THE UNIVERSITY *of* EDINBURGH

Cost-effectiveness of Patient-Matched Pre- and On-Treatment Biomarkers in Cancer Therapy Response Prediction

Giovanni Tramonti

THE UNIVERSITY *of* EDINBURGH

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Edinburgh, August 2022

Abstract

BACKGROUND: The development of new biomarkers allows for the accurate prediction of breast cancer therapy response in the neoadjuvant setting. The implementation of these tools in routine NHS operations could have the potential for better clinical outcomes and a more cost-effective use of resources. **AIMS:** A decision-analytic modelling platform was created to evaluate the clinical effectiveness and cost-effectiveness of on-treatment biomarker-based predictive tools in a routine NHS implementation. The biomarker-based tool used in the simulation was EER4, developed at Edinburgh Institute of Genetics and Cancer. **METHODS:** Patient simulation models were constructed by integrating molecular biomarker data with clinical outcomes and healthcare cost data. The simulation benefits from the incorporation of NHS patient data and it features two complementary sub-models: a Discrete Event Simulation for the neoadjuvant setting and a Markov cohort model for the adjuvant setting. The results of the simulation compare the per-patient costs and health outcomes, expressed as Quality Adjusted Life Years (QALY), for each of the evaluated strategies. Moreover, this study includes a decision and budget impact analysis of OncotypeDX after its introduction in Edinburgh's Breast Clinic. **RESULTS:** Treatment-decision strategies that utilize EER4 are likely to be cost-effective. Specifically, EER4 in conjunction with PREDICT shows lower costs and marginally superior QALYs, compared to PREDICT alone or OncotypeDX with PREDICT. At a threshold of 20,000£/QALY, EER4 with PREDICT has an 86% probability of being a cost-effective alternative to the current standard of care. EER4 in conjunction with neoadjuvant Letrozole increases breast-conserving surgery rates, displacing radical mastectomy by 16%. The Probabilistic One-way Sensitivity Analysis shows that results are robust to the uncertainty of EER4 per-unit costs and clinical performance. The decision and budget impact analysis of OncotypeDX indicates that while this technology might reduce the number of chemotherapies administered, the unit cost is greater than any savings produced by chemotherapy displacement. **CONCLUSION:** The early cost-effectiveness analysis shows that these biomarker-based technologies are likely to be cost-effective. However, further research is needed to assess the clinical effectiveness of EER4. The simulation platform developed in this study has the potential for further evaluation of decision-making tools for other subtypes of breast malignancies.

Lay Summary

New predictive health technologies have the potential to improve the lives of breast cancer patient and make a more efficient use of NHS resources: by identifying early which patients will better respond to a specific treatment and which patients will not, the opportunity to select the most appropriate therapy can increase cancer survival rates and quality of life. If a new breast cancer test can deliver better health outcomes than the current standard of care, before it can be introduced to routine care, its cost-effectiveness must be assessed. That is, it is necessary to estimate the potentially higher costs of the new test relative to the better performance it can deliver compared to the standard of care.

This project uses “EER4” as an example of predictive test for breast cancer, currently under development at Edinburgh’s Institute of Genetics and Cancer (previously known as Institute of Genetics and Molecular Medicine). It can accurately predict how a patient will respond to pre-operative hormone therapy two weeks after diagnosis, granted that the patient has started treatment in the meantime. The response to treatment, as described by the test, is also predictive of long-term survival.

In order to estimate the cost-effectiveness of a new test, we simulate its implementation in clinical care and record the estimated costs and health benefits. Given the limited amount of data available, the simulation makes use of previous research and of Scottish patient clinical records.

The results of the simulation indicate that EER4 is likely a cost-effective use of NHS resources if implemented in routine operations. Furthermore, while there is still some uncertainty regarding the true costs and health benefits of EER4, the simulation platform suggests that it can likely deliver better health outcomes at a lower cost, compared to the standard of care.

Overall, the results of the analysis show that tests predicting response to treatment before surgery are likely to be beneficial for patients and the NHS, and that further research and data collection is needed to reduce the uncertainty surrounding the impact of introducing this kind of technology in routine care.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AML	Acute Myeloid Leukaemia
AQUA	Automated Quantitative Analysis
BCSS	Breast Cancer Specific Survival
BNF	British National Formulary
CEA	Cost-Effectiveness Analysis
CEAC	Cost-Effectiveness Acceptability Curve
CEAF	Cost-Effectiveness Acceptability Frontier
CES	Chemo-Endocrine Sensitivity
CHEERS	Consolidated Health Economic Evaluation Reporting Standards
CHF	Congestive Heart Failure
DES	Discrete Event Simulation
DFS	Disease-free Survival
EP	EndoPredict
ER	Oestrogen Receptor
EVPI	Expected Value of Perfect Information
FISH	Fluorescence in situ hybridization
FOV	Field of View
HER2	Human epidermal growth factor receptor 2
HR	Hazard Ratio
HR	Hormone Receptor
HRQoL	Health-related Quality of Life
HTA	Health Technology Assessment
IBC	Inflammatory Breast Cancer
ICER	Incremental Cost-Effectiveness Ratio
IHC	Immunohistochemistry
INMB	Incremental Net Monetary Benefit
NACT	Neoadjuvant Anti-Cancer Therapy
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NPI	Nottingham Prognostic Score
OS	Overall Survival
OTDX	OncotypeDX
PAM50	Prediction Analysis of Microarrays 50
PCNA	Proliferating Cell Nuclear Antigen
pCR	Pathological Complete Response
PEPI	Preoperative Endocrine Prognostic Index
PHS	Public Health Scotland
PLICS	Patient Level Information and Costing System
POSA	Probabilistic One-way Sensitivity Analysis
PR	Progesterone Receptor
QALY	Quality Adjusted Life Year
QPI	Quality Performance Indicators
RCT	Randomized Controlled Trial
RFS	Recurrence-free Survival
RWD	Real-World Data
RWE	Real-World Evidence
SACT	Systemic Anti-Cancer Therapy
SCAN	South East Scotland Cancer Network
SOC	Standard of Care
TIL	Tumour Infiltrating Lymphocytes
TNBC	Triple-Negative Breast Cancer

1. Introduction

1.1 Background

Breast cancer is the most common type of cancer in the United Kingdom, accounting for 15% of all new cancer cases and 31% of all new female cancer cases (1). Similarly to other tumour types, breast cancer is not a single, monolithic disease, but a diversified collection of tumour subtypes, where each subtype has its own specific characteristics, cellular composition, and clinical outcomes (2). Research and commercial interest in methods to better identify and define patients with high-risk disease has been growing over the last twenty years and several prognosticators and decision-making tools have entered current clinical practice of many healthcare services across the world. By comparing the genomic profiles of breast cancer specimens from patients with, and those without disease recurrence, studies based on multi-gene panels have led to the development of several prognostic assays that should enable more accurate predictions of clinical outcomes, compared to the use of conventional approaches. In contrast, the traditional clinical and pathological methods of estimating the probability of breast cancer recurrence use standard descriptors and physical characteristics, such as patient age, tumour size, histological features, and number of involved axillary lymph nodes. Oestrogen receptor (ER) and progesterone receptor (PR) expression are evaluated using immunohistochemistry (IHC), and both IHC and Fluorescence In Situ Hybridization (FISH) are used to determine HER2 (human epidermal growth factor 2) status (3). Expression of these biomarkers in tumour specimens obtained after surgery or biopsy sampling is used as both a prognostic and predictive marker to identify patients who are more likely to develop a recurrence and might benefit from endocrine therapy, chemotherapy, or anti-HER2-directed therapies (4, 5).

A wide range of clinical decision-making tools based on molecular markers have been proposed for selecting the most appropriate treatment for breast cancer patients. There is significant uncertainty on how decisions based upon these tests impact costs in routine NHS care pathways. Yet, the implementation of these diagnostic and prognostic tools in routine NHS operations could have the potential for better clinical outcomes and a more cost-effective use of resources. New biomarkers are under development that may predict benefit from specific drug

therapies for breast cancer. In particular, the development of these new biomarkers allows for the accurate prediction of breast cancer therapy response in the neoadjuvant setting. When used before surgery, effective drugs may shrink a cancer, increase the rate of successful surgery, and improve the chance of increasing recurrence-free survival (6).

The aim of this study is to investigate the current research landscape of neoadjuvant biomarkers and, with the use of decision-analytic modelling, assess the likely impact of introducing this type of technology within the NHS. This will take the form of a flexible decision-analytic modelling platform for the early economic evaluation of on-treatment biomarker-based predictive tools to compare the clinical effectiveness and cost-effectiveness in a potential routine NHS implementation. The specific example of biomarker-based tool used in the simulation is EER4, currently under development at the Edinburgh Institute of Genetics and Cancer (previously known as Institute of Genetics and Molecular Medicine), details of which can be found in Chapter 2. Moreover, the decision-analytic modelling platform will make use of NHS patient records, thus integrating evidence from previous literature and Real-World Data (RWD). Details on the use of RWD within the process of generating cost-effectiveness evidence can be found in Chapter 3.

1.2 Current Guidelines for the treatment of early breast cancer in the UK

The contemporary treatment of early breast cancer is complex, and it usually involves local therapy modalities, such as surgery and radiotherapy, in combination with systemic anticancer treatments and supportive measures, all of which can be delivered in diverse sequences. The following paragraphs offer a summary of the National Institute of Health and Care Excellence (NICE) guidelines for the treatment of patients diagnosed with early breast cancer (7). The guidelines were initially published in 2009, and have been updated several times since then, reflecting the outcomes of relevant technology appraisals, with the most recent update in 2018.

Surgery, with or without radiotherapy, remains the mainstay of early breast cancer treatment. Surgical treatment for breast cancer may consist of an excision of the tumour with surrounding normal breast tissue (breast conserving surgery, lumpectomy), or complete removal of the breast (mastectomy). In addition, surgery

can also include axillary node clearance, or alternatively sentinel lymph node biopsy. For patients that had a mastectomy and are at high risk of recurrence, chest wall radiotherapy is recommended.

Systemic anti-cancer therapy administered after surgery is referred to as adjuvant therapy, and depending on the tumour's molecular subtype and risk of recurrence, the choice of adjuvant systemic anti-cancer therapy can vary.

Patients with ER positive tumour are likely to receive adjuvant anti-oestrogen endocrine therapy, in the form of tamoxifen and/or aromatase inhibitors. Tamoxifen is indicated for ER-positive pre/perimenopausal patients, low risk patients, or patients with contraindications or severe reactions to aromatase inhibitors. Aromatase inhibitors are indicated as the first hormone therapy for post-menopausal patients at medium to high-risk of recurrence. If the patient is receiving adjuvant chemotherapy, endocrine therapy is usually deferred until all chemotherapy courses have been given. Patients with hormone receptor negative disease do not generally receive endocrine therapy. Adjuvant hormone therapy is normally indicated for a duration of at least 5 years, but can be extended up to 10 years, with either tamoxifen or an aromatase inhibitor depending on menopausal status. NICE's guidelines recommend treating all early ER-positive breast cancer patients with surgery and appropriate systemic therapy, rather than endocrine therapy alone, unless significant comorbidities are preventing surgery.

Adjuvant chemotherapy is recommended for patient at sufficient risk of recurrence, and administered as regimens that contain combinations of taxane and anthracycline-based chemotherapies. For patients with HER2 positive tumours, adjuvant trastuzumab (biological therapy) is recommended.

Systemic anti-cancer therapy delivered before surgery is referred to as neoadjuvant treatment. Neoadjuvant hormone therapy and neoadjuvant chemotherapy can be recommended as an option to reduce tumour size before surgery, especially in cases where the size reduction can result in breast conserving surgery rather than complete mastectomy. For HER2 positive tumours, neoadjuvant pertuzumab (biological therapy) can be offered as an option, in line with NICE's technology appraisal guidance on pertuzumab. For patients with triple-negative invasive breast

cancer, a neoadjuvant chemotherapy regimen that contains both platinum agents and anthracycline is recommended.

1.3 Predicting risk and treatment decisions

Section 1.2 often mentions that different treatment option can be offered depending on the patient's risk of recurrence. When it comes to estimating a tumour's risk of recurrence, two different broad strategies are available: prognostication tools that use clinical and pathological factors, or molecular tests that measure the level of expression of specific genes. The first approach generally stratifies risk based on variables that are routinely collected as part of the diagnostic and prognostic course of the patient; the most common tools for this approach are the Nottingham Prognostic Index, Adjuvant!, and PREDICT.

The Nottingham Prognostic Index (NPI) is a prognostic scoring system based on a large English cohort of patients with early breast cancer and it is based on tumour size, grade, and lymph node status (8, 9) In its first formulation, it categorized patients into three groups with significantly different survival; the index has been validated and updated since, with its most recent formulation allocating patients to one of six prognostic groups (10, 11).

Adjuvant! Online is an evidence-based online tool that enables the recurrence risk prediction for breast cancer patients using standardized clinical and pathological variables, based on data from large US databases (12). An individual's 10-year risk of disease recurrence can be used to estimate the overall risk reduction provided by adjuvant therapy, assuming a constant proportional risk reduction.

PREDICT is an online prognostication and treatment benefit tool for patients with early-stage breast cancer that can inform decisions on various combinations of therapy after breast cancer surgery (13). Similarly to Adjuvant! Online, PREDICT provides prognostic information as 10-year survival estimates by using clinical and pathological variables. The survival estimates can then be combined with the estimated therapy benefits of various adjuvant treatment combinations (chemotherapy, hormone therapy and trastuzumab). The first online version of the tool was published in 2010, followed by a series of updates that added new prognostic factors.

PREDICT, as an online tool, is limited in its capacity: producing survival estimates for only one patient at a time, it can make it difficult and time-consuming to generate survival estimates for large cohorts of patients. Given that this thesis makes use of a large cohort of breast cancer patients, and it uses PREDICT as a risk stratification tool extensively, a software package for R Programming Language (14) was created for the purpose of calculating PREDICT estimates for large numbers of patients with data recorded in tabular form. The package, “nhs.predict”, is available on the Comprehensive R Archive Network, and it is currently being used by the SCAN Audit team for quality performance indicators audit purposes (15). The functions within the package apply the very same algorithm used by PREDICT, granted that any given patient observation occupies a single row. The functions automatically check for missing data, apply the correct modifiers for unknown values, and attach the survival estimates to the corresponding observation.

As for the second approach to estimating risk of recurrence and treatment benefit, the strategy involves the pathological analysis and/or genomic sequencing of tumour tissue samples. This has led to the development of clinically useful assays to identify each intrinsic subtype and improve prognostication and risk stratification in patients with early-stage breast cancer, attracting considerable commercial and clinical interest. The majority of commercially available multigene assays focus on the evaluation of patients with hormone receptor-positive, HER2-negative disease, which includes OncotypeDX, Prosigna, MammaPrint, EndoPredict, and Immunohistochemistry 4 (IHC4).

OncotypeDX was one of the earliest clinically validated and commercially available molecular tests for patients with early-stage breast cancer (16). The assay measures the relative expression of 21 genes, 16 that are tumour-associated and five that serve as controls, with the results expressed as a computed quantitative recurrence score (RS) ranging 0–100, and it assumes that patients are given adjuvant endocrine therapy. Tumour samples with RS <18 are classified as low-risk disease, those with RS between 18 and 30 as intermediate-risk disease, and those with RS ≥31 as high-risk disease, where intermediate and high-risk patients might benefit from receiving adjuvant chemotherapy.

The Prediction Analysis of Microarrays 50 (PAM50) established a set of 50 gene transcripts used to identify intrinsic breast cancer subtypes with a high level of prognostic validity (17). The expression of 50 classifier genes and five control genes can be used to categorize breast cancer samples into one of four intrinsic subtypes (luminal A, luminal B, HER2-enriched, and basal-like). Subsequently, an algorithm (Prosigna) determines a risk-of-recurrence score ranging 0-100, and assigns risk categories (low, intermediate, or high) that reflect both the 10-year risk of distant recurrence of patients with early-stage hormone receptor-positive breast cancer and the potential benefit of adjuvant systemic anti-cancer therapy.

MammaPrint is a 70-gene assay that uses DNA microarray technology for quantification of gene expression associated with long-term survival (18). Mammaprint is intended for use on stage 2, hormone-positive HER2-negative tumours. The genes that form this signature are predominantly associated with tumour progression and metastasis, the analysis of which produces a binary result, high-risk or low-risk classification. Patients classified as high-risk might benefit from adjuvant chemotherapy.

EndoPredict involves the quantification of the expression of eight cancer-related and three control genes in order to calculate a risk score (19). This approach enables stratification of patients with ER-positive or HER2-negative early-stage breast cancer into a low or high 10-year recurrence risk group, when treated with adjuvant endocrine therapy alone. The results of an EndoPredict tests are usually combined with tumour size and nodal status to enable the calculation of a comprehensive risk score, EPclin.

The combination of ER, PR, HER2, and Ki-67 expression levels, as determined using IHC, have been integrated into the IHC4 score (20). IHC4 score was initially developed as a simplified comparator and surrogate for OncotypeDX, and was found to provide a similar level of information and accuracy as the 21-gene signature. However, the widespread use of IHC4 requires standardization of the IHC methods used to measure proliferative activity (Ki-67), a topic that is covered more in-depth in Chapter 2.

1.4 Pros and Cons of Neoadjuvant Therapy

Neoadjuvant systemic anti-cancer therapy has seen an increased use in clinical practice. With the same overall survival and recurrence-free survival rates are as post-operative therapy (21), neoadjuvant therapy has the advantage of potentially serving an in vivo sensitivity test for adjuvant therapies, while increasing the rate of breast conserving therapy (22). Moreover, it offers the opportunity of further studying tumour biology. These advantages often come at a cost: treatment before surgery modifies the stage and creates the potential for over-treatment. There is also the potential risk that residual intraductal components may be left behind after breast conserving surgery.

A growing body of evidence indicates that tumour response to neoadjuvant therapy may predict long-term outcomes of patients on adjuvant endocrine therapy (23-25), which argues in favour of its wider application in treating hormone receptor-positive patients. From the research perspective, neoadjuvant endocrine therapy provides a unique opportunity for studying therapy response and the development of novel therapies and predictive/prognostic markers.

1.5 Cost-Effectiveness Analysis and Analytical Models

This thesis aims to assess the impact of a novel biomarker on NHS routine practices by performing a cost-effectiveness analysis (CEA) using a two-stage decision analytic model. The following paragraphs provide a brief explanation on the nature of CEA and definitions for the types of models used in the analysis.

CEA is a quantitative approach that examines the costs and health outcomes of one or more interventions. By comparing two or more interventions, one of which is usually the standard of care (SOC), it estimates how much it would cost, at the margin, to gain a unit of a health outcome, such as a life-year gained or a death prevented, when considering a more effective and more costly option. The results of the analysis are usually presented as an Incremental Cost-Effectiveness Ratio (ICER): net cost divided by changes in health outcomes. Examples can include cost per case of disease prevented or cost per life-year gained. However, if the net costs of the intervention to be compared to the SOC are negative (which means a more

effective intervention is less costly), this results in a negative ICER, which can be interpreted as net cost savings.

In addition to ICERs, the results of a CEA can be presented as Net Monetary Benefits (NMB), which are calculated by subtracting the total costs of implementing the new intervention from the total monetary benefits that result from it. Monetary benefits can either include direct cost savings as measured in natural units (e.g. reduced hospitalizations or number of cancers detected early), as well as indirect benefits, such as improved quality of life. In particular, this study makes use of Quality-Adjusted Life Years (QALY) to calculate NMB: QALY is a measure of health outcomes that takes into account both the quantity and the quality of life, by multiplying the number of years gained by the quality of life of those years. Quality of life is expressed as a weight between 0 and 1, and it is typically measured with standardized tools and questionnaires. The monetary benefits calculated using QALYs are then obtained by multiplying the QALY of an intervention with a cost-effectiveness threshold, which measures the financial value of additional QALY and is expressed as a specified amount of currency per QALY gained. This operation effectively converts health outcomes into financial outcomes, conditional on the specified cost-effectiveness threshold. Calculating the difference between the NMB of the new intervention and the NMB of the SOC produces Incremental Net Monetary Benefit (INMB): depending on the specified cost-effectiveness threshold, positive INMB signify that the new intervention is cost-effective compared to the SOC, while negative INMB indicate that the new intervention should not be implemented.

As the novel biomarkers considered by this study are still under development, the CEA for the intervention can be considered an early cost-effectiveness study. This type of decision modelling refers to analyses that are conducted early in the technology's development process (26) and eventually could guide the performance of the technology under development. More specifically, the focus of early modelling is often on the commercial viability of new interventions, as to allow the developers of a new intervention to stop further development if the results indicate that the product is unlikely to become cost-effective. The primary question that early modelling is trying to answer is whether or not the development of an intervention should proceed, and the secondary question is, if the development is to go ahead, what the conditions for the new intervention to operate cost-effectively. As a

consequence, early CEAs are characterized by higher levels of uncertainty and by assumptions that define which minimum criteria the technology needs to meet for it to be cost-effective. Early economic modelling can yield numerous advantages for all the stakeholders involved in the development of a new intervention (27). The primary benefit for the developers of the new intervention is to estimate the range of values of key parameters within which the new technology would be deemed cost-effective. Moreover, early CEA can inform clinicians and final users of potential enhancements in the clinical pathway and treatment plan, particularly for patient subpopulations wherein the new technology is cost-effective. For payers, regardless of whether it is a universal health service or a private health insurance provider, early CEA can assist in the timely identification of promising new technologies that may result in quicker reimbursement decisions.

This study makes use of two analytical frameworks to achieve the results of the CEA: a semi-Markov model (Chapter 5), which simulates the adjuvant setting of a breast cancer cohort, and a discrete event simulation (DES), which accounts for the neoadjuvant setting and it integrates the results of the semi-Markov model (Chapter 6). The following definitions are summarised from Brennan et al 2006 (28).

