**	**time horizon**	**discounting rate**	**Target Population**
Karnon and Jones [42]	CEA /CUA	UK NHS	none	Letrozole vs. Tamoxifene as a first-line hormonal therapies in postmenopausal women diagnosed with advanced breast cancer	ICER = £6068/LYG, £2927-3969/QALY gained with letrozole	LYG, QALY	Full life time	6% for resources. 1.5% for life years applied to analysis	Advanced breast cancer. Diagnosis to death.
Elkin, Weinstein [48]	CUA	Not stated (USA)	none	Hercept Test, FISH alone and both followed by trastuzumab and chemotherapy	Base case: HercepTest with FISH $125,000/QALY. Initial FISH only $145,000/QALY	QALY and QALM (month)	lifetime	3% annually	65 year old MBC patients
Marchetti, Caruggi [49]	CUA	Italian National Health Service	none	Three Hormone therapies: 1) anastrazol followed by tamoxifen, then megestrol; 2) letrozole followed by tamoxifen, then megestrol; 3) tamoxifene followed by anastrazole, then megestrol.	1 vs 3 = €10795/QALY. 2 vs 3 = €16886/QALY	QALY	8.3 years (100 month)	3% to resources and Life-years gained	Post-menopausal women with estrogen positive MBC
Cameron, Camidge [43]	CEA /CUA	NICE/Scottish medicine Consortium (UK NHS)	none	Fluvestrant used as a second-line hormonal therapy for Advanced Breast cancer (Cohort A vs. B) Fluvestrant used as a third-line hormonal therapy for Advanced Breast cancer (Cohort A vs. B).	Fluvestrant used as a second-line hormonal therapy: £6500/LYG, and £7300/QALY. Fluvestrant used as a third-line hormonal therapy: cost saving of £430 per patient	LYG and QALY	Life time (10 years)	3.5% annually	ER-positive advanced MBC

Karnon, Delea [28]	CEA /CUA	UK NHS	none	lestrozole vs. anastrazole by comparing aromataze inhibitors to tamoxifen. Letrozole vs. tamoxifen, and anastrozole vs. tamoxifen	Letrozole vs. tamoxifene £ 10,502/LYG and incremental cost per QALY is £10,379. Anastrazole vs. tamoxifen £11,703/LYG and incremental cost per QALY is £11,428	LY, QALY	5 years	3.5% annually both costs and LY/QALY	Women of 61 years old with early invasive Breast cancer
Lidgren, Wilking [50]	CUA	Societal (Sweden)	Average annual inpatient costs and average annual informal care costs	HER-2 testing and trastuzumab in combination with chemotherapy (strategy 2) compared with Chemotherapy alone (strategy 1)	561207 SEK/QALY	QALY	Life-time	not done	Hypothetical cohort of 65 years old MBC patients
Benedict, Cameron [44]	CEA /CUA	UK NHS	none	Docetaxel vs. Paclitaxel regimens in treatment of MBC previously with an Anthracycline	ICERs for docetaxel were d12 032/QALY Versus Pac3w, 4583/QALYversus Pac1w and d14 694/QALYversus Nab-P.	LYG, and QALY	Life time (10 years)	3.5% annually	MBC patients progressed after anthracycline-containing chemotherapy
Dedes, Matter-Walstra [51]	CUA	Swiss Health Care System	none	Paclitaxel/Bevacizumab+Paclitaxel	Pac+Bev versus Pac alone was €189.427QALY gained	QALY	Lifetime	not done	MBC patients