The Markov model is probably the most common type of model used in economic evaluation of healthcare interventions. Markov models use disease states to represent all possible consequences of the intervention of interest. These states are mutually exclusive and exhaustive, so that every individual represented in the model can only occupy a single health state at any given time. For example, disease states that could be used to model cancer interventions could be: disease-free, recurrence, death. Individuals transition between disease states as their modelled condition changes over time. Time is modelled as discrete in the form of cycles, which typically last for months or years. Movements from one state to another in each cycle are modelled as transition probabilities. Time spent in a disease state for a single model cycle is associated with costs and health outcomes, which can be aggregated for all simulated patients in order to provide a summary of the cohort experience. This aggregate sum of costs and health outcomes can then be compared with the aggregate experience of a similar cohort that received a different treatment, such as the comparator. This comparison can then result in an ICER, thus providing an estimate of cost-effectiveness of the relevant comparators.

Markov models typically operate under the Markovian assumption: the probability of moving to a subsequent disease state is independent of the time spent in previous states. Moreover, the usual formulation assumes that the transition probabilities are constant over time. To improve the generalizability of models, whenever the Markovian assumption might be at odds with the progression of a disease, there are several approaches available: for example, changing transition matrices as time progresses, or redefining states based on certain attributes. Models that adopt these modifications are referred to as semi-Markov.

Discrete event simulation (DES) is a modelling technique in which individual patient experience is simulated over time, the simulation is driven by events occurring to the patient and the consequences of such events are tracked and summarised. Events occur at varying times, rather than during cycles of fixed length, and event likelihoods are determined by individual patient characteristics, which are recorded at baseline and may be updated as the patient experience accumulates. The state of the modelled system includes the current entities (patients), their attributes, and a list of events that can occur either at the current simulation time or that are scheduled to occur in the future. The aggregate results, which include the number and type of events, attributes, and initial/final characteristics, are then associated with costs and health outcomes. DES is likely to be useful for modelling complex pathways with many possible types of events, or situations where the patient's history may impact on future events (for example, the long-term effects of neoadjuvant therapy on breast cancer patients).

2. A Review of Current Molecular Markers Predicting Response to Neoadjuvant Therapy

2.1 Introduction and Review Criteria

Several molecular-based predictive tests are available today for guiding post-operative treatment decision in breast cancer. Yet, a gap undoubtedly exists for technologies predicting treatment response in the neoadjuvant setting: with the increased tendency to treat pre-operatively (29), clinicians are now facing the problem of identifying which patients will benefit the most from each type of treatment. The need of being able to predict response to neoadjuvant therapy becomes more pressing when considering that the benefits of treating patients preoperatively are not limited to the short-term only but are reflective of therapy response after surgery (30). With this consideration, the purpose of the review is to offer the current picture of the research landscape for breast cancer neoadjuvant predictive tests, and identify which technologies are the most likely to fill the clinical need in the near future. While the overarching objective of this thesis is to develop a simulation platform for the economic appraisal of a select few examples of biomarker-based predictive technology, it is necessary to explore the differences and similarities in the development process of other molecular markers as to observe in which direction the research is headed and what characteristics make a predictive test a potentially successful candidate for clinical use. The prognostic and predictive models described in this review make use of different strategies and molecular markers that are associated with pathological response to neoadjuvant treatment.

The information for this review was compiled by searching the PubMed and MEDLINE databases for articles published between January 2013 and May 2019, as to identify the most recent molecular markers. The primary search terms used featured “breast cancer” in association with, but not limited to, “neoadjuvant”, “preoperative”, “molecular marker”, “gene-signature”. Only articles published in English and with full text available were considered. When possible, primary sources have been quoted. Full articles were obtained, and references were checked for additional material when appropriate. The predictive tests described in the articles were grouped together in categories based on the type of molecular signature used in the test. The technologies were grouped together as follows: Ki-67 measurement,

Immune Signatures, Established Adjuvant assays in neoadjuvant context, Pre- and on-treatment, Mutations, Other Signatures. The review concludes with a mention of the obstacles during the translational process of these technologies into clinical care and how early Health Technology Assessment could overcome them.

2.2 Assessment of Proliferation

Proliferation of cancer cells is a fundamental defining hallmark, closely associated with tumour grade. It can be measured by several markers, but the most widely used in breast cancer is Ki-67, a nuclear protein present in all active stages of the cell cycle except for G0(quiescent phase) (31). Proliferation is arguably the most important feature for classifying oestrogen receptor positive (ER+) early-stage breast cancer (32), which are now often classified into Luminal A and Luminal B subtypes, with the latter being more proliferative and associated with poorer prognosis (33). Although the Luminal A/B subtypes arose from gene-expression analysis, a threshold value for Ki-67 (14%) has been proposed as a surrogate for the distinction between the two (34). Moreover, measurements of Ki-67 Labelling Index have shown promising results with respect to the prediction of clinical response to chemotherapy and survival (35). The systematic clinical implementation of Ki-67 scoring, both as a surrogate for Luminal status and as a predictive biomarker, has been slowed down due to the controversial lack of reproducibility in counting the percentage of cells staining positive in immunohistochemistry (36). Overall, proliferation is also higher in luminal than basal and ER- subtypes.

The measurement of Ki-67 is performed by counting the percentage of stained positive (tumour) cells in immunohistochemistry (IHC) (37), and then a prediction is formulated based on cut-off values. On one hand, the method for counting the percentage of positive nuclei varies between averaging the value of the whole area of the IHC and choosing “hotspot” areas (where positive nuclei are clustered) then calculating the percentage of positive nuclei in the area of interest (37). On the other hand, the threshold values for predicting chemotherapy response can vary between 1% and 29%, according to the literature (38). This variability in the methods makes it difficult to replicate and validate predictive tools based on Ki-67.

The study by Brown et al. (2013) (39) aimed at objectively measuring the level of Ki-67 proliferation by removing the inherent subjectivity of a pathologist, by using quantitative immunofluorescent automated quantitative analysis (AQUA), and by using the scores of the average and maximum fields of view (FOVs) in the range from 5 to 115. The study showed a direct correlation between Ki-67 expression and pathological complete response (pCR) to neoadjuvant chemotherapy, using AQUA scores. Analysis by means of AQUA eliminates the need for a cut-off value and removes the subjectivity of a pathologist's judgement: it predicts the likelihood of achieving pCR using continuous variables. As such, the study did not recommend a cut-off value for guiding therapy decision. However, this method and AQUA appear not to have been adopted by subsequent studies on Ki-67 as predictive marker for pCR: this could be potentially due to the wide availability of standard IHC performed by a pathologist and the fact that the discussion around cut-off values is far from settled.

2.3 Relative or On-Treatment Assessment of Proliferation

The analysis of Ki-67 expression has recently shifted its development focus from using a baseline measurement for prediction to measuring the absolute and relative differences in proliferation before and after neoadjuvant therapy. These differences appear to be an important predictor of pCR, of Overall Survival (OS) and of early metastasis (40). The results from Yoshioka et al. (2013) have shown that patients whose tumours contain high Ki67 expression effectively responded to neoadjuvant chemotherapy, but that high expression of Ki67 in residual tumours was strongly correlated with poor overall survival, regardless of subtype. The evaluation of changes in Ki-67 in response to neoadjuvant chemotherapy can also significantly predict early metastasis; the results from Tokuda et al. (2017) indicated that increased Ki-67 expression is related to the early development of metastasis (41). This may allow the identification of this category of patients early, so that more alternative treatments can be used earlier.

While the expression of Ki-67 in Hormone-receptor positive (HR+) breast cancer patients has been extensively investigated, there is actually little evidence for the association between Ki-67 proliferation and clinical outcome within cohorts of HR

negative patients (42). According to the results from Tan et al. (2014), the pattern is similar to HR+ tumours: tumours that contained high Ki67 expression pre-treatment effectively responded to neoadjuvant chemotherapy, high Ki67 expression post-treatment was associated with poor disease-free survival (DFS). The results suggest that Ki-67 level is a predictive marker for pCR, but the expression levels had no correlation with Overall Survival. The Ki-67 level of proliferation in HR-negative patients can improve the assessment of pCR after neoadjuvant chemotherapy, while also being and independent prognosticator of DFS in residual tumours.

The proliferation index for Ki-67 can potentially be used to guide chemotherapy treatment decision if coupled with neoadjuvant endocrine therapy (43). In the recent ACOSOG Z1031 trial, specifically in cohort B, ER+ breast cancer patients were initially assigned to neoadjuvant aromatase inhibitors. Ki-67 LI was measured at 2 or 4 weeks after commencement of treatment. If the Index was greater than 10%, the patient would be switched to neoadjuvant chemotherapy. The failure to suppress Ki-67 proliferation by AI therapy would allow the identification of patients with AI-resistant tumours. While previous studies highlighted a direct correlation between AI-resistance and chemotherapy efficacy, Ellis et al. (2017) found that ER+ AI-resistant tumours were less responsive than expected to chemotherapy, given the low pCR rate of those who switched.

Measurements of Ki-67 expression can also be used to compute the PEPI (Preoperative Endocrine Prognostic Index) score, along with tumour size, node status and ER status (44). The PEPI score of a patient is associated with the risk of relapse, specifically the three score groups (0, 1-3, ≥4) are respectively associated with 10%, 23% and 48% risk of relapse. Ellis et al (2017), using the data from Z1031B sought to validate the estimate for risk of relapse associated with PEPI=0 and found that patients with PEPI=0 had relapse risk over 5 years of 3.6% without chemotherapy, supporting the idea of adjuvant endocrine monotherapy in these patients.

The upper threshold of Ki-67 measurement for PEPI=0 is 2.7%, which was further confirmed and validated by Goncalves and al. (2017), using the data of Z1031 (45). The analysis led to the development of a clinical trial assay in the context of neoadjuvant endocrine therapy that uses an efficient and reproducible Ki-67 scoring

system. In contrast with previous threshold estimates, this assay uses a Ki67 LI 10% cut-off point for the genomic surrogate identification of Luminal B tumours. The methods described in the study may in the future allow for the identification of ER+ patients for whom adjuvant chemotherapy can be safely spared.

While the overall findings of these studies show promising results, the variability of the methods for Ki67 measurements remain problematic. Recent efforts toward standardisation continue, including validation and reproducibility studies aimed at increasing the scoring concordance and reducing the analytical variability between laboratories (46, 47). Moreover, there exist potential alternatives to Ki67 as a proliferation marker. For example, PCNA (proliferating cell nuclear antigen) and MCM (minichromosome maintenance) proteins have become standard markers of proliferation, used to assess the growth cells (48), while the measurements of AURKA (aurora kinase A), coupled with ER and HER2 (thus a three-gene signature), can robustly classify Luminal-like breast cancers.

Recent evidence from the POETIC trial highlighted the use of pre-operative aromatase inhibitors coupled with paired measurements of Ki67 (at baseline and two weeks on treatment) to predict response to adjuvant therapy, and therefore individually select the most appropriate post-operative course of treatment (49). While the trial has not shown any improvement in treatment outcomes due to neoadjuvant aromatase inhibitors, it has shown that this treatment can be safely utilized to aid in the selection of appropriate adjuvant therapy. In general, patients with low Ki67 at baseline or low Ki67 at two weeks on-treatment exhibit positive responses to adjuvant standard endocrine therapy, while individuals with persistently high levels of Ki67 at two-week on-treatment may benefit from supplementary adjuvant therapy.

2.4 Immune Signatures and Tumour Infiltrating Lymphocytes

Within the context of the development of molecular markers, Tumour Infiltrating Lymphocytes (TILs) and, more generally, Immune Signatures have been studied due to the association between cancer outcomes and immune activity in the patients (50). TILs are white blood cells that migrate to the tumour site from the bloodstream, and they are involved in the process of killing tumour cells (51); their presence and

activity within the tumour can change according to the cancer progression and according to the treatment applied (52). With the same principles, immune activity and immune-related gene-expression changes according to how cancer develops, and which therapy the patient undergoes. These changes can be monitored and measured to produce models that, coupled with neoadjuvant treatments, can predict the pathological response of patients to therapy.

In the context of triple-negative breast cancer, the measurement of TILs levels and status are significantly associated with Relapse-free survival (RFS) and Breast Cancer Specific Survival (BCSS). In particular, $CD8^+$ TILs and $FOXP3^+$ TILs show the most prognostic significance when measured before and after neoadjuvant chemotherapy (53). In the study by Miyashita et al. (2015), triple-negative breast cancer patients with high $CD8^+$ TIL levels or a high $CD8^+/FOXP3^+$ ratio in residual tumours after NACT had better RFS and BCSS, with the former measure being a highly significant predictor of BCSS. Greater rates of change in the TILs ratio were also associated with better RFS and BCSS. Moreover, the parameters accurately predict improved prognosis in TNBC patients with non-pCR following NACT.

Yet, the evaluation of TILs levels is based on histopathological measurements, which intrinsically possess limited accuracy and reproducibility (54). Thus, an alternative proxy marker has been proposed, which would retain the prognostic value of TILs while increasing accuracy: the signature of TILs methylation (MeTILs). According to the results of Jeschke et al. (2017), MeTIL signature predict survival and response to chemotherapy in BC better than histopathological assessment of TILs markers (55). While the profiling of DNA methylation and lymphocyte methylation used in the study requires specialist equipment and prohibitive amounts of resources for a possible clinical application, MeTIL markers can be determined economically by bisulphite pyrosequencing of small amounts of DNA from FFPE samples, which would make clinical applications of this method more feasible.

In addition to TILs, other immune signatures include those using immune-related gene expression data, which have been used to predict pCR to neoadjuvant chemotherapy (56). IRSN-23 is a 23-gene signature that can predict pathological complete response in patients treated with neoadjuvant chemotherapy. The assay is capable of classifying patients as either genomically predicted responders or non-

responders, and it can predict pCR independently of intrinsic subtypes and chemotherapy regimens. Yet the clinical value of IRSN23 appears to be limited because its diagnostic accuracy is not sufficiently high: 79% and 75% in training and validation datasets, 70% in the external validation public datasets. Evidence around the clinical utility of IRSN-23 is still insufficient, and prospective studies are needed in this respect.

In the case of Inflammatory Breast Cancer, no molecular marker has been identified as being able to predict response to neoadjuvant chemotherapy or predict survival, until the work of Bertucci et al. (2013) (57). This study managed to demonstrate that even in the case of inflammatory breast cancer, immune-related processes can be associated with response to neoadjuvant chemotherapy. By examining the correlation with predictive and prognostic gene expression signatures published in nIBC-GES the study found that three of the five nIBC-GES tested prognostic gene expression signatures and two tested nIBC-GES gene-expressed signatures discriminated between IBC with and without pCR. Moreover, the researchers were able to identify a 107-gene signature enriched for immunity-related genes that was able to separate IBC patients into responders and non-responders to neoadjuvant chemotherapy. The model for the 107-gene signature was validated internally, reaching 75% accuracy, and then externally validated, with 81% accuracy. Unfortunately, the study was not able to find a robust signature that would associate neoadjuvant chemotherapy response in Inflammatory Breast Cancer and distant metastasis-free survival.

The review process highlighted a lack of pre-operative molecular markers in HER2-positive breast cancer. Given the highly heterogeneous nature of this cancer subtype, reliable markers for therapy response have been elusive (58). Nevertheless, research efforts aimed at investigating the association between response to therapy and HER-2 positivity have continued, mainly focusing towards immune response and TILs. A secondary analysis of the neoALTTO trial highlighted that the presence of TILs at diagnosis is an independent prognostic marker for pCR and event-free survival in HER2-positive early breast cancer that has been treated with neoadjuvant anti-HER2 agents and chemotherapy (59). Furthermore, a meta-analysis by Denkert et al. (2018) has shown that increased TIL concentration can predict response to neoadjuvant chemotherapy in all breast cancer molecular

subtypes, and was also associated with survival benefit in HER2-positive cancers and TNBC (60).The research by Varadan et al. (2016) (61) suggests that intrinsic subtype and immune cell infiltration may predict response for trastuzumab-based therapy (62). As a significant proportion of HER2-positive patients do not respond to this therapy, identifying responders and non-responders early on is vital to achieve better clinical outcomes. Moreover, from a health economics perspective the high costs of HER2 antigen-based therapy (compared to endocrine therapy and chemotherapy), create a compelling argument towards the development of technologies able to correctly identify patients benefitting the most from this type of treatment. Recent NICE guidance regarding pertuzumab in combination with trastuzumab and docetaxel reported a 0% probability that the treatment is cost-effective at a threshold of 30,000£/QALY (63).

A subsequent analysis by Varadan et al. (2016) (64) applies this method for the first time in vivo. Their study observed the changes in immune signatures after exposure to preoperative trastuzumab and the achievement of pCR following trastuzumab paired with chemotherapy in HER2 positive breast cancer. The results have shown that immune activity (summarised as Immune Index) had significantly increased after exposure to trastuzumab, and this increase was predictive of response in HER2 positive tumours. This suggests that a brief dose of preoperative trastuzumab has the potential to uncover clinically useful immune signatures for predicting response to biological therapy. The early identification of responders and non-responders, and thus more appropriate treatment decisions, should help achieving better clinical outcomes. Although not within the scope of this review's initial timeframe, recent advancements in research have yielded promising methods to predict response in HER2-positive subtypes. The development and validation of HER2DX by Prat et al. (2022) demonstrates that by combining data on tumour features, pathology features, and immune features in a single assay, it is possible to predict both long-term survival and the likelihood of achieving pCR (65). Specifically, HER2DX incorporates tumour size, nodal staging, and 4 gene expression signatures tracking immune infiltration, tumour cell proliferation, luminal differentiation, and the expression of the HER2 amplicon, into a single score. On top of significantly predicting long-term outcomes, the study suggests that HER2DX risk score might be able to identify patients with early-stage, HER2-positive breast cancer who do not need additional

HER2-targeted therapies, such as pertuzumab, trastuzumab emtansine or neratinib, due to an already favourable survival outcomes with chemotherapy and trastuzumab.

2.5 Moving Established Adjuvant Assays into the Neoadjuvant Setting?

Established adjuvant assays have validated prognostic power, but their efficacy has rarely been tested in the less common neoadjuvant setting. Evidence for their ability to predict benefit from systemic therapies in the adjuvant setting is weak, but the neoadjuvant setting is an opportunity to do better. As noted in the technologies explained before, oestrogen receptor positive breast cancer shows heterogeneous characteristics that determine different sensitivities to treatment.

Bertucci et al. (2014) used EndoPredict score classification to predict pathological complete response to neoadjuvant chemotherapy for ER+/HER2- cancer (66). Specifically, the scoring system classifies patients as either low or high-risk. EndoPredict low-risk patients can be potentially treated with post-operative endocrine therapy alone, while EndoPredict high-risk patients may necessitate adjuvant chemotherapy in addition to endocrine therapy. The objective of this study was to determine the predictive ability of EndoPredict for pathological complete response to neoadjuvant chemotherapy, while also determining whether EP high-risk patients are more or less sensitive to chemotherapy than low-risk patients.

The findings indicate that the EndoPredict classification was associated with a pCR rate of 7% in the low-risk group and 17% in the high-risk group. In the high-risk group, many of the upregulated genes of the expression profile were involved in cell proliferation, whereas in the low-risk group other upregulated genes were involved in ER signalling. Despite the high chemosensitivity, the high-risk group was associated with worse disease-free survival. Consistent with previous findings, the study concluded that ER+ tumours in the high-risk group benefit the most from chemotherapy, corroborating the justification to avoid treating low-risk patients (which are also low chemo-sensitive) with chemotherapy.

Prat et al. (2016) derived a PAM50-based chemoendocrine score (CES) to predict the sensitivity to chemotherapy and endocrine therapy in $HR^+/HER2^-$ breast cancer (67). The predictive ability of the score was assessed on 4 independent neoadjuvant datasets. Based on the results of the test, patients can be classified in three subgroups:

- CES-E(endocrine sensitive), patients that benefit the most from endocrine therapy
- CES-U(uncertain), for which it is unclear which treatment is likely to achieve pCR
- CES-C(chemotherapy sensitive), patients that benefit the most from chemotherapy

Thus, the results of the study were the first to confirm, in a randomized setting, an inverse relationship between endocrine and chemotherapy sensitivity, this is summarised in Figure 1: risk of recurrence is plotted against relative chemoendocrine sensitivity, displaying the intrinsic cancer subtypes (adapted from Prat et al. 2016). In terms of clinical application of the chemoendocrine score, while there is minimal benefit for $HR^+/HER2^-$ in the low and high-risk categories, there is potential gain for patients in the intermediate score, as it would better guide treatment decisions. While this neoadjuvant application of a PAM-50 based score is capable of estimating chemoendocrine sensitivity beyond intrinsic subtype of cancer, there are several limitations to the study, including its retrospective nature, the use of the research version of PAM50, and the lack of association between CES and survival data in the adjuvant setting.

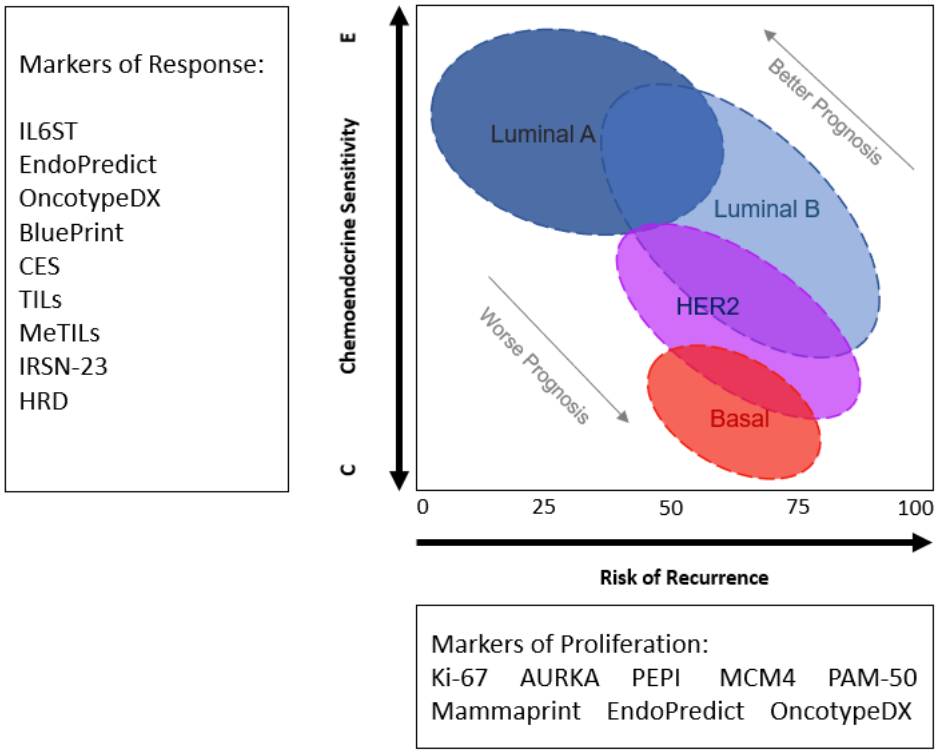


Figure 1: Summary of the relationship between intrinsic molecular subtype and gene signatures or other markers used for clinical decision-making

2.6 Measurements of Pre- and On-Treatment Molecular Markers

The early identification of patients not responding to neoadjuvant treatment would make it possible to switch to an alternative treatment strategy that could be more effective. Thus, identifying appropriate molecular markers and measuring their level at baseline and on-treatment could prove a successful strategy for the early prediction of response to neoadjuvant therapy.

With regards to ER-positive breast cancers, a recent 4-gene signature is able to predict response to neoadjuvant endocrine therapy (68): the EER4 test measures the gene expression level in tumour samples from two matched biopsies, one pre-treatment, the other 2-weeks on treatment. The model developed by Turnbull et al. (2015) predicts clinical response by measuring the level of 2 genes pre-treatment (IL6ST associated with immune signalling, NGFRAP1 with apoptosis), and the level of proliferation of 2 genes on-treatment (ASPM, MCM4).

Using the measurement from these gene-signatures, the model is able to predict clinical response in patients treated with Aromatase Inhibitors, by separating patients into three categories, defined by the change in tumour size: quick stable response, slow responders, non-responders. The 4-gene signature model used for this test has 96% accuracy in correctly identifying responders and non-responders. An independent validation of the model on a dataset with a similar cohort was 91% accurate. More than 80% of tumours had changed in volume on-treatment. The molecular markers' predictions are associated with long-term recurrence free survival and overall survival (68)

Magbanua et al. (2015) observed changes in pre- and on- treatment biomarkers levels, showing the potential of on-treatment measurements (69). This study collected expression data at baseline, during treatment (between 24 and 96 hours from the start of neoadjuvant chemotherapy) and at surgery. Expression data was compared between baseline and treatment and between baseline and surgery. The subtypes were assigned using PAM50 gene signature and the differences in early gene expression changes between pre- and on-treatment between responders and non-responders were evaluated. According to the results of the study, significant differences of expression profiles were identified between baseline and on-treatment, and baseline and surgery. During treatment, there was a significant downregulation of genes associated with proliferation and immune-response; moreover, gene expression changes between baseline and on-treatment for cell cycle inhibitors were correlated with worse response. Furthermore, the positive change between baseline and surgery in interferon signalling and high expression of proliferation genes in residual tumours were associated with reduced relapse-free survival. The study concludes that the serial gene expression analysis identified pathways associated with immune-response and proliferation that can predict response to neoadjuvant chemotherapy and recurrence.

Bownes et al. (2019) identified a novel single-gene on-treatment marker that predicts response to neoadjuvant chemotherapy (70). The predictive marker, AAGAB, showed a testing accuracy of 100% (data from Edinburgh NEO trial) by using a biopsy sample mid-chemo, and a validation accuracy of 78% (validated with data from the I-SPY 1 trial). AAGAB is also predictive of long-term survival in both datasets, as defined by treatment response. The single-gene marker outperforms, in

terms of predictive accuracy, established predictive assays (PAM50 and Mammaprint) tested on-treatment in the neoadjuvant setting. A recent study by Selli et al. (2019), focusing on extended neoadjuvant treatment to observe differences between acquired resistance and tumour dormancy, suggests that promising markers for response prediction may be evident relatively early on treatment (71). A possible timeline for the use of on-treatment biomarkers is shown in Figure 2.

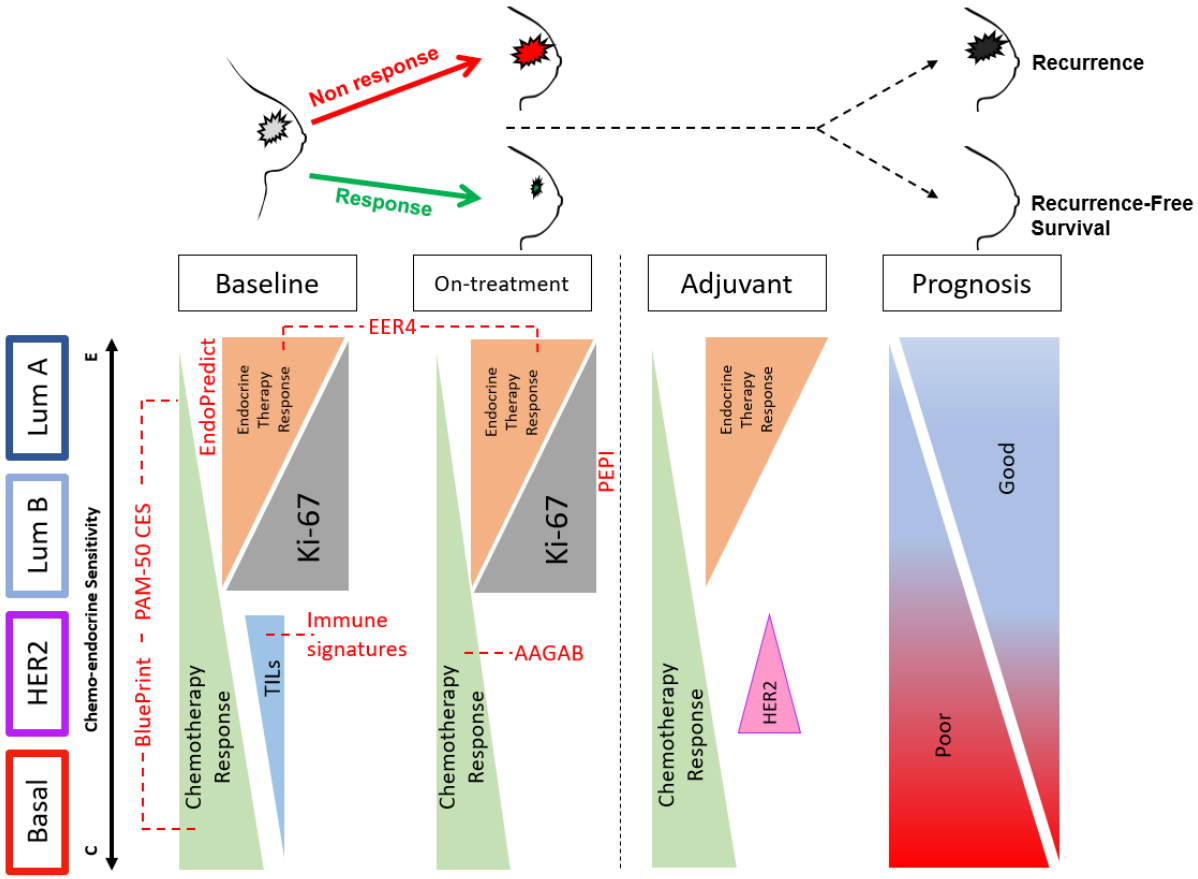


Figure 2: Timeline of a breast tumour from Diagnosis to Prognosis, for the four intrinsic subtypes

2.7 Mutations Do Not Seem to Be the Answer for Now

There have been attempts to analyse mutations as biomarkers to identify responders and non-responders in the neoadjuvant setting. Unfortunately, analysis of mutations does not seem to provide insights for the early identification of non-responders (72). In the study by Lips et al. (2015), deep sequencing of tumour DNA prior to neoadjuvant chemotherapy in triple-negative breast cancer patients did not highlight any mutation that can reliably predict response to treatment. The mutations observed

in the study were diverse, and few recurrent mutations were detected. Mutation rates were similar in responders and non-responders and no recurrent mutations were associated with chemotherapy response or relapse. Perhaps in the future, after having characterised rarer mutations, it might be possible to use this information to guide treatment decisions.

Further efforts are being made to better understand the significance and the mechanisms behind mutations: a recent analysis by Bertucci et al. (2019) investigated the distribution of mutational signatures to identify which processes might be driving late tumour progression (73), as metastatic breast cancer is generally more genetically complex than early breast cancer. After comparing the mutational burden and clonal diversity between the two disease settings, the analysis identified genomic alterations that are enriched in advanced breast cancer, and the results suggest that activation of mutational processes could contribute to disease aggressiveness and genome evolution.

A study by Jiang et al. (2019) regarding triple-negative breast cancer mutations shows that the copy number or the mutational cluster are not able to predict recurrence-free survival (74). The findings further confirm that using mutations is not viable for predicting response and guiding treatment, as of now and with current knowledge.

2.8 Other Signatures

This section of the review features a brief description of the predictive technologies using molecular markers that do not otherwise fit in the previous categories.

A study by Whitworth et al. (2014) sought to validate the chemosensitivity prediction ability of BluePrint, when coupled with MammaPrint. BluePrint is an 80-gene panel that classifies breast cancer patients into four molecular subtypes (Luminal A, Luminal B, HER2 and Basal) (75). The study compared the accuracy of BluePrint, coupled with MammaPrint, with conventional IHC and FISH subtyping and measured the reassignment rate; that is, the number of patients that were reclassified by BluePrint from one subtype to another after being initially classified by IHC and FISH. In the study, chemosensitivity is defined by pCR, while endocrine sensitivity is defined by partial response. The results indicate that BluePrint reclassifies 22% of

tumours by assigning more responsive patients to HER2-positive and Basal subtypes. It also reassigns less responsive patients to the Luminal category. The findings suggest that compared to more conventional IHC and FISH, BluePrint is able to identify more accurately the molecular subtype and patients likely to respond to neoadjuvant chemotherapy.

In the context of BRCA mutated breast cancers, and in sporadic cases, triple negative breast cancers, the tumours have DNA repairs defect and are sensitive to DNA-damaging therapeutics (76). The genomic instability of these tumours can be measured with three independent scores, defines as: loss of heterozygosity (LOH) (77), telomeric allelic imbalance (TAI) (78) and large-scale state transitions (LST) (79). The study by Telli et al. (2016) uses un unweighted sum of these three measures to compile a homologous recombination deficiency score (HRD) to predict response to neoadjuvant platinum-containing chemotherapy in triple negative breast cancer (80). The HRD score was able to predict residual cancer burden score and pathological complete response with high significance. The HRD score remained a significant predictor of residual cancer burden even when the model was adjusted for clinical variables, including in the multivariate model. The findings thus suggest that the HRD score is able to identify the triple negative breast cancers, including those in patients without germline BRCA mutations, more likely to respond to neoadjuvant chemotherapy that contains platinum.

2.9 Clinical Utility and Health Technology Assessment

As its primary objective, this review provided a snapshot of the landscape of potential future molecular markers that could play a role in clinical decision making in the neoadjuvant setting. When presenting the technologies, the available evidence of analytical and clinical validity was also discussed. Yet, even after the analytical and clinical validity of a biomarker have been demonstrated, there needs to be evidence of sufficient clinical utility before the technology can be used to guide decision-making.

Ultimately, the decision to adopt a new marker in clinical routine may rest on the outcomes of the Health Technology Assessment (HTA) process. As noted in Buchanan et al. (2013) (81), the main challenges of performing an economic

evaluation on genomic technologies lie in outcomes measurements and the differences in guideline standardization. While this analysis is focused on genomic technologies, the caveats surrounding the HTA process still remain valid for the molecular markers presented in this review. Furthermore, effective economic evaluation necessitates a great volume of data, which could prove difficult to collect, especially in the context of early technologies. Additionally, translating the development of a novel marker into a clinically available decision-making tool goes well beyond demonstrating a superior outcome relative to comparators: there are several organizational aspects that need to be resolved before a technology can be approved for routine use (mainly in the form of guidelines and standardised recommendations). Miquel-Cases et al. (2017) (82) suggested that these novel technologies should undergo HTA during the early stages of research and development, as performing an economic evaluation at those points may facilitate the translational process of potentially successful technologies, while screening out those that are unlikely to provide added value in a clinical context. Even in the case of established adjuvant assays, which are generally considered to be cost-effective (83), further analysis might be required for optimizing their use and role in clinical care. As underlined in Hall et al. (2017) (84) the value of information analysis shows the importance of HTA modelling as a guide for further trials for improving research efficiency, and thus better and faster implementation (or discarding) of a predictive technology.

This study conducted the initial scoping work for a literature review aimed at identifying recent methodologies that are tailored to early CEA of medical tests, with the intention of collecting a limited number of case studies to present in a qualitative manner. To achieve this goal, existing economic models published between 2015 and 2018 were examined, and the PubMed database was utilized to retrieve relevant articles. The search strategy incorporated MeSH terms for early health technology assessment, decision-analytic models, and medical tests. A total of 97 articles with full-text available in English were initially identified, and after filtering out articles that were not relevant, the total number of potentially eligible articles was fifteen (85-99). However, the review was stopped at this point due to two primary reasons. Firstly, the included articles exhibited heterogeneity in both methods and the reporting of methods, which would have shifted the review's focus towards the investigation of

methodological issues in early economic modelling. This deviation was not consistent with the study's overall aim, so further efforts towards the review ceased. Secondly, continuing the review would have resulted in replicating similar works, leading to the risk of redundancy, such as part of the work from IJzerman et al (2017) and the work from Frempong et al (2018) (100,101). The conclusion of their reviews, coupled with the commentary surrounding the topic of early economic modelling and its methodological/reporting issues (27, 102), shows an overall consensus: lack of standardised guidance on the methodology to be used for early modelling, lack of transparency and repeatability of the studies, lack of consistent reporting of methodology.

2.10 Discussion

In both the neoadjuvant and the adjuvant setting, clinical and pathological variables still remain the standard of care for guiding decisions on systemic therapy in breast cancer. In routine care, their use in deciding the course of treatment is inexpensive, but there are still many patients who fail to respond, or whose cancer recurs after treatment directed by these established factors. Chemotherapy overtreatment or unnecessary/ineffective therapy for HR+ patients non-responsive to endocrine therapy currently presents a significant morbidity burden for patients. In recent years, adjuvant assays were introduced in a limited role in routine care; adoption has presented health authorities with difficult decisions in the face of uncertain evidence (85); a gap undoubtedly still exists for new predictive technologies that can guide therapy decisions. Given the increased tendency to treat breast cancer patients before surgery in recent years, this gap could be filled by novel predictive markers in the neoadjuvant setting.

This review highlighted some of the advantages and flaws associated with each type of technology, and several potential issues still remain before any implementation in standard care can be considered. For example, while measurements of proliferation based on Ki-67 are routinely used and performed on IHC, this particular marker suffers from a lack of consensus on the optimal cut-off point in the literature. Even though established adjuvant assays used in the preoperative setting might be reliable, their cost might outweigh the clinical benefits of their predictions, rendering them unfit for filling the gap. Moreover, other gene panels aimed at predicting

therapy and immune response do not reach a level of accuracy suitable for clinical applications.

Given a high performance in terms of predictive capability, the technologies most likely to be implemented in routine care are those that can be translated into simpler and cheaper techniques without loss of accuracy (such as IHC or bisulphite pyrosequencing). Most of the technologies outlined in this review are still in the early stages of development, and a recurrent theme across the studies is the need for prospective validation before clinical applications can be considered.

3. Using Real-World Data to Inform an Economic Evaluation

3.1 Introduction

To avoid causing more harm than good, economic evaluations of new healthcare technologies have to rely on a robust evidence base. Traditionally, health technology assessment is informed by Randomised Controlled Trial (RCT) data. Given their strong internal validity and minimisation of bias, RCTs are considered the gold-standard for evidence aimed at demonstrating medical safety and efficacy. Additionally, in terms of economic evaluation, they can have the advantage of measuring benefits and costs within the same setting. However, the high-quality data from RCTs comes at the cost of reduced generalizability of results: strict selection criteria produce highly selected participants and environments, which, on top of high protocol adherence and monitoring, might not reflect current clinical practice. Moreover, the increasing administrative and operational costs of RCTs (86), combined with potentially low cross-country and external validity, have generated a growing interest for more representative data source alternatives (87).

Real World Data (RWD), routinely collected patient records in particular, can offer a less expensive and more readily available data source, but with several caveats. Routinely collected clinical data possess the potential advantage of better reflecting the patient population and current clinical practices, thus potentially improving the generalisability of results (88). The administrative and operational costs are reduced, as the data is collected regardless, is readily accessible, and offers long follow-up times. Additionally, RWD can typically provide large study samples, and thus the opportunity to study less common adverse effects more easily (89). When such data are analysed, the information produced can be referred to as real-world evidence (RWE). The NICE Decision Support Unit currently defines RWD and RWE as:

“RWD is a commonly used term to describe data generated from sources that relate to everyday clinical practice, generally outside the artificial constraints of randomized controlled trials. In its broad definition, RWD can include data generated as part of pragmatic controlled trials, however most RWD does not produce randomized evidence of treatment effect. In the context of Health Technology Assessment (HTA),

RWD typically presents as observational data from registries, administrative databases, and surveys." - Bell et al. 2016 (90)

While all the advantages of RWD might make it a suitable, available, and inexpensive option compared to RCT data, there are several obstacles to consider. While RCT data can minimise bias by balancing observable and un-observable confounders through randomisation, the main limitation of using routinely collected clinical data is that interpretations and subsequent decisions are conditional on the amount of bias and uncertainty present in the data (91). If not addressed, the potential bias within RWD poses the risk of inadvertently informing healthcare decisions into allocating resources to technologies that might be detrimental to patient welfare.

A review of 113 relevant single technology appraisals submitted to NICE using RWD/RWE from Bullement et al 2020 (92) found that while sources of RWE were routinely criticized as part of the appraisal process, only two cases were explicitly rejected. In the majority of cases, RWE was accepted in cancer drug submissions to NICE. Key criticisms of RWE in these submissions are rarely aimed at the use of RWE itself, but rather at specific data sources and the applicability of these to the decision problem. The review found that out of the 113 submissions, 71% used RWE to inform HRQoL estimates, 46% used it to inform costs, 40% to quantify healthcare resource utilization. Only a small number of submissions used RWE to inform efficacy, with 20 submissions using it for efficacy of both the intervention and the comparator, 17 for efficacy of the comparator, and 6 for the efficacy of the intervention. In all of these cases, RWE was used to supplement the estimate of efficacy from trials, rather than supplant it entirely.

Ultimately, integrating RWD into health technology evaluation presents a trade-off between generalisability of results for a wider patient population, and the risk of introducing bias in cost-effectiveness estimates. Combined with established literature data, the use of RWD can inform prospective research questions and provide valuable evidence for decision-makers (93). In particular, its use as data source for early-CEA could be beneficial for deciding whether or not a new technology has the potential to justify seeking more robust evidence.

3.2 Data Sources of electronic patient records

For the primary data source of breast cancer diagnoses, this study makes use of Scottish national records, in particular SMR06 and the SCAN Audit Data.

The Scottish Cancer Registry (SMR06) began collecting personal, demographic, and cancer diagnosis information in 1958. From 1997, additional clinical information has been included (such as staging and treatment information). Public Health Scotland (PHS) is responsible for the data collection and maintenance of the registry. Data collection is performed annually, and it includes all new cases of cancer affecting Scottish residents. For breast cancer, the information contained in the registry includes, but is not limited to: clinical/pathological staging, morphology, grade, subtype, date of diagnosis, date of treatment, type of treatment etc.

The Audit dataset is created for the purpose of looking at Health Board variation in Quality Performance Indicators. Maintained by the South East Scotland Cancer Network (SCAN), the Audit dataset has the purpose of supporting service improvements, ensure national standards are met and that clinical practice is delivered to an equitable standard. Similarly to the Cancer Registry with respect to breast cancer, the Audit Dataset contains all the diagnosis information mentioned previously, with the addition of more specific details regarding treatment and surgery.

3.3 Extraction and Processing of NHS Lothian Patient Records

The data for this study was obtained through the Scottish National Safe Haven, with the following specifications: all patients (NHS Lothian residents at date of diagnosis) diagnosed with early ER+ breast cancer (T1-4, N0-3, M0) within NHS Lothian from 01/01/2001 to 31/12/2017, with data linkage using Scottish national datasets, in particular, SMR06 and SCAN Audit Data.

The data linkage produced a total of 10,558 patients diagnosed between 2001 and 2017 with varying degree of information available depending on the year of diagnosis, due to new record-keeping practices implemented in the time frame. After excluding for duplicate, missing, or incomplete records, the viable cohort resulted in a total of 8,733 observations. Based on the difference of data available across the time span between 2001 and 2017 due to new record-keeping practices, this cohort was divided in two groups: a primary cohort, with 3,264 patients diagnosed between

2013 and 2017, and a secondary cohort, with 5,469 patients diagnosed between 2001 and 2012.