Le and Hay [52]	CUA	US Societal (USA)	Patient Time cost and travel costs estimated and included	Capecitabine monotherapy vs. Capecitabine plus Lapatinib combination therapy	$ 166,113 per QALY (or approximately $13,843 per QALM)	QALY, and QALM (month)	Life-time	3% time not specified	Advanced breast cancer patients age of 53
Lux, Hartmann [45]	CEA /CUA	German health care system	none	Therapy sequence including second-line fulvestrant (Cohort A) to corresponding cohort whose sequence did not include fulvestrant (Cohort B)	Cost per LY for Cohort A: $6,645 Cohort B: $7,160 Cost per QALY for Cohort A: $11,614 Cohort B: $12,319. Cohort A dominates	QALY and LY	Life time	3% annually	Post-menopausal women with Advanced breast cancer
Frias, Cortes [46]	CEA /CUA	NHS (Spain), 3rd party payer	none	CE of Docetaxel vs. Paclitaxel	€ 190/LYG €295/QALY	LYG and QALY	5-years	3% period not specified	Patients with MBC
Matter-Walstra, Dedes [53]	CUA	Swiss health care system	none	Trastuzumab plus capecitabine vs. capecitabine alone	€98,329 per QALY	QALY	not stated	not stated	Patients with MBC
Hedden, O'Reilly [47]	CEA /CUA	Payer perspective (Canada)	none	Adjuvant trastuzumab for operable, HER-2/neu-positive early breast cancer, accounting for differences in costs and health outcomes associated with trastuzumab and standard of care treatments in adjuvant and metastatic settings	ICER $8,479; $13,095/QALY, $15,492/LYG	QALY and LYG	28 years	both costs and outcomes 3% annually	50 year old women with early HER-2/neu-positive breast cancer

11 Appendix 5

Table 4 - Quality appraisal of decision analytic modeling in selected Economic evaluations of metastatic breast cancer

Study		Karnon et al. 2003	Elkin et al. 2004	Marchetti et al. 2004	Cameron et al. 2008	Karnon et al. 2008	Lidgren et al. 2008	Benedict et al. 2009	Dedes et al. 2009	Le et al. 2009	Lux et al. 2009	Frias et al. 2010	Matter-Walstra et al. 2010	Heddenet et al. 2012	Percentage of positive answers %/	(no.) Studies satisfied the whole dimensions
Dimension of Quality																
Structure (S)																
S1 Statement of decision problem/objective	a)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	10
	b)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	
	c)	yes	no	yes	yes	yes	yes	yes	yes	no	yes	yes	yes	no	76.92	
S2 Statement of scope/perspective	a)	yes	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	92.3	10
	b)	yes	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	92.3	
	c)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	
	d)	yes	no	yes	no	yes	yes	no	yes	yes	yes	yes	yes	yes	76.92	
S3 Rationale for structure	a)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	0
	b)	no	yes	no	no	yes	yes	yes	yes	yes	no	yes	yes	yes	69.23	
	c)	no	no	no	no	no	no	no	no	no	no	no	no	no	0	
	d)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	
	e)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	
S4 Structural assumptions	a)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	11
	b)	yes	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	92.3	
S5 Strategies/comparators	a)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	1
	b)	no	yes	no	no	no	no	no	no	no	no	no	no	no	7.69	
	c)	no	yes	no	no	no	no	no	no	no	no	no	no	no	7.69	
S6 Model type	a)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	13
S7 Time horizon	a)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	no	yes	92.3	12