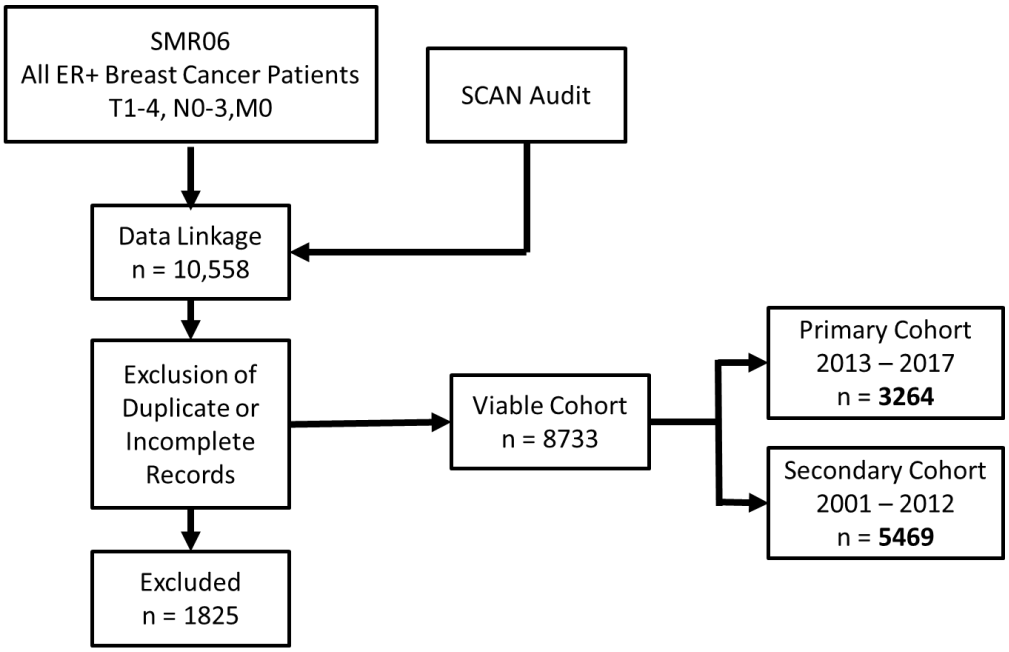


Figure 3: Flow Diagram of Inclusions and Exclusions of Patient Records

3.4 Risk Stratification

The two cohorts are stratified by risk, using available clinical data. The main risk stratification tool used is PREDICT, version 2.1 (13). Through the input of routinely collected clinical variables, PREDICT generates the survival probability of breast cancer patients at ten years from diagnosis. Specifically, the algorithm uses age at diagnosis, mode of tumour detection, ER status, HER2 status, KI67 status, tumour size and grade, number of nodes involved, to generate a recurrence score, which in turn produces the annual risk of recurrence for that specific patient. From the annual risk, the algorithm derives the predicted annual event rates which are then transformed to cumulative event rates. The cumulative event rates are converted to the 10-year survival probability, on the basis of how patient with similar characteristics fared. The survival probability is conditional on surgery alone as treatment, but PREDICT also calculates the probability increments due to other treatment combinations, producing the treatment benefits for hormone therapy, chemotherapy, trastuzumab, and bisphosphonates. The PREDICT probabilities for

these cohorts were generated in R (14), using the package “nhs.predict” (15). Further details on PREDICT and nhs.predict package can be found in Chapter 1.

To simulate the test scores of OncotypeDX, which this study considers as one of the comparators in the cost-effectiveness analysis, a proxy was used. As the patients in these cohorts were never administered OncotypeDX, a method using available clinical data to simulate the test results was implemented: GR-PR (94), which simulates OncotypeDX risk categories using tumour grade and PR status. Specifically, patients are assigned a score between 0 and 2, receiving one point if either the tumour grade is three or if PR-positivity is above 20%, two points for both. This simulation assumes an OncotypeDX low-risk cut-off of 18, which translates to a GR-PR score of 1 (94). This method is 73% accurate in assigning the correct OncotypeDX risk category in patients who were tested.

3.5 Descriptive Statistics

Table 1 reports the clinical characteristics and risk stratification of the primary and secondary cohort, respectively. The secondary cohort appears to have a larger share of high-risk patients, compared to the primary cohort, as it can be observed both from the PREDICT probabilities and chemotherapy benefits, and on the average tumour size, grade, and number of involved nodes. While relatively small in magnitude, the differences between the two cohorts are significant to a critical value of 0.1%, highlighting a change of the composition of the breast cancer patient population over time. The differences between the two cohorts could be explained by a combination of likely factors, including changes to treatment pathways and lower cancer discovery rates through screening. The primary cohort is used as the main data source for the cost-effectiveness analysis, as described in Chapter 5, as it is more reflective of current patient population and clinical practice. The secondary cohort is used to inform an alternative scenario analysis, to infer how the evaluated health technologies perform with a higher-risk patient population.

Table 1	**Primary Cohort 2013 – 2017**	**Secondary Cohort 2001 – 2012**
Number of Patients	3264	5469
Age, mean (range), y	62 (24 – 96)	60 (21 – 94)
Number of Involved Nodes, No. (%)		
None	2424 (74.3)	3505 (64.1)
1-3	633 (19.4)	1288 (23.5)
4-9	148 (4.5)	439 (8.0)
>10	59 (1.8)	237 (4.4)
Histological grade, No. (%)		
1	616 (18.9)	885 (16.2)
2	1826 (55.9)	2772 (50.7)
3	822 (25.2)	1812 (33.1)
Largest tumour size, mean (range), mm	19 (1–90)	21 (0.5–90)
≤30 No. (%)	2796 (85.7)	4588 (83.9)
>30 No. (%)	468 (14.3)	881 (16.1)
Relapse-free Survival, median(range), 10y	75 (0.3 – 96)	69 (0.1 – 97)
Chemotherapy Benefit, No. (%), 10y		
Endocrine therapy only	2712 (83.1)	3923 (71.7)
Endocrine and Chemotherapy	552 (16.9)	1546 (28.3)
GRPR Score, No. (%)		
GRPR < 1	1511(45.4)	2073(37.9)
GRPR ≥ 1	1753(54.6)	3396(62.1)

Table 1. Clinical characteristics and risk stratification of the Primary and Secondary Cohorts.

Figure 4 displays the density distributions of PREDICT estimates of survival at 10 years from diagnosis for the two cohorts: within a 95% confidence interval, the recurrence-free survival at 10 years of a patient form the Primary Cohort is expected to be between 3% and 8% higher than the survival of a patient from the Secondary Cohort.

In line with the characteristics of typical RWD, the dataset, contains a portion of missing data. A total of 1825 diagnoses was excluded on the basis of record duplicates or incomplete records (incomplete with respect to key variables needed for the calculation of PREDICT scores). The reason why certain variables might be missing from the records could be due to early mortality, or due to frailer patient

being unfit for diagnostic work-up. This group of patients might primarily contain older patients or patients affected by severe comorbidities. While the decision to exclude patients with missing records might skew the distribution towards the relatively healthier and younger breast cancer population, it can be reasonably discarded as a potential bias, as these patients, being unfit even for diagnostic work-up, would be less likely to undergo surgery or anti-cancer therapy.

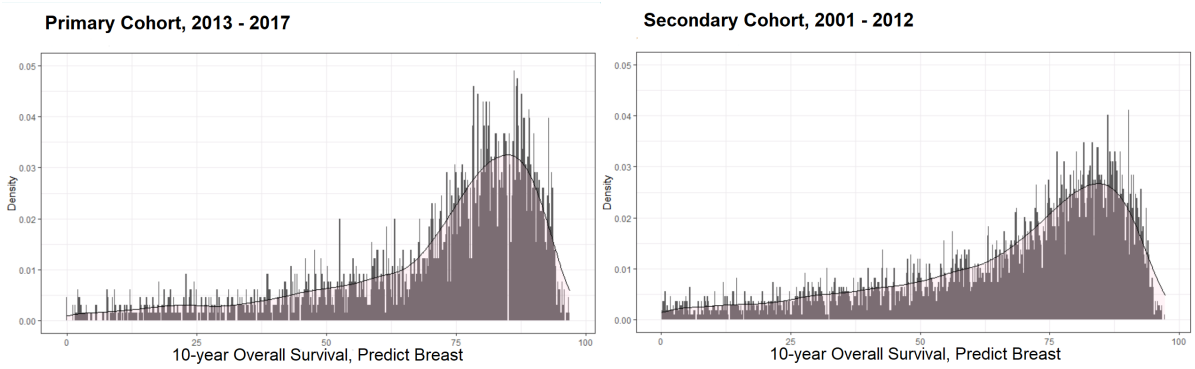


Figure 4: Distribution density of 10-year Overall Survival for the Primary and Secondary Cohorts

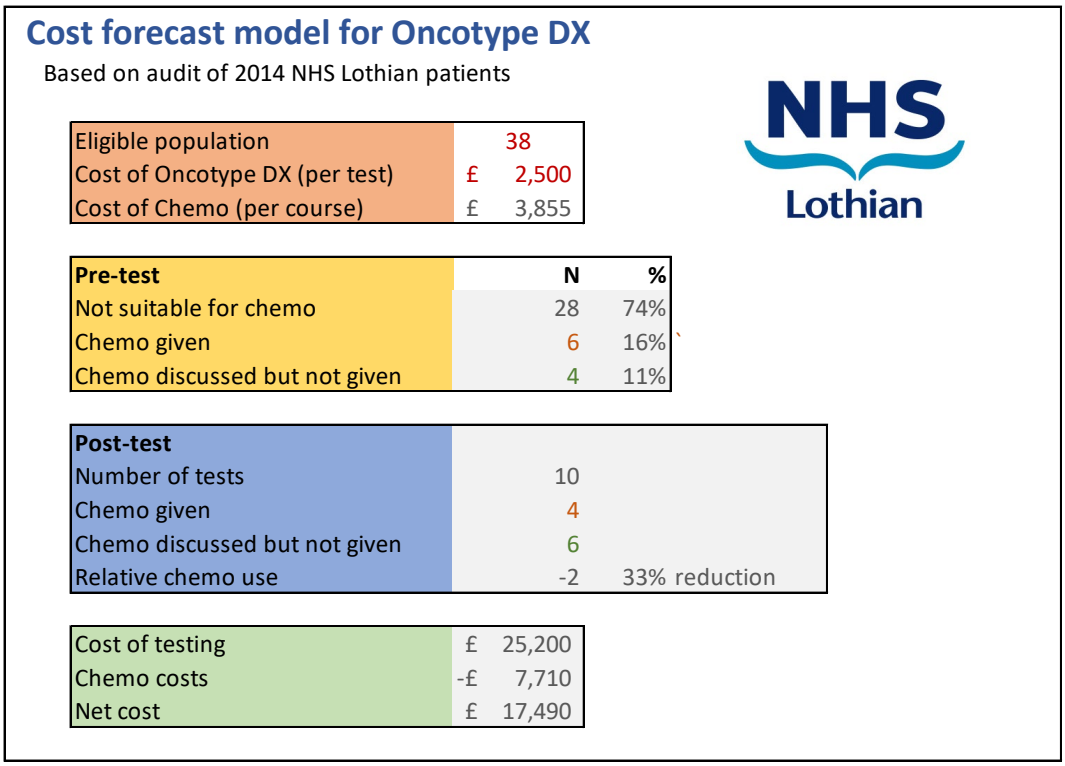
3.6 Demonstrating the Use of Real-World Data

The following chapter (Chapter 4) presents a demonstration of the use of RWD for assessing the impact of introducing a new technology in a clinical setting. The analysis describes the decision and budget impact of the use of OncotypeDX within NHS Lothian in 2016, using the small number of patients that were administered the test and a representative cohort. The results show the likely impact that the test might have within the same decision parameters adopted for those who were tested. The analysis was performed in 2018 as part of an NHS Lothian business audit and makes use of a currently outdated version of PREDICT (v1.2) (13).

4. Decision Impact of OncotypeDX: a Natural Experiment Using Routine Data

4.1 Introduction and Background

OncotypeDX is a gene expression profiling test that can provide additional information on the biological features of an individual's early breast cancer which can refine prognostication of the course of the disease and help guide decisions based on whether adjuvant chemotherapy is likely to be of sufficient benefit to justify the risks of treatment. Using a 21-gene expression panel, OncotypeDX produces a continuous recurrence score, ranging 0-100, indicating the 10-year risk of recurrence (assuming 5 years of Tamoxifen), and the prediction of adjuvant chemotherapy benefit. The higher the score, the higher the risk of recurrence and higher the benefit from chemotherapy. In January 2016, the Molecular Pathology Evaluation Panel assessed and evaluated this test and accepted OncotypeDX on grounds of effectiveness and cost-effectiveness for use in NHS Scotland, subject to a discounted price. The test was "recommended for patients with ER-positive, HER2-negative, lymph-node-negative breast cancer with NPI ≥ 3.4 (8) in whom current risk stratification would lead to a recommendation for adjuvant chemotherapy but in whom the benefits of chemotherapy are considered uncertain by the MDT". Before the adoption of OncotypeDX, a cost forecast model for the genomic test was drawn, based on audit of 2014 NHS Lothian patients (Table 2). A prospective audit is necessary to assess the use and impact of OncotypeDX in NHS Lothian, for the reappraisal of its role in routine care.



Cost forecast model for Oncotype DX

Based on audit of 2014 NHS Lothian patients

NHS Lothian

Eligible population	38
Cost of Oncotype DX (per test)	£ 2,500
Cost of Chemo (per course)	£ 3,855

Pre-test	**N**	**%**
Not suitable for chemo	28	74%
Chemo given	6	16%
Chemo discussed but not given	4	11%

Post-test		
Number of tests	10	
Chemo given	4	
Chemo discussed but not given	6	
Relative chemo use	-2	33% reduction

Cost of testing	£ 25,200
Chemo costs	-£ 7,710
Net cost	£ 17,490

Table 2. Estimates for eligible population, chemotherapy use and Oncotype DX uptake. Extracted from NHS Lothian Internal Report for OncotypeDX 2014 audit

4.2 Data and Methods

The audit relies on national datasets available through the QPI audit, supplemented by national datasets curated by the NSS, the contents of which are described in Chapter 3. The dataset includes all the patients treated for Breast Cancer by NHS Lothian between 2010 and 2017, while OncotypeDX became available from 2016 onwards. The first phase of the analysis features descriptive statistics regarding the patients that were tested with OncotypeDX, including the recurrence score from the test, their PREDICT scores and NPI. Moreover, statistics regarding chemotherapy use are presented for the patients within the same PREDICT v1.2 interval of the tested patients, for the same years and for the years before OncotypeDX availability. The second phase features a logistic regression model for the evaluation of probability of receiving chemotherapy before and after the availability of OncotypeDX. This allows the estimation of the relative reduction in the probability of receiving chemotherapy for those who are tested with OncotypeDX. The third phase shows a simulation of a representative Scottish cohort for the estimate of the budget impact of OncotypeDX, given the results of the second phase and the actual use of the test in NHS Lothian.

4.3 Summary of Assumptions

The following list is a summary of the main assumptions used throughout the budget and decision impact of the analysis.

- PREDICT chemotherapy benefit scores are a reflection of clinical decision-making with respect to chemotherapy assignment and prognostication.
- Holding other factors equal, patients with the same PREDICT chemotherapy benefit have the same probability of being prescribed chemotherapy.
- The estimated costs associated with treatment, including toxicities, are on average an accurate reflection of the costs incurred by the NHS when treating a breast cancer patient with chemotherapy.
- Cost neutrality of a decision-making tool is achieved when at the cohort level the sum of the cost of testing is equal to the sum of savings due to chemotherapies spared.

4.4 Descriptive Statistics

Between 2016 and 2017, a total of 36 patients were tested with OncotypeDX, the 26% of which were treated with adjuvant chemotherapy. According to the guidelines, all patients tested were eligible for OncotypeDX, with the exception of two patients: one whose electronic records report ER-negativity, and another patient who was node positive. The tables below report the descriptive statistics for PREDICT v1.2 chemotherapy benefit score, NPI score, and OncotypeDX score for the 36 patients tested with the genomic test.

	Minimum	Median	Mean	Maximum	Std. Deviation
PREDICT 1.2	0.81	3.13	3.22	6.17	1.35
NPI	6.2	7.8	8.2	12.2	1.79
OncotypeDX	6	20.5	22.5	52	10.5

Table 3. Risk scores distributions of tested patients. Summary measures of dispersion for the distributions of PREDICT chemotherapy score, NPI, and OncotypeDX.

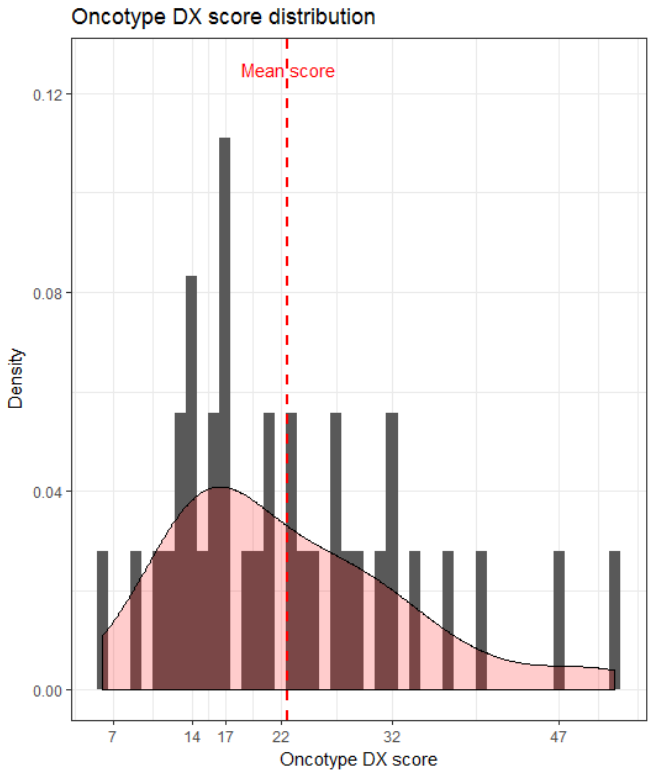


Figure 5: OncotypeDX Score Distribution

Figure 5: OncotypeDX Score Distribution

Score distribution of the 36 patients that were tested with OncotypeDX between 2016 and 2017. Mean score and density are displayed.

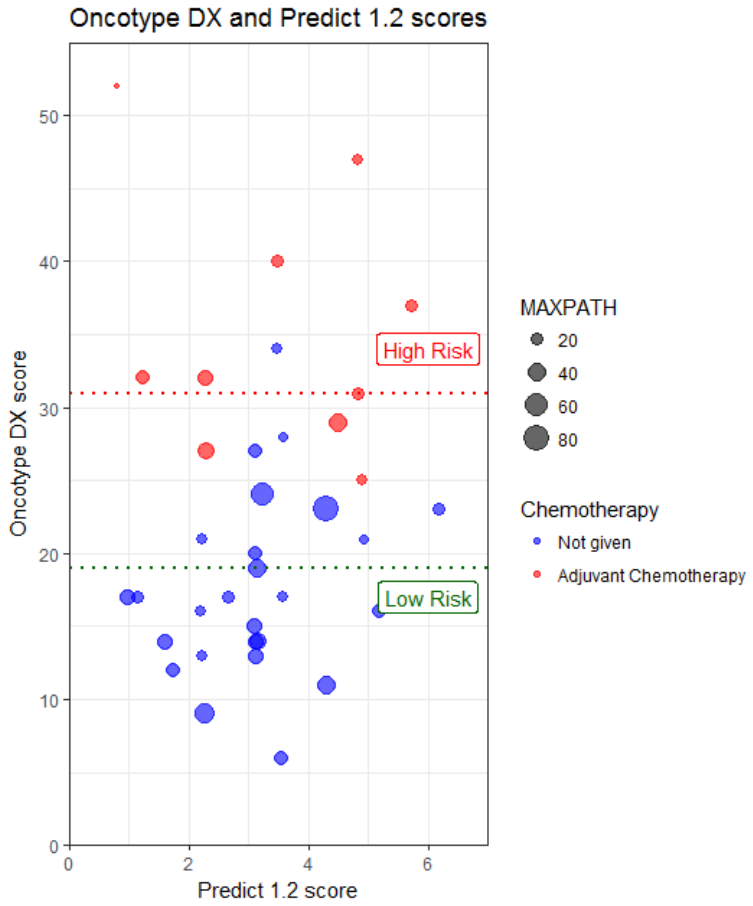


Figure 6: OncotypeDX against Predict Breast v1.2 scores, stratified by tumour size and treatment.

Plot of Predict v1.2 against Oncotype-DX scores. The colour of the data-point identifies the treatment, while the size of each individual observation is proportional to the size of the tumour (in millimetres).

Figure 6: OncotypeDX against PREDICT v1.2 scores, stratified by tumour size and treatment

Using the PREDICT v1.2 chemotherapy benefit score interval and the characteristics (including ER status and Node status) of the patients tested with OncotypeDX, a control group was identified. A total of 575 patients with similar characteristics to the OncotypeDX group was extracted from the 2016-2017 dataset. The table below reports the use of the test and of chemotherapy in the group of patients identified by the PREDICT v1.2 interval (both patients tested and controls).

Table 4.	No Test	OncotypeDX	Row sum
No Chemotherapy	**499 (81.6%)**	**26 (4.3%)**	**525 (85.9%)**
Chemotherapy	**76 (12.5%)**	**10 (1.6%)**	**86 (14.1%)**
Column Sum	**575 (94.1%)**	**36 (5.9%)**	**611 (100%)**

Table 4. Two-way table for OncotypeDX and chemotherapy. Two-way table reporting absolute numbers and proportions for patients that were tested and the control group.

4.5 Decision Impact Analysis

The difference in the probability of receiving chemotherapy between patients tested with OncotypeDX and the controls has been evaluated with a logistic regression, using a binary variable Test/No-Test and the PREDICT v1.2 score as regressors.

Model : P(chemotherapy) = OncotypeDX + Predict Breast v1.2

Table 5.

Coefficients	**Estimate**	**Exponentials**	**Std. Error**	**P value**
Intercept	-4.624	0.009	0.336	0.000***
OTDX	-0.899	0.407	0.513	0.079
PRED1.2	1.269	3.555	0.124	0.000***

Table 5. Logistic Regression Coefficients, for estimating the probability of receiving chemotherapy, conditional on the use of OncotypeDX and PREDICT v1.2 score.

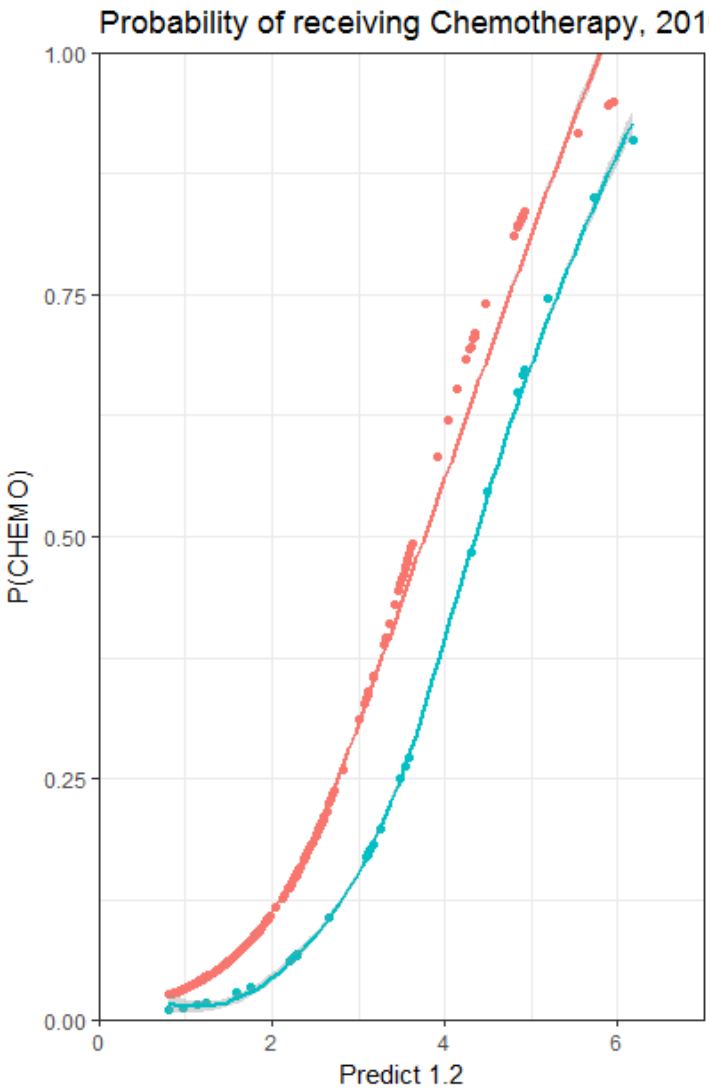


Figure 7: Logistic Regression, visualised.

Plot of the fitted probability of receiving chemotherapy, against Predict 1.2, using the coefficients from the modelled logistic regression.

The red curve follows the fitted probabilities of patients that were not tested with Oncotype-DX. The blue curve outlines the fitted probabilities of tested patients.

Figure 7: Logistic Regression, visualised

To assess whether the probability of receiving chemotherapy has changed across the years, independently of the introduction of OncotypeDX, a logistic regression using PREDICT v1.2 as regressor was run on a 2013-2015 group of patients, selected with the same criteria as the previous controls. The estimate for the PREDICT v1.2 coefficient was then compared to the coefficient of the 2016-2017 group. The table below displays the comparison of the coefficients, including 95% confidence intervals.

Coefficients	Estimate	2.5%	97.5%
Pred1.2 (2013-2015)	**1.337**	**1.086**	**1.614**
Pred1.2 (2016-2017)	**1.268**	**1.03**	**1.52**

Table 6. Estimated coefficients of PREDICT v1.2 scores as predictors of the probability of receiving chemotherapy. Comparison of the estimated coefficients of PREDICT as a regressor for the probability of receiving chemotherapy.

The coefficients from the original logistic regression were then applied to the PREDICT v1.2 scores of the 2013-2015 group, and fitted probabilities were calculated. Figure 8 displays the plot of fitted probabilities using both the coefficients of 2016-2017 and of 2013-2015.

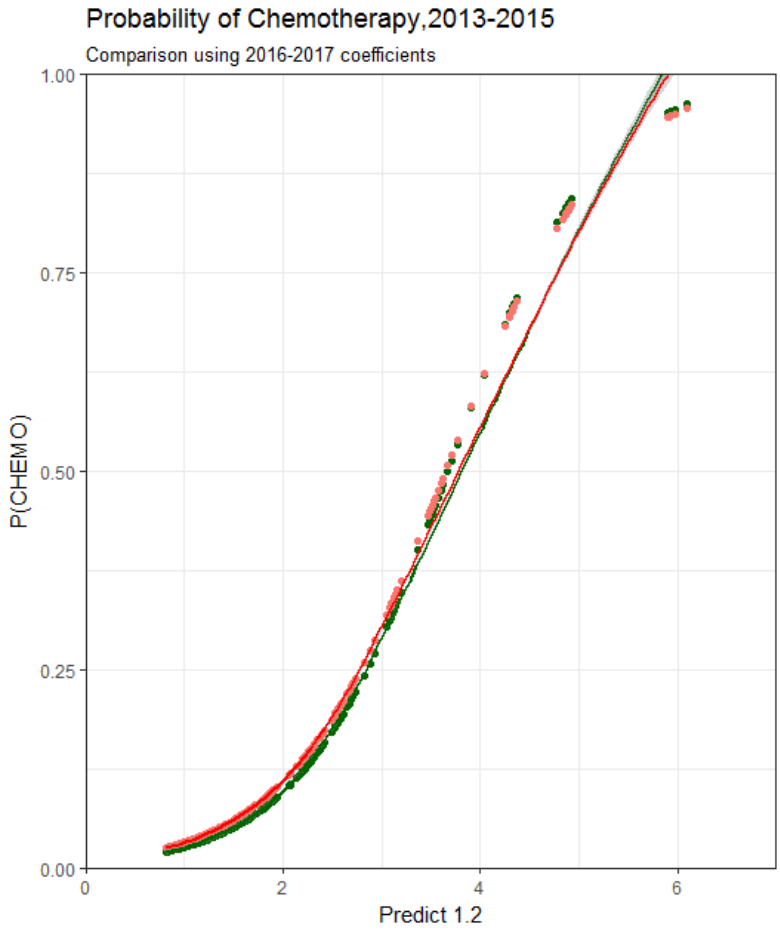


Figure 8: Comparison of fitted probabilities

Figure 8: Comparison of fitted probabilities

Plot of fitted probabilities, using 2013-2015 data. In red, fitted probabilities using 2016-2017 coefficients. In green, fitted probabilities using 2013-2015 coefficients.

Given the results of the comparison, there is not enough evidence that the probability of receiving chemotherapy has changed across the years, independently of the availability of OncotypeDX in the 2016-2017 group. As such, it is possible to solely rely on the coefficients of the 2016-2017 group to calculate the relative difference in the probability of receiving chemotherapy between patients tested with OncotypeDX and patients not tested, using as baseline the group of patients that were actually tested with the 21-gene signature.

Fitted probabilities for the 36 tested patients were calculated, with and without the effect of OncotypeDX; Figure 9 shows the plot of said fitted probabilities and the change in the difference as the PREDICT v1.2 score increases.

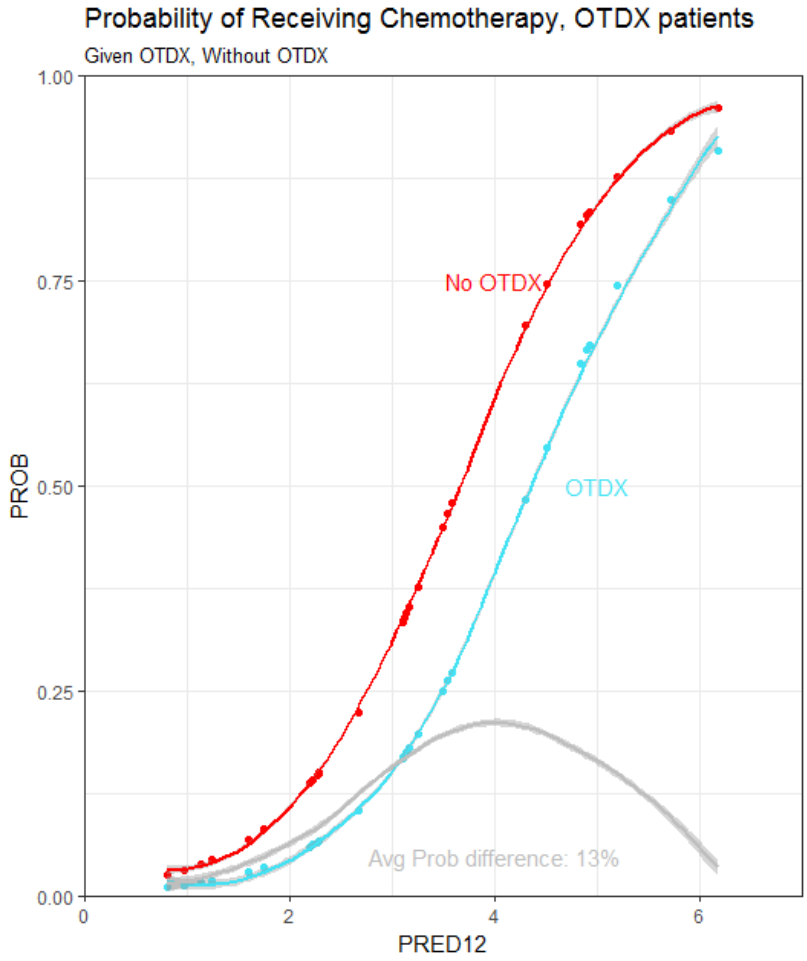


Figure 9: Impact of OncotypeDX

The blue line shows the fitted probability of receiving chemotherapy, given the fact that the patients were tested. The red line outlines the fitted probability of receiving chemotherapy as if those patients were never tested.

The grey line is the difference between the first two curves.

Figure 9: Impact of OncotypeDX

The average reduction in the probability of receiving chemotherapy as a consequence of being tested with OncotypeDX is 13%. Meaning that if there had been no test, there would have been 13% more patients treated with chemotherapy. Given that the actual use of OncotypeDX is ~6% of the patients within the same PREDICT v1.2 interval, a 13% reduction in chemotherapy use in the tested population means a global reduction of chemotherapy of 0.767% for the eligible population.

4.6 Budget Impact Analysis of OncotypeDX

Chemotherapy costs were calculated as the expected value of the cost of chemotherapy given the proportional use of 8 different drug regimens, each with associated supportive medication and toxicity costs. The prices for chemotherapy and supportive drugs, and the costs for toxicity events were drawn from the electronic Market Information Tool, the British National Formulary, and NHS reference costs 2015/2016. The average total cost of regimen per patient is £ 4,159, calculated using NHS Reference Costs, with the addition of “Patient Level Information and Costing System” (PLICS) cost for Cancer-related Febrile Neutropoenia. The non-confidential cost of OncotypeDX is £ 2500. The following calculations assumes a representative Scottish cohort of 600 breast cancer patients, with similar characteristics to the groups identified in the previous phase: ER+, N0 breast cancer with PREDICT v1.2 between 0.81 and 6.17. The proportion of patients undergoing chemotherapy has been calculated as the probability of chemotherapy in absence of the test from the 2013-2017 data. The proportion of patients being tested is reflective of the actual use of OncotypeDX between 2016 and 2017. The overall reduction chemotherapy use due to the test is calculated using the difference in the probability of receiving chemotherapy as shown in the previous phase (13%), and the actual use of OncotypeDX (5.9% of the cohort). Cost-neutrality, given these parameters, is achieved at an illustrative price of £595.

Scottish Representative Cohort

Total number of patients	600	P(chemo) without test	0.13
Chemotherapy cost	£4,159	Proportional reduction in chemo	0.00767
OncotypeDX cost	£2,500	use due to OncotypeDX	
Chemotherapy use in absence of test		Chemotherapy use with test	
		Number of tests:	35
Patients given chemotherapy:	78	Patients given chemotherapy:	73
Patients not given chemotherapy:	522	Patients not given chemotherapy:	527
Relative chemotherapy use:	5	Cost of testing:	£87,500
Chemotherapy savings:	£20,795	Net cost of testing:	£66,705

Table 7. Budget Impact of OncotypeDX on a Scottish Representative Cohort

As the difference in probability of receiving chemotherapy due to the test is a function of PREDICT v1.2, the results have been broken down by chemotherapy benefit interval.

For PREDICT v1.2 chemotherapy benefit between 3 and 5%, assuming a cohort of 90 patients with those characteristics, the probability of receiving chemo independently of the test is 54%. With a test rate of 23.6%, the overall reduction in the probability of receiving chemotherapy is 4%, given an average reduction in probability of chemotherapy for tested patient at 18%.

For chemotherapy benefit below 3%, assuming a cohort of 500 patients, the probability of receiving chemotherapy in absence of the test is 7%, while the rate of testing is ~2.5%. For this subgroup, the average reduction in probability of chemotherapy due to the test is 6%, meaning an overall reduction for the cohort of less than 0.2%.

For patients with chemotherapy benefit above 5%, the rate of testing is 37%, while the baseline chemotherapy probability is 62%. The average reduction of chemotherapy probability for the tested patients is 9%. In a simulated cohort of 10 patients, there would be 4 tested patients and 6 chemotherapies. Given that the overall reduction in chemotherapy probability is 3.3%, there would be no chemotherapy savings. The overall cost of testing would be £ 10,000.

4.7 Conclusion

The budget impact analysis indicates that with the testing rates observed between 2016 and 2017 within NHS Lothian, the commercial price of OncotypeDX is far above the cost-neutrality level, calculated as £595 in the representative Scottish cohort example. While the estimated chemotherapy use reduction due to OncotypeDX is not statistically significant in this sample, the overall trend is in line with results from other decision impact analysis informed by UK patient data (95, 96). This could be a further indication that the number of patients tested with the genomic assay is far below the number necessary to observe a chemotherapy use reduction comparable to what has been reported by Holt et al 2013 and Loncaster et al. 2017. Moreover, the budget impact analysis suggests that the net cost of testing is lowest when OncotypeDX is used in the PREDICT v1.2 interval between 3% and 5%, where the estimated benefits of chemotherapy are most uncertain. Due to the low number of patients tested with OncotypeDX, the significance of the results of the decision impact analysis is uncertain. Further evidence or higher testing rates are required for a more precise calculation of the impact that the 21-gene signature had on the probability that a patient receives chemotherapy. Further research would ideally use an updated version of PREDICT, and a larger cohort containing higher testing rates. Moreover, it would be beneficial to compare survival outcomes of the tested and non-tested groups, matched by patient-characteristics: since the evidence in clinical practice indicates that the current commercial price is not balanced against chemotherapy reduction, maybe long-term outcomes might justify the higher price.

5. Cost-Effectiveness Modelling of a 4-gene signature for therapy response prediction in breast cancer

5.1 Introduction

As discussed in Chapter 1, the implementation of new biomarker-based predictive tests has the potential for better clinical outcomes and a more efficient resource allocation within the NHS. The evaluation of the cost-effectiveness of a new biomarker in a routine NHS setting is complex, relying on impact through clinical decision making and modifications of the clinical pathway followed by a patient. Moreover, as discussed in Chapter 3, while the use of routinely collected patient data may better reflect real-world practices, the analysis and the results might be biased because of it, due to the non-randomised design and the presence of confounding factors.

The following economic evaluation will take as example a 4-gene signature developed within the Edinburgh Cancer Research Centre and the Institute of Genetics and Cancer. The signature, EER4 (68), has been developed into a promising method for accurately predicting response to pre-operative endocrine therapy in early breast cancer using patient-matched pre- and on-treatment biomarkers; further information is available in section 2.6. This method out-performs all established pre-treatment only assessment methods, but potentially modifies the patient's clinical pathway. The cost-utility model will examine whether or not this new technology would be a cost-effective use of NHS resource, quantify the impact of its introduction clinical care, and under which conditions and assumptions the new technology can be chosen as a better alternative for the current standard of care.

5.2 Data

A full description of the data and patient cohort is covered in Chapter 3. The study population comprises all patients diagnosed with early ER-positive breast cancer between January 2001 and December 2017. The dataset was divided in two groups, based on the availability of additional linked datasets and local record-keeping practices. After excluding patients with missing observations or incomplete clinical records (with respect to several key variables), this resulted in two cohorts:

- The primary cohort, patients diagnosed between 2013 and 2017, for a total of 3264
- The secondary cohort, patients diagnosed between 2001 and 2012, for a total of 5469

The primary cohort is the main data source informing the model, supplemented with literature sources for utility weights, costs and other parameters that could not be inferred from the patient cohort. The secondary cohort is used for providing estimates of cost-effectiveness in a higher-risk patient population. Details and descriptive statistics of the two cohorts are available in section 3.4. As previously noted in section 3.3, the decision to separate the dataset into two cohorts was driven by changes in record-keeping practices for the primary data sources, SMR06 and SCAN Audit data. This resulted in key risk stratifiers for the calculation of PREDICT being unavailable pre-2013, leading to greater uncertainty in risk stratification and recurrence-free survival estimation in the secondary cohort. As shown in Table 1 of section 3.5, the secondary cohort had a higher proportion of patients in higher-risk categories, based on factors such as positive lymph nodes, tumour size, and tumour grade, compared to the primary cohort. This trend is likely due to increased uncertainty and missing data in the secondary cohort, as well as increased earlier diagnosis rate and treatment rates in the primary cohort.

5.3 Methods

The cost-effectiveness model was developed in line with the Consolidated Health Economics Evaluation Reporting Standards (CHEERS) guidelines (97). A time-dependent discrete-state transition model was used to assess the cost utility of four strategies for guiding treatment decision in a representative cohort of Scottish patients diagnosed with ER-positive early breast cancer. The model was developed using R Programming Language (14). The full code of the model is available in Appendix 1 of the thesis. The strategies considered were:

1. Predict.breast (Baseline): After diagnosis, the patient undergoes surgery, and the PREDICT score for chemotherapy benefit is calculated. If the score is

below 3%, the patient receives endocrine therapy only; if the score is above 3%, the patient receives endocrine and chemotherapy.

2. Predict.breast and OncotypeDX: If the chemotherapy benefit score is below 3%, patient receives endocrine therapy only. If the score is between 3% and 5%, the patient is tested with OncotypeDX. If the patient tests low risk, endocrine therapy only; if high risk, endocrine and chemotherapy. If the chemotherapy benefit is above 5%, the patient receives endocrine and chemotherapy.
3. EER4 Only: After diagnosis, the patient undergoes two weeks of neoadjuvant endocrine therapy and two biopsies, and then surgery. If the patient tests as responder, endocrine therapy only; if they test non-responder, endocrine and chemotherapy. Responders benefit from a mean hazard ratio of 0.35 in RFS against non-responders.
4. EER4 and Predict.breast: After diagnosis, the patient undergoes two weeks of neoadjuvant endocrine therapy and two biopsies, and then surgery. If the patient's chemotherapy benefit score is below 3%, patient receives endocrine therapy only. If the score is between 3% and 5%, the patient is assigned chemotherapy based on the test result of EER4. If the chemotherapy benefit is above 5%, the patient receives endocrine and chemotherapy. Responders benefit from a mean hazard ratio of 0.35 in RFS against non-responders.

Figure 10 provides a visual breakdown of the four strategies.

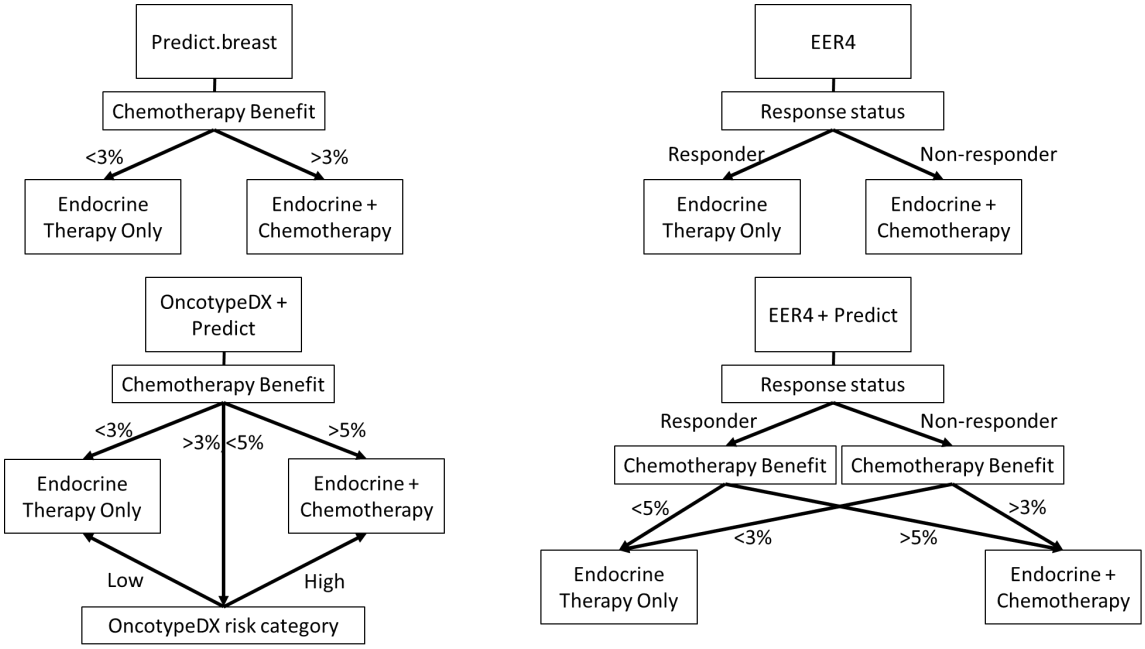


Figure 10: Flow Diagram of the four testing strategies

The model was developed from the payer perspective of the National Health Service, and assumes a life-time horizon with 1-year cycle length. Both costs and health states were discounted at a rate of 3.5% per year. All prices were captured in 2016-pound sterling for health states costs and for treatment costs.