	b)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	no	yes	92.3	
	c)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	no	yes	92.3	
S8 Disease states/pathways	a)	no	yes	no	no	yes	yes	yes	yes	yes	no	yes	yes	yes	69.23	9
S9 Cycle length	a)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	13
Data (D)																
D1 Data identification	a)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	2
	b)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	
	c)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	
	d)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	
	e)	no	no	no	no	no	yes	yes	no	no	yes	no	no	no	23	
	f)	yes	no	no	yes	yes	yes	no	no	no	yes	no	no	no	38	
D2 Pre-model data analysis	a)	no	no	no	no	no	no	no	no	no	no	no	no	no	0	0
D2a Baseline data	a)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	0
	b)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	
	c)	yes	no	no	no	yes	no	no	no	no	no	no	no	no	15	
	d)	no	no	no	no	no	no	no	no	no	no	no	no	no	0	
D2b Treatment effects	a)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	6
	b)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	
	c)	yes	yes	yes	yes	yes	yes	yes	yes	yes	no	yes	yes	yes	92	
	d)	yes	yes	no	yes	no	no	yes	yes	yes	no	no	yes	yes	62	
	e)	yes	yes	no	yes	no	no	yes	yes	yes	no	no	no	no	46	
D2c Costs	a)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	10
	b)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	
	c)	yes	yes	yes	yes	yes	no	yes	no	yes	yes	yes	no	yes	77	
D2d Quality-of-life weights (utilities)	a)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	6
	b)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	
	c)	no	yes	no	yes	yes	yes	no	yes	no	yes	no	no	no	46	
D3 Data incorporation	a)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	12
	b)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	

	c)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	
	d)	yes	yes	yes	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	92	
	e)	yes	yes	yes	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	92	
D4 Assessment of uncertainty	a)	no	no	no	yes	no	no	no	no	no	yes	no	no	no	15	2
	b)	no	no	no	no	no	no	no	no	no	no	no	no	no	0	
D4a Methodological	a)	no	no	no	yes	no	no	yes	no	no	yes	no	no	no	23	3
D4b Structural	a)	no	yes	no	yes	no	yes	no	yes	yes	yes	no	yes	no	54	7
D4c Heterogeneity	a)	no	yes	no	yes	no	yes	no	yes	no	yes	yes	yes	no	54	7
D4d Parameter	a)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	13
	b)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	
	c)	no	no	no	yes	no	no	no	no	no	no	no	no	no	8	
Consistency (C)																
C1 Internal consistency	a)	no	no	no	no	no	no	yes	no	no	no	no	no	no	8	1
C2 External consistency	a)	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	100	0
	b)	no	no	no	no	no	no	yes	no	no	no	yes	no	no	15	
	c)	no	no	no	no	no	yes	no	yes	yes	no	no	yes	no	31	
	d)	yes	yes	yes	yes	yes	yes	yes	no	no	yes	yes	yes	yes	85	

12 Appendix 6

Table 5 - Quality of decision analytic modeling of reviewed studies

Study		Karnon et al. 2003	Elkin et al. 2004	Marchetti et al. 2004	Cameron et al. 2008	Karnon et al. 2008	Lidgren et al. 2008	Benedict et al. 2009	Dedes et al. 2009	Le et al. 2009	Lux et al. 2009	Frias et al. 2010	Matter-Walstra et al. 2010	Hedden et al. 2012	Average % per quality dimension
Quality dimension															
	Total no. Qs														
Structure (ovarall)	23														
Positive scores (no.)		18	17	18	17	20	20	19	20	19	18	20	17	19	
Quality %		78	74	78	74	87	87	83	87	83	78	87	74	83	81
Data															
Pre model data analysis	15														
Positive scores (no.)		13	13	10	13	12	10	12	12	12	10	10	10	11	
Quality %		87	87	67	87	80	67	80	80	92	67	77	77	73	78
Data (overall)	35														
Positive scores (no.)		24	25	20	27	23	24	26	24	23	26	21	22	21	
Quality %		68	71	57	77	65	68	74	68	65	74	60	62	60	67
Consistency (ovarall)	5														
Positive scores (no.)		2	2	2	2	2	3	4	2	2	2	3	3	1	
Quality %		40	20	25	25	25	50	75	25	25	25	50	50	25	35
Three dimensions combined (ovarall)	63*														
Positive scores (no.)		44	44	40	46	45	47	47	46	44	46	44	42	42	
Quality %		69	69	63	73	71	74	74	73	69	73	69	66	66	70

*Calculated using totals for each dimension: 63=23+35+5 with the sub-dimension for ‘pre-model data analysis’ included as a portion of the ‘data’ dimension.