5.3.1 Simulating Test Scores

As OncotypeDX was not routinely administered to these cohorts, the test risk categories were simulated using clinically observable variables, according to the GR-PR score: by assigning a score based on tumour grade and progesterone receptor status, this proxy simulates the risk categories of OncotypeDX with 73% accuracy, depending on the chosen cut-off point (94). For this analysis, the low cut-off of 18 for low-risk OncotypeDX was chosen (further details in section 3.3).

EER4 is assumed to be an independent predictor of response with respect to clinical variables. As such, the simulated test outcome for response is assigned through a random distribution reflective of the proportion of responders and non-responders (0.7 vs 0.3) described in the study (68). Response status is used in the model as a proxy for risk categories, with responders assigned to the low-risk group and non-responders assigned to the high-risk group.

The version of PREDICT used by this study is v2.1, which calculates individual overall survival plus treatment benefits of hormone therapy, chemotherapy, trastuzumab, and bisphosphonates. The chemotherapy benefit score is used as a decision rule to assign patients to risk categories, with patients scoring less than 3% belonging to the low-risk group, between 3% and 5% to medium-risk, and above 5% to the high-risk group. The evaluated baseline strategy (PREDICT only) combines the medium and high-risk groups into a single high-risk group.

Moreover, this study assumes a constant relationship between overall survival and relapse-free survival; a hazard ratio was applied to the annual event rates calculated by PREDICT v2.1, thus producing individual relapse-free survival estimates plus treatment benefit scores. These estimates were subsequently used to generate the transition probabilities for the model.

5.3.2 Model Structure

The model features seven health states:

1. Disease-free
2. Local Recurrence
3. Disease-free after Local Recurrence
4. Distant Recurrence
5. Congestive Heart Failure
6. Acute Myeloid Leukaemia
7. Death

The model structure and the seven modelled health states are illustrated in Figure 11. In each strategy, patients are allocated to either a high-risk or a low-risk group, where the high-risk group receives adjuvant chemotherapy. All patients are assumed to receive identical adjuvant endocrine therapy. The model development did not feature expert consultation, as the model structure is in line with previous studies in similar clinical settings (84, 98).

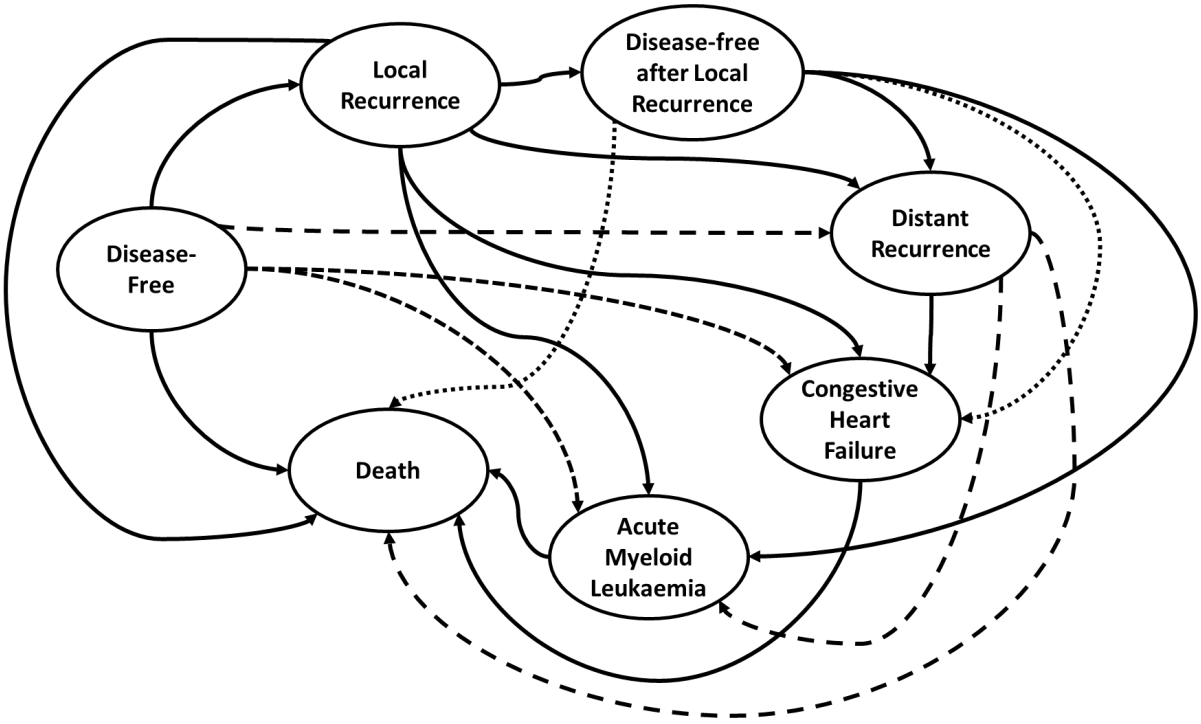


Figure 11: Visualisation of Health States and Markov Cycle

The probability of a patient moving from one state to another during each cycle is used to predict the proportion of patients in each state at each cycle during and after treatment. The total Quality Adjusted Life Years (QALYs) associated with a given strategy is then calculated by combining the proportion of the cohort in each state per cycle with the utility associated with these states. Costs are calculated in the same fashion, with the addition of the test costs, chemotherapy costs (if administered), and terminal costs (if patient dies from breast cancer).

The simulation differs from a standard Markov model in that the probability of staying in the disease-free state varies depending on how many cycles were spent in the disease-free state itself and on whether chemotherapy was administered.

Patients enter the model at the start of adjuvant therapy and are assumed to be disease-free. Disease-free patients can either stay disease-free, or develop a recurrence (local or distant), or develop congestive heart failure, or acute myeloid leukaemia, or die. Patients that develop a local recurrence can be cured, but cannot reverse to “disease-free”, thus they can move to a separate “disease-free after local recurrence” state. Patients that develop a distant recurrence stay in the distant recurrence state until death. Patients that develop CHF stay in that state until death from CHF. Patients that develop AML stay in the AML state until death from other

causes or AML mortality. A half-cycle correction was applied to adjust for patients that do not transition at the end of each yearly cycle, thus providing a calculation of QALYs and costs reflecting an average transition in the middle of each cycle. Moreover, the QALYs and costs are weighed by a compliance parameter set at 95%, to better reflect a clinical setting: if chemotherapy is assigned, 95% of patients will follow the recommendation, while 5% will be prescribed endocrine therapy only; vice versa, if endocrine therapy only is assigned, 95% of patients comply, and 5% will also be prescribed chemotherapy.

5.3.3 Model Input Parameters

Assuming a constant relationship between Overall Survival and Recurrence-free Survival, cancer recurrence rates were estimated using the 10-year relapse-free survival estimate of PREDICT. The effect of chemotherapy on relapse-free survival is based on the hazard ratio used by the PREDICT algorithm: the specified hazard ratio is then applied to the annual event rate derived from the Recurrence-free Survival estimate. Moreover, the impact of EER4 was reflected in the chemotherapy effect by applying an additional hazard ratio to the event rate, in accordance with the estimates of Turnbull et al. (68). Chemotherapy procurement, delivery, and toxicity costs were taken from BNF (99) and NHS Reference costs (100), while chemotherapy assignment proportions were taken from the OPTIMA prelim trial (101). Chemo-related toxicities considered were: febrile neutropoenia, allergic reactions, nausea, diarrhoea, anaemia, thrombocytopaenia, and stomatitis. The costs associated with breast cancer-specific health states were taken from Hall et al. 2017 (84), and adjusted for inflation after the distributions were generated (i.e. the distribution shape parameters are retained, then the inflation multiplier is applied to the output). Cost of testing with EER4 was determined with a log-normal distribution selecting for credible value based on the cost of testing of similar technologies. The cost of testing with OncotypeDX was fixed at half of the reported commercial price. The choice of using a lower cost for OncotypeDX is motivated by two reasons: the simulation is using a proxy for the test results, which might be inherently less accurate, and the simulation assumes that OncotypeDX might be offered at a discounted price to the NHS (as described in Section 4.1). Baseline utility values were taken from Kind et al. 1998 (102), and utility weights and decrements were

obtained from Campbell et al. 2011 (103). Utility values for congestive heart failure and acute myeloid leukaemia were taken respectively from Kirsch et al. 2000 (104) and Castejòn et al. 2018 (105). A comprehensive list of input parameters and sources can be found in tables 8, 9, and 10.

Table 8. Model Input Parameters

Parameter	Base case	Shape Parameters	Distribution	Description	Source
Effect of chemotherapy on risk of recurrence	0.73	S1: 71.2115 S2: 26.3385	Beta	Hazard ratio of the risk of recurrence for patients treated with adjuvant chemotherapy	Wishart et al, 2010 (13)
Proportion of endocrine therapy responders	0.7	S1: 209 S2: 89	Beta	Proportion of patient that respond to neoadjuvant endocrine therapy	Turnbull et al, 2015 (68)
Endocrine therapy effect on responders identified by EER4	0.35	S1: 397 S2: 738	Beta	Hazard ratio for recurrence rates of responders vs non-responders as identified by the test	Turnbull et al 2015 (68)
Proportion locoregional vs distant recurrence	0.31	S1: 292 S2: 663	Beta	Proportion of recurrences that are local to the primary cancer	Baum et al, 2003 (106)
Death after distant recurrence, with chemotherapy	0.30	S1: 1.0 S2: 2.35	Beta	Annual probability ER-positive tumours	Walkington et al, 2012 (107)
Death after recurrence, no chemotherapy	0.14	S1: 0.42 S2: 2.58	Beta	Annual probability ER-positive tumours	Albain et al, 2009 (108)
Background mortality	Age-specific		fixed	Life tables	Office of National Statistics, 2009 (109)

Chemotherapy excess mortality (first year)	0.0024	S1: 1.6 S2: 677.6	Beta	Excess mortality due to chemotherapy toxicities	Albain et al, 2012 (110)
Background rate CHF	Age-specific		fixed	Annual age-specific female incidence of CHF	Townsend et al, 2012 (111)
Relative risk of CHF with anthracycline treatment	1.61	S1: 0.458 S2: 0.191	Log-normal	Applied as a constant lifetime risk	Albain er al, 2012 (110)
Death after CHF	0.6	S1: 136 S2: 84	Beta		Cowie et al, 2000 (112)
Background annual rate AML	0.0029	fixed	fixed		Bhayat et al, 2009 (113)
Relative risk of AML after chemo	7.6	S1: 1.71 S2: 0.79	Log-normal		Wolff et al. 2014 (114)
Relative 5-year survival for AML (female,age 65+)	0.0383	S1: 1524 S2: 60	Beta	Assumes constant relative survival, applied to background mortality rate	Oliver et al, 2013 (115)

Table 9. Model Cost Parameters

Cost Parameter	Base Case (£)	Shape Parameters	Notes	Source
Disease free (annual cost excluding chemotherapy-related costs)	1,203.03	Log Normal S1: 6.91 S2: 0.004	Inflation multiplier: 1.2	Hall et al 2017 (84)

Disease free after local recurrence (annual)	1,607.17	Log Normal S1: 7.2 S2: 0.11	Inflation multiplier: 1.2	Hall et al, 2017 (84)
Local recurrence (first year)	7,371.97	Log Normal S1: 8.72 S2: 0.08	Inflation multiplier: 1.2	Hall et al, 2017 (84)
Distant recurrence (annual)	2,022.85	Log Normal S1: 7.43 S2: 0.019	Inflation multiplier: 1.2	Hall et al 2017 (84)
Terminal disease (final 3 months)	2,058.84	Log Normal S1: 7.63 S2:0.003	Inflation multiplier: 1.2	Hall et al 2017 (84)
Cost of treating with chemotherapy	3536.43		Weighted average of the six chemo regimens considered	BNF (99) NHS Reference Costs (100) Optima trial (101)
Cost associated with chemotherapy toxicities	312.44		Weighted average based on the 6 chemo regimens considered	BNF (99) NHS Reference Costs (100) Optima trial (101)
Cost of testing with EER4	1,674.23	Log-normal S1: 7.42 S2: 0.03	Based on test costs of similar technologies	
Cost of testing with OncotypeDX	1,250.00	fixed		

Table 10. Model Utility Parameters

Utility Parameter	Base Case	Shape Parameters	Distribution	Source
Starting utility	Age group specific 60-64 = 0.81 65-74 = 0.78 75-100 = 0.71	fixed		Kind et al 1998 (102)
Disease free (no chemo)	-0.003	S1: -8.117 S2: 2.148	lognormal	Campbell, 2011 (103)
Disease free (on chemo, first year)	-0.099	S1: -2.365 S1: 0.325	lognormal	Campbell, 2011 (103)
Local recurrence	-0.108	S1: -2.29 S2: 0.359	lognormal	Campbell, 2011 (103)
Distant recurrence	-0.303	S1: -1.317 S2: 0.496	lognormal	Campbell, 2011 (103)
Congestive heart failure	0.528	S1: 103.29 S2: 92.34	beta	Kirsch, 2000 (104)
Acute Myeloid Leukaemia	0.38	S1: 0.311 S2: 0.507	beta	Castejòn et al, 2018 (105)

5.3.4 Sensitivity Analysis

Probabilistic sensitivity analysis (PSA) was conducted using a Monte Carlo simulation to sample from distributions assigned to model input parameters (117). Probability distributions were fitted to each input parameter using mean values and shape parameters, according to the specific type of distribution. The beta distribution was used for binomial proportions, the Dirichlet distribution for multinomial proportions. The lognormal distribution was used for health state costs and test cost, while chemotherapy and toxicity costs are summarised as a weighted average of chemotherapy regimens, which include the use of normal and log normal

distributions. Baseline utility is dependent on the age group, while utility decrements use the lognormal distributions. Utility weights for CHF and AML states are fitted to beta distributions.

5.3.5 Probabilistic One-Way Sensitivity Analysis

While deterministic sensitivity analysis (DSA) remains an important method for characterising uncertainty and providing an easily accessible interpretation of CEA results, classic DSA methodologies may lead to wrong conclusions due to a lack of information regarding marginal effects, likelihood, and correlation between input parameters (117). Policy makers might be interested in specific values of a parameter, including whether these values are possible, their likelihood, and how they would affect a decision; this type of comprehensive information cannot be produced by DSA.

Probabilistic One-way Sensitivity Analysis was recently proposed as a solution to bridge this information gap (118). This approach generalizes the concept of "conditional net benefit" for evaluations comparing more than two strategies. The method also proposes the "conditional net benefit frontier" as a way to identify the most cost-effective option, out of a set of strategies, conditional on a specific value of an input parameter of interest.

In order to identify the values of key parameters that might change the decision regarding testing strategies, a probabilistic one-way sensitivity analysis (POSA) was carried out, according to the methods outlined in McCabe et al. 2020 (119).

After selecting alternative distributions for the cost of EER4 and the HR of patients testing as Responders with EER4, the values of these key parameters were fixed to a range of centiles (0.01,0.99, and from 0.1 through 0.9), and a Monte Carlo simulation was run for each selected centile and the net monetary benefit for each strategy were recorded. The resulting data is then used to produce a Cost-effectiveness Acceptability Frontier based on conditional net monetary benefits (conditional on the centile of the parameter cumulative distribution). The alternative distributions reflect higher uncertainty and include extreme/unlikely values, so that the analysis might identify the values for which a decision is likely to change. The alternative distribution for the cost of EER4 increases the baseline price to £1600,

with a minimum of £300 and a maximum of £8500. The alternative distribution for the Responders' HR maintains a base value of 0.35, but extends the range to the extremes of the Beta distributions, thus including scenarios where Responders are virtually no different from Non-responders in terms of risk of recurrence, and scenarios where Responders' risk of recurrence disappears. This range of values should reveal the minimum value of the ratio for which a testing strategy using EER4 is cost-effective compared to alternatives.

5.3.6 Value of Information Analysis

To quantify the benefits of reducing uncertainty with respect to selecting the most cost-effective strategy, the sensitivity analysis includes an estimate of the population Expected Value of Perfect Information (EVPI). The population EVPI estimate for EER4 is based on a Scotland annual incidence of 3700 new cases of ER+ breast cancer in females and the lifetime of the technology is assumed to be ten years, with a 3.5% annual discount rate applied.

5.3.7 Summary of Model Assumptions

The following list is a summary of important assumptions and implications used throughout the economic model.

- Using PREDICT survival estimates to derive recurrence rates is an accurate reflection of the recurrence rates that would be observed for this population.
- Using PREDICT chemotherapy benefit scores is an accurate proxy for the set of clinical decisions leading to administering chemotherapy.
- The modelled use of OncotypeDX, in conjunction with PREDICT, reflects the use of OncotypeDX in NHS routine practice (i.e. for cases where the chemotherapy benefit is uncertain).
- The modelled use of EER4, especially when used in conjunction with PREDICT, is an accurate approximation of how the test would be used in NHS routine operations.
- The estimated costs for treatment, and associated toxicities, reflect the true cost of treating a patient in NHS routine practice.
- Responders treated with neoadjuvant Letrozole enjoy a survival benefit over non-responders
- The health states used in the model accurately reflect the long-term patient pathway.

5.4 Results

5.4.1 Base case

The base case probabilistic analysis of the model produces an ICER of £6837 per QALY for therapy guided by OncotypeDX in conjunction with PREDICT (Strategy 2), over PREDICT alone. On the other hand, strategies that include EER4, both alone and in combination with PREDICT, produce negative ICERs: £-1109 per QALY and £-2078 per QALY, respectively. As it can be observed in the incremental cost-effectiveness plane, the majority of simulations of strategies that include EER4 fall in the south-east quadrant, suggesting that testing with EER4 is likely to produce better clinical outcomes at a lower cost, compared to using PREDICT alone. The mean cost per patient on PREDICT alone is £13,948 compared with £14,050 of OncotypeDX with Predict, £13,634 with EER4 alone, £13,246 with EER4 with PREDICT. The mean QALYs per patient for PREDICT alone is 7.69, compared with 7.71 with OncotypeDX, 7.98 with EER4 alone, 8.04 with EER4 and PREDICT. At a threshold of £20,000 and in terms of net monetary benefits, the probability of EER4 in conjunction with PREDICT being cost-effective is 86%, while OncotypeDX with PREDICT is 10%.

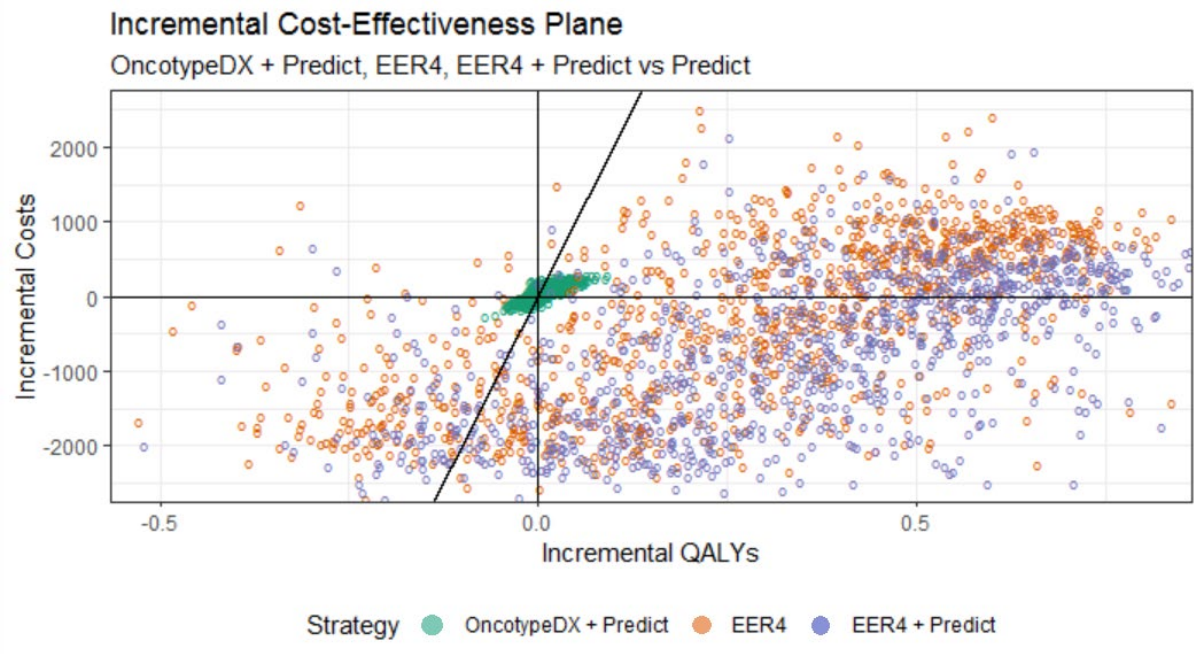


Figure 12: Incremental Cost-Effectiveness Plane

As it can be observed from the INMB density plot (Figure 13) in conjunction with the Incremental Cost-Effectiveness plane (Figure 12), the incremental benefits for OncotypeDX are heavily concentrated around the origin/zero-net benefit, while the two strategies featuring EER4 are distributed all along the scale (due to higher uncertainty surrounding costs and benefits) but show the highest density around the 10,000 mark.

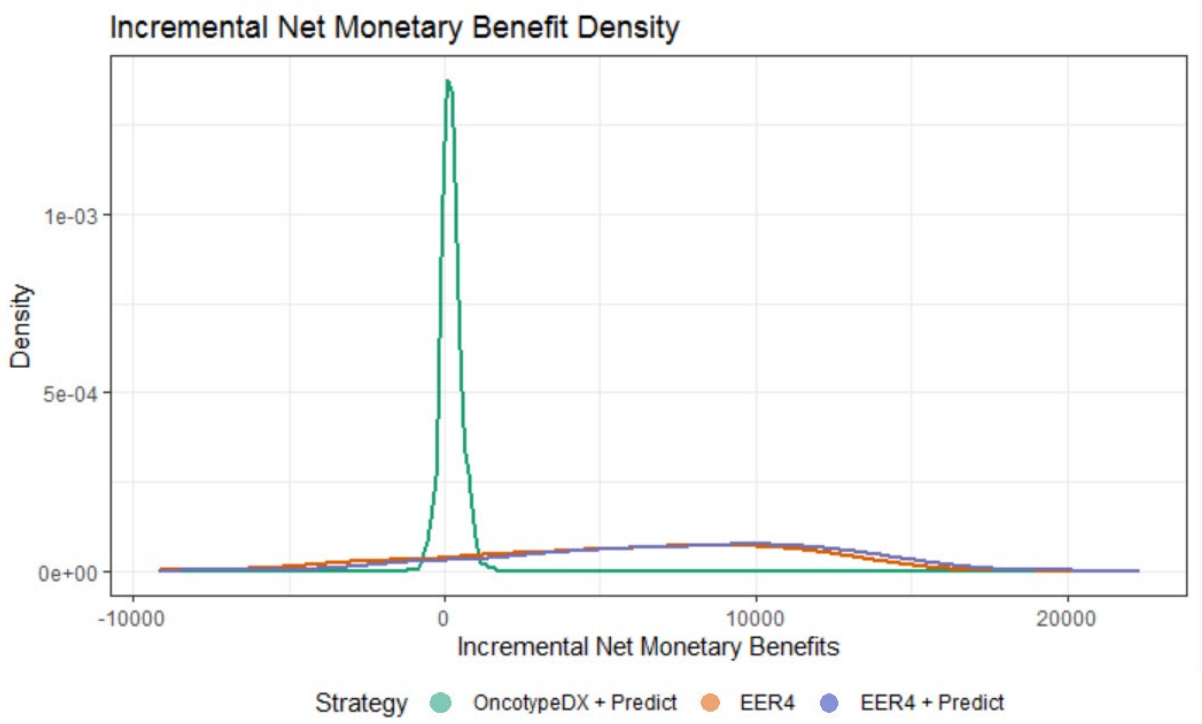


Figure 13: Incremental Net Monetary Benefit Density over Predict Breast alone

In the Cost-effectiveness Acceptability Curve (Figure 14), each strategy is individually evaluated against guiding therapy with PREDICT alone. Based on the proportion of simulations in which each alternative strategy's Incremental Net Monetary Benefits are positive, the analysis produces the probability that each alternative is cost-effective given a range of relevant values of the threshold.

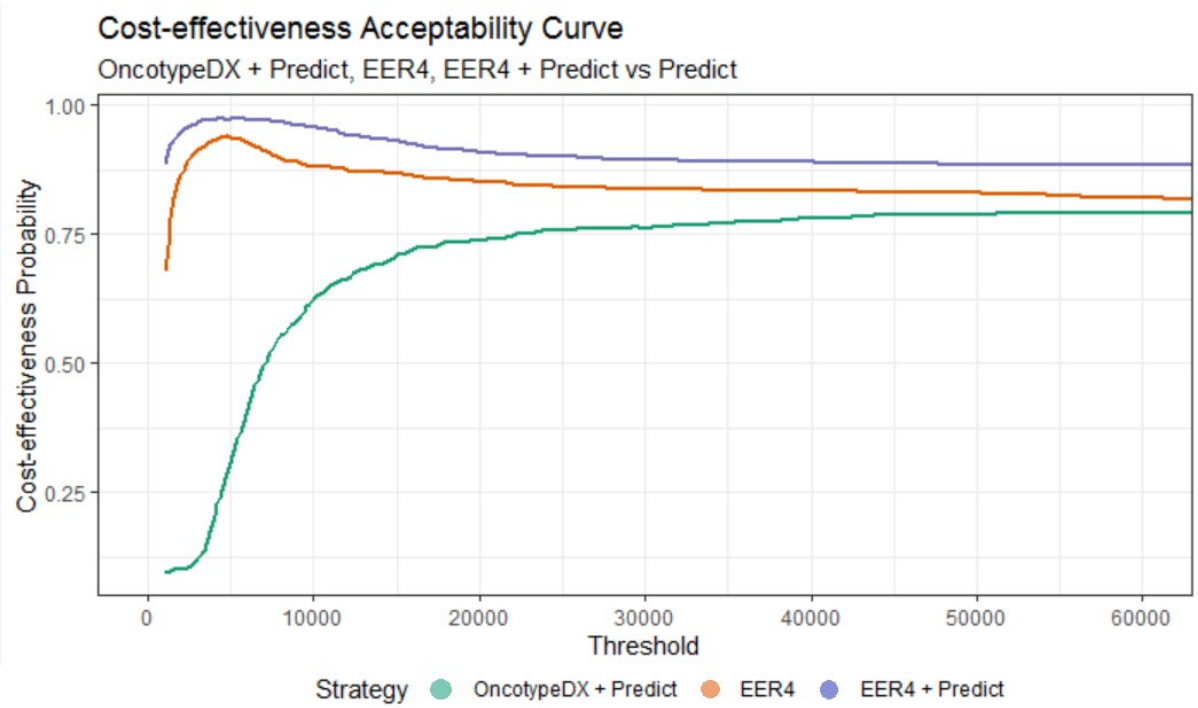


Figure 14: Cost-Effectiveness Acceptability Curve

At any value of the threshold, guiding therapy through EER4 with PREDICT is the most likely cost-effective strategy. This is also reflected in the Cost-effectiveness Acceptability Frontier (Figure 15), where all strategies are jointly evaluated on the basis of Net Monetary Benefits (i.e. for any value of the threshold, the sum of all strategies' probabilities of being cost-effective is equal to 1). EER4 in conjunction with PREDICT is the preferred alternative for any relevant value of the threshold.

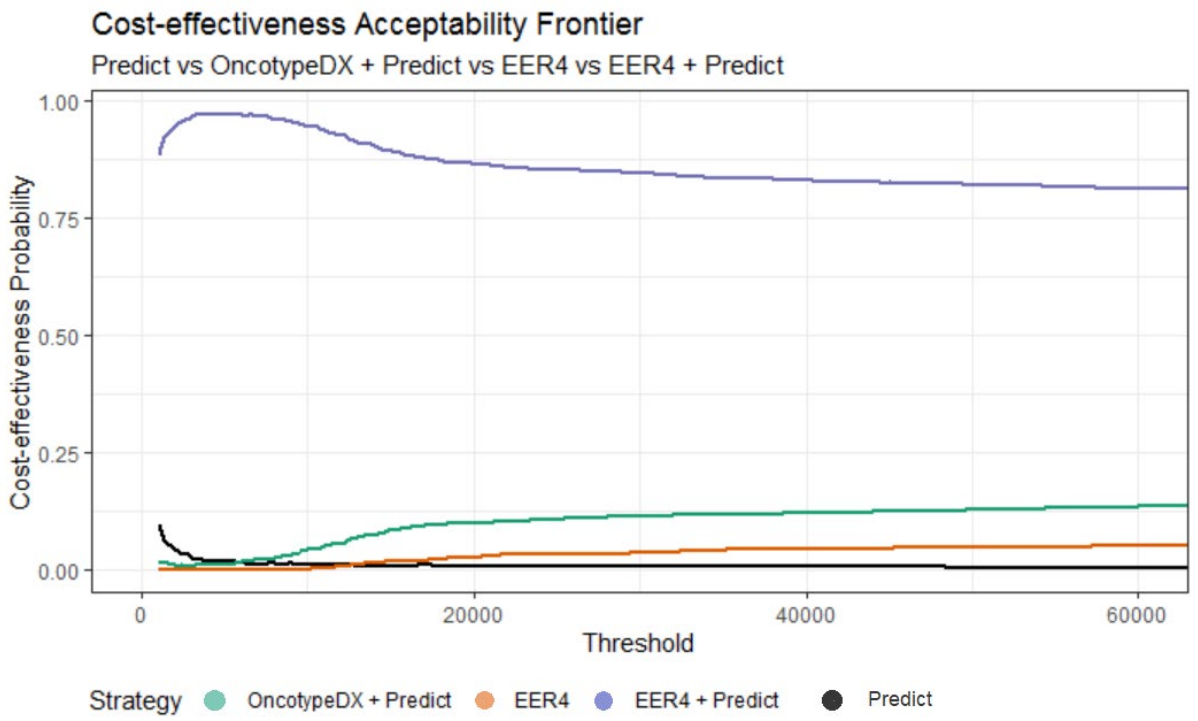


Figure 15: Cost-Effectiveness Acceptability Frontier

The simulated recurrence-free survival appears to be in line with the RFS estimated by PREDICT: a rule-out test performed on the distributions (simulated survival and estimated survival) resulted in a p-value of p=0.34, meaning that it cannot be excluded that the two sample are virtually drawn from an equivalent distribution. This effectively ensures the internal consistency of the simulation. Moreover, examining the simulated survival for each strategy (Figure 16) reveals that PREDICT alone displays the lowest survival, while EER4 alone and EER4 with PREDICT produce a virtually identical estimate.

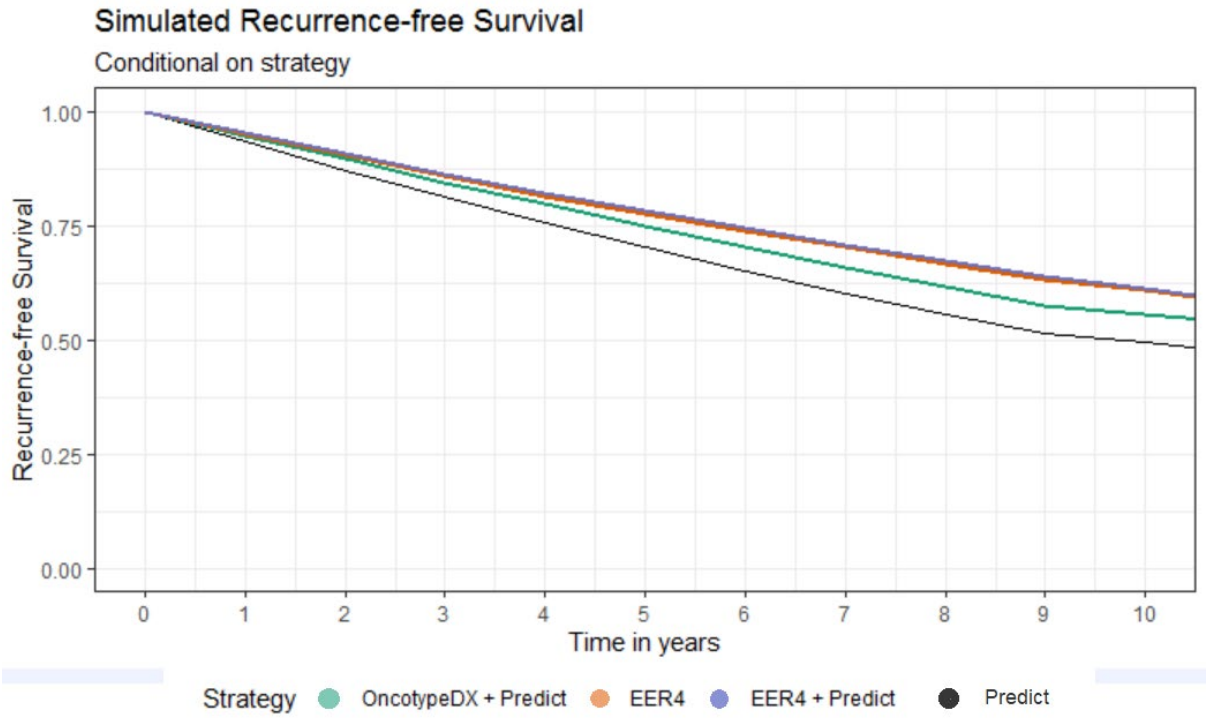


Figure 16: Simulated Recurrence-free Survival, by Testing Strategy

The individual Expected Value of Perfect Information is £238 at a threshold of £20,000 per QALY, and given the assumption of 3700 new cases of ER+ breast cancer case in Scotland every year, this translates to a 10-year population EVPI of £7,329,827. Figure 17 shows how the EVPI changes as the cost-effectiveness threshold is increased.

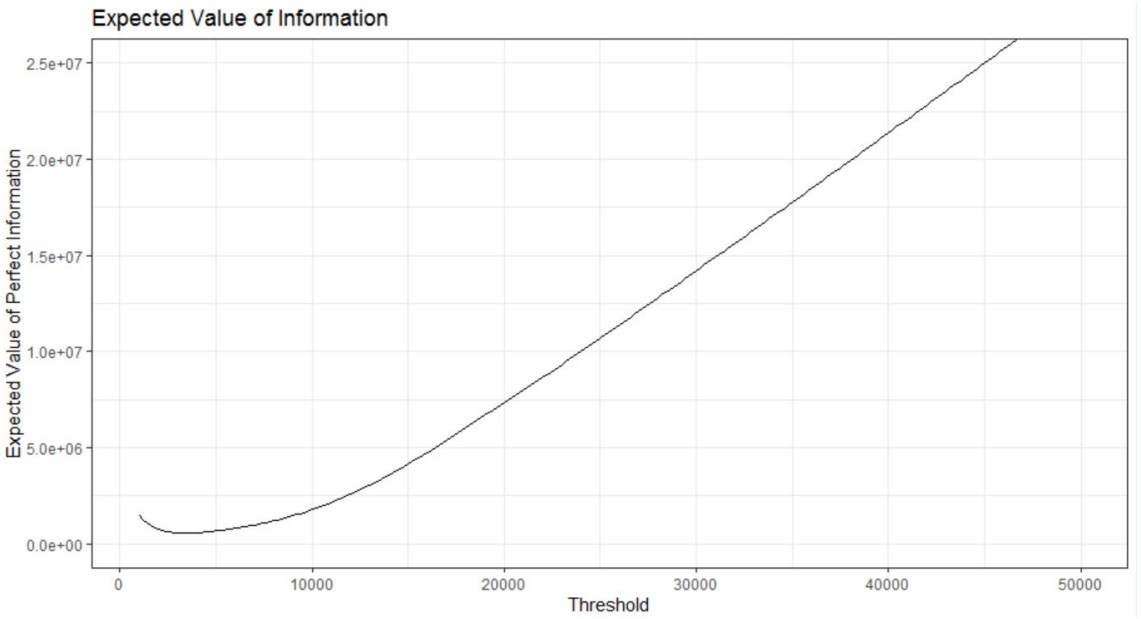


Figure 17: Expected Value of Perfect Information.

5.4.2 Probabilistic One-Way Sensitivity Analysis

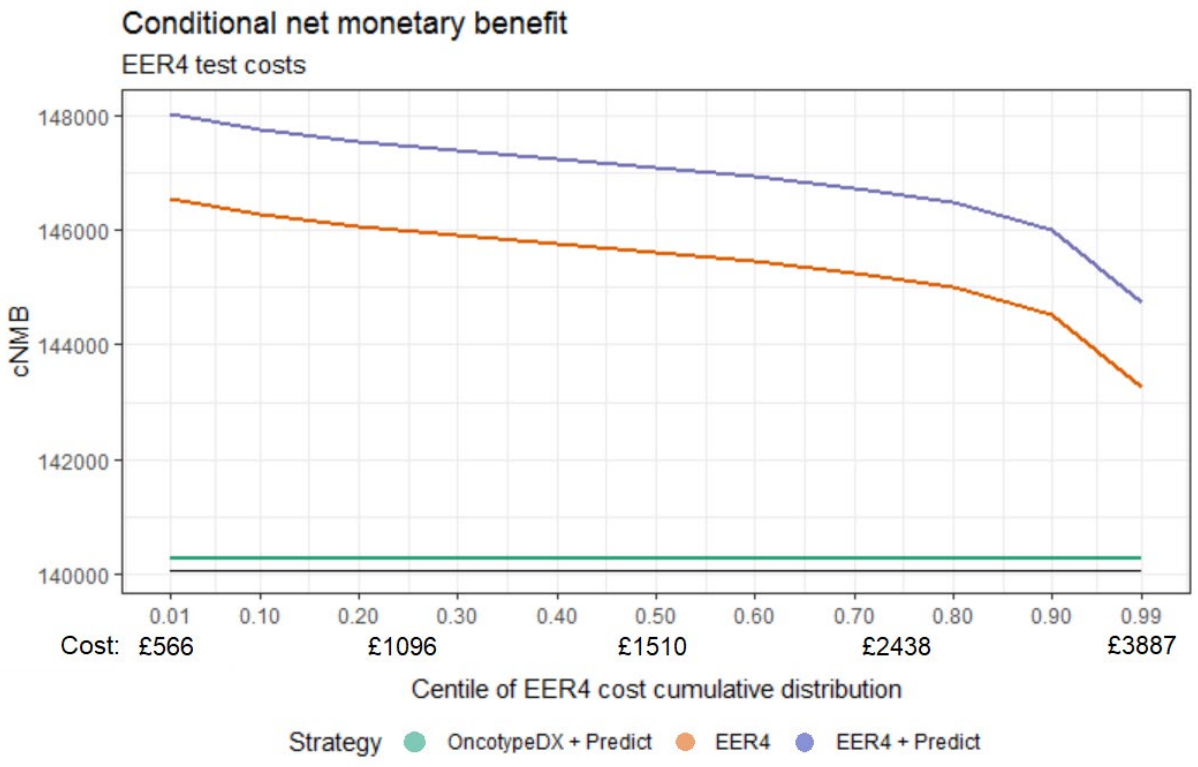


Figure 18: Conditional Net Monetary Benefit, Cost of Test

The CEAF conditional on the value of EER4 cost (Figure 18) displays the expected Net Monetary Benefit for each strategy at each centile of the cumulative distribution, assuming a £20,000 per QALY threshold. As expected, the NMB of PREDICT alone and OncotypeDX with PREDICT do not vary with the change in EER4 costs. The POSA suggests that strategies featuring EER4 are likely to be cost-effective compared to alternatives, even when the costs of the test could be far above any of the commercially available predictive assays in similar clinical settings.

Moreover, the POSA conditional on the value of Responders' Hazard Ratio (Figure 19) suggests that unless the HR in RFS between responders and non-responders is actually above 0.8, then any strategy featuring EER4 is likely to be cost-effective at a threshold of £20,000 per QALY.

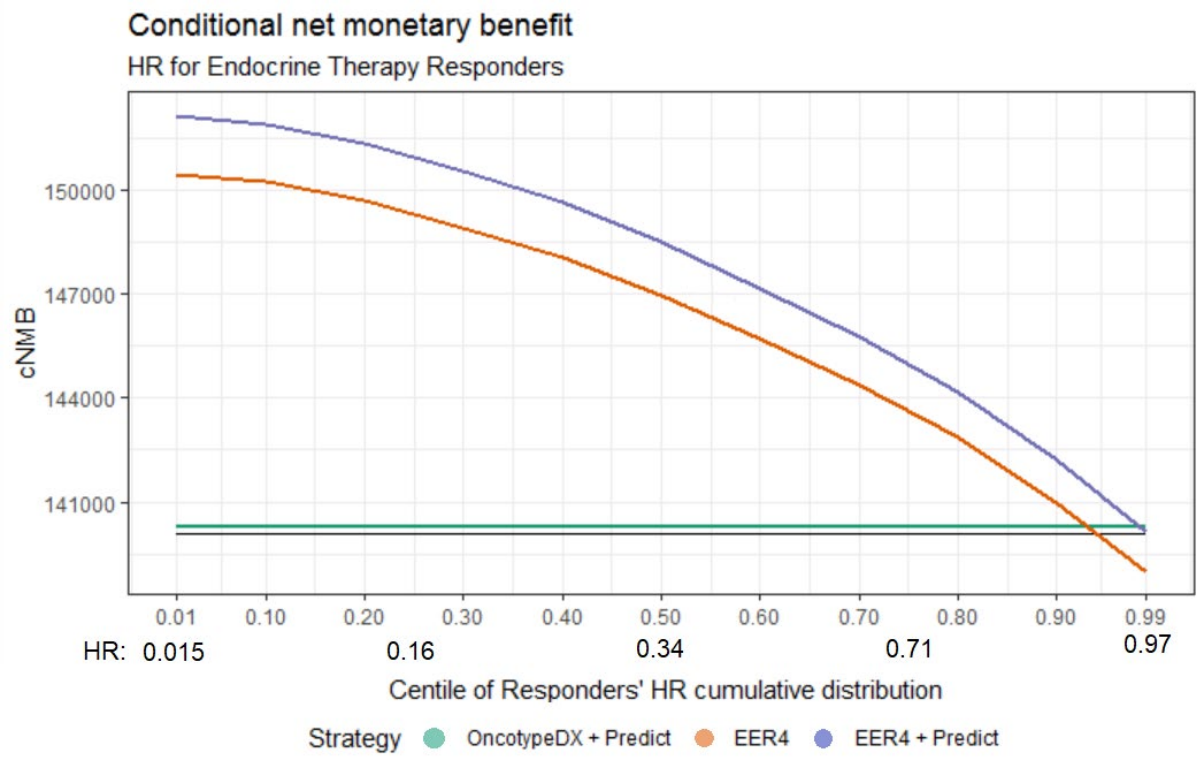


Figure 19: Conditional Net Monetary Benefit, Responder Hazard Ratio

5.4.3 Alternative Scenario: Higher Risk Patients.

As discussed in Chapter 3, the Secondary Cohort presents a larger share of higher risk patients, compared to the Primary Cohort. The simulation was run by populating

the model with the Secondary Cohort. The Incremental Cost-Effectiveness Plane (Fig. 20) presents a higher dispersion of values compared to Figure 12, while the density plot of net monetary benefits (Fig 21) indicates a similar pattern to its counterpart in the Primary Cohort (Fig 13), with the incremental net monetary benefits of OncotypeDX + PREDICT concentrating around 0, and the benefits of strategies with EER4 distributed across a wider interval and peaking around the 10,000 mark. The Cost-Effectiveness Acceptability Curve reports the same ranking of alternatives as in the simulation informed by the Primary cohort (Figure 22).

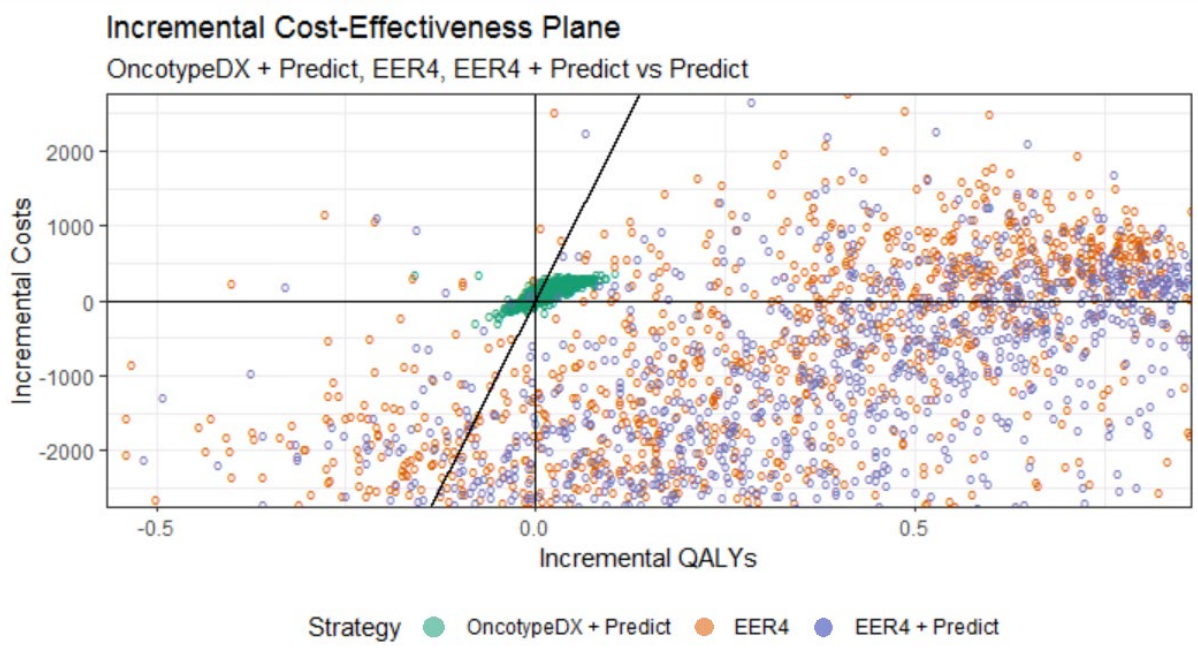


Figure 20: Incremental Cost-Effectiveness Plane, Secondary Cohort

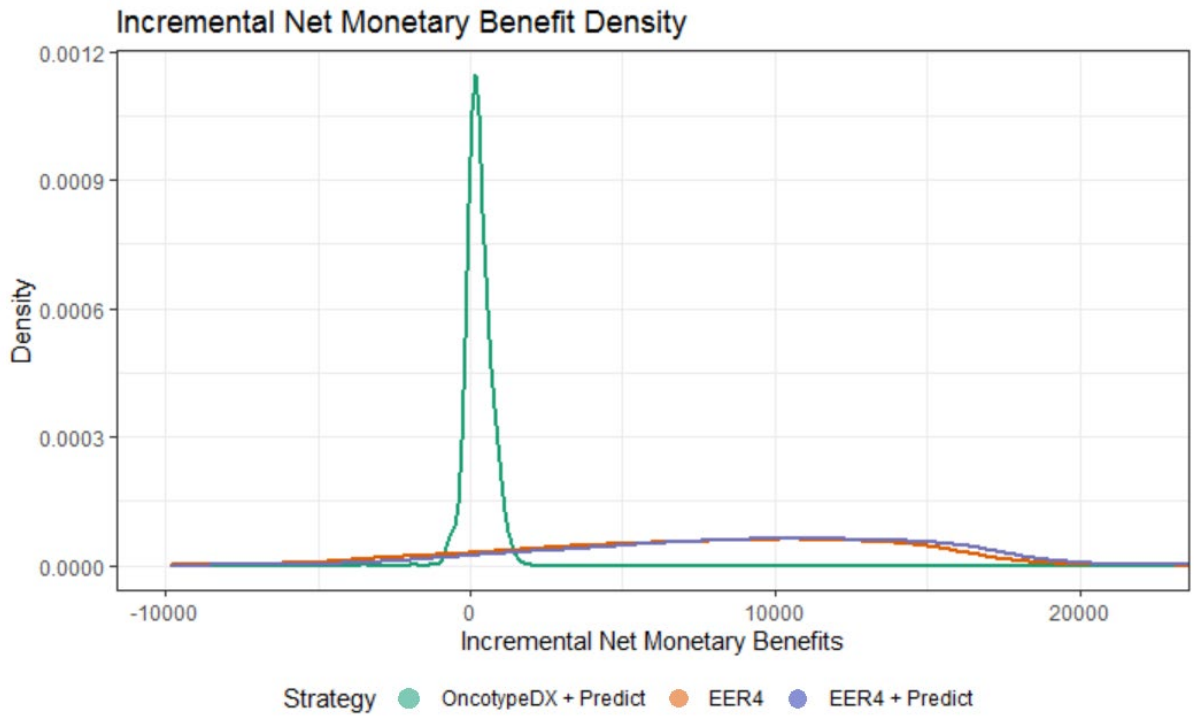


Figure 21: Density Plot of Incremental Net Monetary Benefit, Secondary Cohort

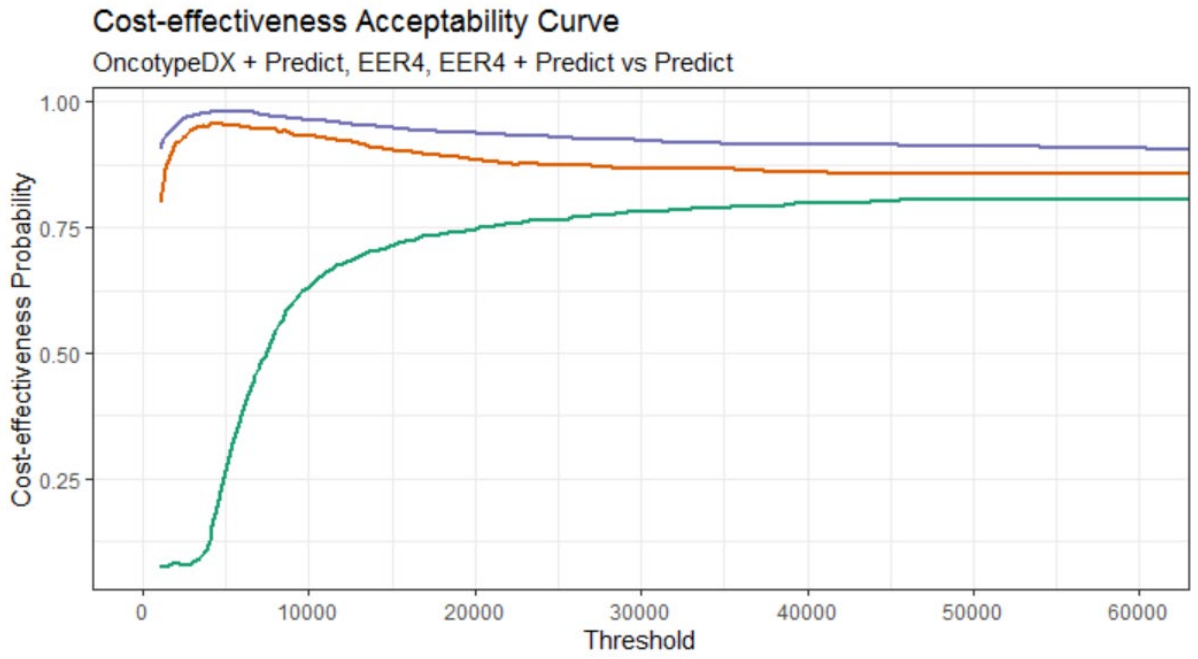


Figure 22: Cost-Effectiveness Acceptability Curve, Secondary Cohort

5.5 Discussion

The results of the simulation strongly indicate that testing strategies utilising EER4 have the potential for delivering better outcomes to early ER+ breast cancer patients than the current standard of care. While the estimates are conditional on the uncertainty surrounding the clinical performance of the 4-gene signature, they suggest that the introduction of EER4 into routine care could deliver marginally superior survival and quality of life at a lower cost, compared to PREDICT alone.

The two main sources of uncertainty are the operational cost of the 4-gene signature, and the actual benefit (expressed as a hazard ratio) of identifying Endocrine therapy responders early. The results from the Probabilistic One-way Sensitivity Analysis point to positive net monetary benefits even in extreme values of the probabilistic distributions of the test cost and hazard ratio. Consequently, even if the current estimates for the test cost and hazard ratio are to be revised in view of new research, the predictions formulated by the simulation should nevertheless hold in terms of ranking of alternatives.

The estimated benefits of OncotypeDX with PREDICT are limited both in magnitude and in variability: this appears to be a consequence of the relatively small role of the test in the simulation: only patients with a PREDICT chemotherapy score between 3% and 5% are tested, thus limiting the number of tests administered to around 15% of the simulated cohort. The decision to simulate OncotypeDX in this manner might not fully reflect its actual clinical use and testing rates, although it is in line with the findings of Chapter 4. In addition, the simulation assumed a lower price for OncotypeDX compared to the listed commercial price.

In line with expectations for economic evaluations of decision-making tools, the simulation estimates relatively small marginal benefits compared to the standard of care: treatment options are the main sources of potential health benefits, and since treatment options are shared among testing alternatives, the test benefits originate from patients switching treatment option under the recommendation of the test. Thus, it appears that the main factors driving the ICERs and net monetary benefits are the costs of testing, including costs net of treatment savings.

Interestingly, even though PREDICT alone is assumed to be a costless testing strategy (net of treatment), the base case results of the simulation estimate EER4 to

be less expensive in absolute terms, reflecting the fact that fewer chemotherapies are administered compared to PREDICT alone.

Molecular testing in breast cancer offers the opportunity to improve survival of patients by selecting the most appropriate course of treatment: sparing chemotherapy when unnecessary and identifying early those who might benefit from it the most. In turn, economic evaluation has the fundamental role of guiding potential decision-makers towards a more cost-effective use of resources through the examination of new technologies.

The analysis suggests that new technologies based on the measurement of pre- and on-treatment markers can potentially deliver better clinical outcomes at a lower cost, but further research and data collection is needed to reduce the uncertainty of their benefits and costs, and to establish exactly their clinical guidelines.

The results of this cost-effectiveness analysis hinge on specific assumptions for their general validity: in particular, that PREDICT is reflective of the current standard of care, and that the semi-Markov process used in the simulation is a sound approximation of the treatment pathway. Following guidelines, necessary steps have been taken to reduce potential weak spots in the simulation and bias in the data. Extensive sensitivity analysis has been carried out to ensure the proper characterization of uncertainty surrounding the input parameters, and appropriate confidence intervals were selected when dealing with cost parameters of EER4-based strategies.

This economic evaluation adds to the growing list of cost-effectiveness analyses that have shown the potential of using Real-World Data, in particular populating a model with routinely collected patient data. This relatively fast way of accessing large amounts of data, coupled with the potential for expanding this simulation, offers the possibility of generating an analytical platform capable of performing the early evaluation of decision-making tools for any breast cancer subtype.

6. Modelling Treatment before Surgery: the Impact of a Predictive Test on Breast-Conserving Surgery Rates

6.1 Introduction

The previous chapter explored the effects of introducing a predictive molecular marker to clinical care. In particular, it investigated the impact that EER4, which predicts response to neoadjuvant endocrine therapy and also long-term outcomes, can have on chemotherapy treatment rates. This chapter explores how a technology such as EER4 can affect patient in the pre-operative settings, and how it changes breast-conserving surgery rates. By combining the effects that this technology can have on a patient pathway before and after surgery, the analysis estimates the overall costs and benefits that a predictive molecular marker can have on routine NHS operations.

Due to the experimental nature of the analysis, the simulation platform required has to be flexible and intuitive: a Discrete-Event Simulation (DES) model is able to fill this role due to its high customisability and the potential to model complex situations at an individual level, compared to Markov models or decision-tree models (120). Instead of health states, as seen in Chapter 5, a DES model makes use of trajectories: a simplification of patient pathways. On these trajectories, patients are seen as entities experiencing events, such as a recurrence, test decisions, or surgery. The simulation output takes the form of an event list, which details the type of events experienced, the simulated time they happened, and the path that the virtual patient or disease has taken (121). This event list can then be used to compute the resource utilisation and health outcomes, which in turn provide an estimate of cost-effectiveness.

As mentioned, the analysis will be experimental and explorative, because of the limited data available pertaining this particular area, and the issue of simulating pathways which are not present in routine care. Therefore, the final cost-utility estimates are to be interpreted within this context and its assumptions.

6.2 Methods and Data

A time-dependent, non-resource-constrained, individual discrete event simulation model was developed to estimate the resource utilisation and health outcomes of three alternative strategies for guiding treatment decision in a representative Scottish cohort of ER+ early breast cancer patients. The model was developed within R Programming Language (14) and the software package “simmer” (122). The full code for executing the model is available in Appendix 2. The three strategies are described as follows:

1. Standard of Care: The patient is diagnosed and then scheduled for surgery. Depending on the size of the tumour, the type of surgery is assigned: if less than or equal to 2 cm, the patient undergoes breast-conserving surgery (lumpectomy), if the tumour is larger than 2 cm the patient undergoes radical mastectomy and immediate reconstructive surgery. In the adjuvant setting, the patient proceeds with the strategy “PREDICT only” as described in Chapter 5.
2. Neoadjuvant Letrozole for All: The patient is diagnosed and then treated with Letrozole until surgery. At the end of treatment, the patient undergoes surgery, and the type of surgery is determined by the size of the tumour after treatment. In the adjuvant setting, the patient proceeds with strategy “PREDICT only” as described in Chapter 5.
3. EER4 for all: The patient is diagnosed, treated with Letrozole for 14 days, and tested for response status. If the patient is a Predicted Responder, they continue treatment until scheduled for surgery, and the type of surgery is determined by the size of the tumour after surgery. If they test as Predicted Non-responder, the patient is scheduled for surgery and surgery is determined by initial tumour size. In the adjuvant setting, the patient proceeds with the EER4 only strategy as described in Chapter 5.

Figure 23 provides a visual breakdown of the strategies and model structure.

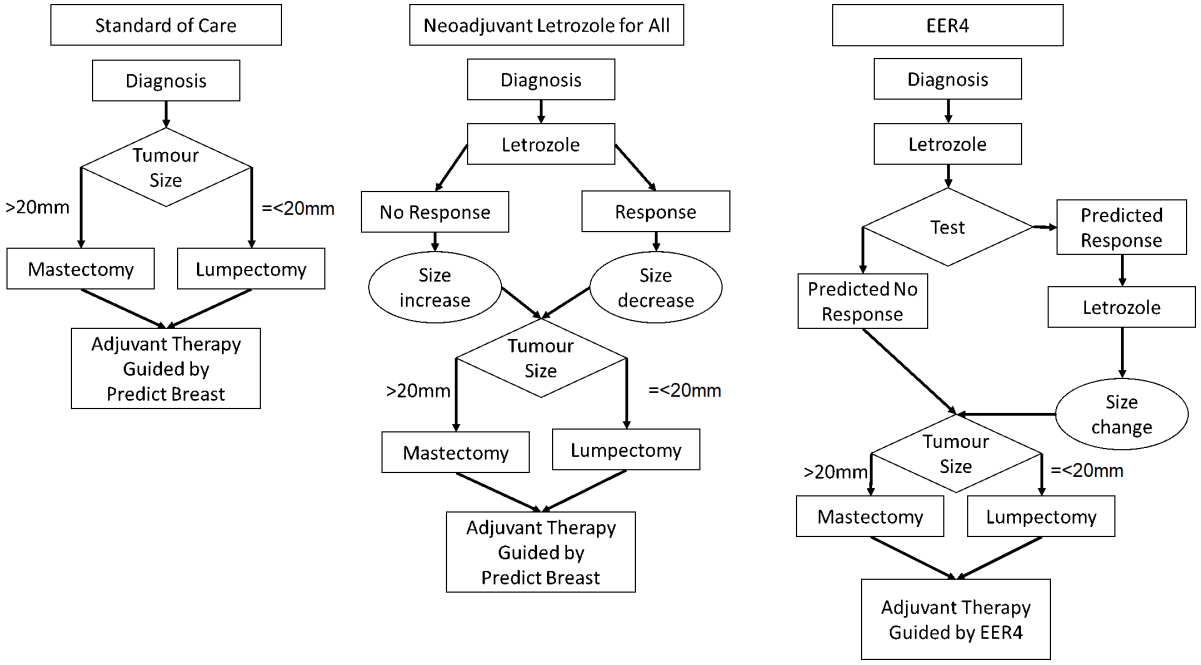


Figure 23: Flow diagram of Strategies and Model Structure

The model was developed from the payer perspective of the National Health Service and assumes a life-time horizon by calculating cost and outcomes of the neoadjuvant setting and surgery, and then integrating them with the results found with the CEA in Chapter 5. The model assumes that breast-conserving surgery (lumpectomy) provides a marginal increase in the quality of life of the patient, over the alternative of radical mastectomy. Once converted in QALY, the effect of lumpectomy is discounted yearly at 3.5% and added to the total number of QALYs calculated in Chapter 5, according to the appropriate strategy. The costs of neoadjuvant treatment and surgery are incurred immediately and are not discounted. The resulting costs are added to the sum estimated in Chapter 5, according to each strategy. The cost of testing with EER4 is consequently already included in the estimates from the Markov model. All prices were captured in 2016-pound sterling.

As for the model input parameters, rates of decrease and increase in tumour volume based on therapy response were obtained from the Edinburgh Neoadjuvant Cohort used for developing EER4 (68). Tumour volume distribution was based on the NHS Lothian cohort described in Chapter 3. Waiting times from diagnosis to surgery were extracted from Redaniel et al. 2013 (123). There is no definitive consensus on the duration of neoadjuvant treatment, however most patients are treated preoperatively for 3 to 6 months (124, 125), and a probabilistic distribution was fitted to

appropriately reflect this uncertainty and variation. The cost of neoadjuvant endocrine therapy was obtained from the BNF (99), while the cost of surgery is extracted from Grant et al. 2018 (126) and NHS Reference Costs (100). The quality-of-life difference between undergoing lumpectomy and mastectomy is taken from Norum et al. 1997 (127). The model assumes that 70% of this patient population responds to neoadjuvant Letrozole, and that EER4 operates with 96% sensitivity and 94% specificity, as per Turnbull et al 2015 (68).

To reflect the overall uncertainty of the model, probabilistic sensitivity analysis was conducted using a Monte Carlo simulation to sample from distributions assigned to model input parameters. Similarly to the input parameters of the Markov model in Chapter 5, probability distributions were fitted to each input parameter using mean values and shape parameters, according to the specific type of distribution. A list of the input parameters and associated distributions can be found in table 11.

Table 11. Model Input Parameters

Parameter	Base case	Shape Parameters	Notes	Source
Tumour size at diagnosis in millimetres	25.1	Log normal S1: 3.1 S2: 0.49		Chapter 3
Decrease in tumour size for endocrine therapy responders	15.5	Log normal S1: 2.63 S2: 0.54		Turnbull et al 2015 (68)
Increase in tumour size for endocrine therapy non-responders	12.1	Log normal S1: 2.33 S2: 0.58		Turnbull et al 2015 (68)
Waiting time from diagnosis to surgery (days)	22.5	Log normal S1: 3.08 S2: 0.3		Redaniel et al 2013 (123)
Time spent on neoadjuvant treatment (days)	150	Normal S1: 150 S2: 30		(124, 125)
Cost of breast conserving surgery	£ 6,157	Log normal S1: 8.699 S2: 0.2		(99,100,126)
Cost of mastectomy plus breast reconstruction	£ 10,939	Log normal S1: 9.3 S2: 0.03		(99,100,126)
Cost of EER4-guided adjuvant therapy	£ 13,634	Log normal	Total cost of using EER4 to guide adjuvant therapy, as	Chapter 5

		S1: 9.52 S2: 0.06	derived in the semi-Markov model	
Cost of SOC-guided adjuvant therapy	£ 13,948	Log normal S1: 9.543 S2: 0.05	Total cost of SOC-guided adjuvant therapy	Chapter 5
QALY difference between BCS and mastectomy	0.03	Log normal S1: -3.056 S2: 0.5		Norum et al 1997 (127)
QALY of EER4-guided adjuvant therapy	7.98	Log normal S1: 2.076 S2: 0.04	Total QALY of EER4-guided adjuvant therapy	Chapter 5
QALY of SOC-guided adjuvant therapy	7.69	Log normal S1: 2.039 S2: 0.03	Total QALY of SOC-guided adjuvant therapy	Chapter 5

Given the small timeframe where the discrete event simulation operates, several simplifying assumptions were made regarding risks and procedures. The model assumes that patients undergoing lumpectomy have the same risk of recurrence of those undergoing mastectomy, all things being equal. Moreover, risk of death before surgery due to breast cancer or other competing risks were not considered. The model assumes that all surgeries, regardless of type, are always successful and the patient incurs no complications. Finally, the patients are assumed to have perfect compliance with respect to treatment decisions, and that breast reconstruction surgery takes place immediately after every mastectomy.

6.3 Summary of Model Assumptions

The following list is a summary of important assumptions and implications used throughout the economic model.

- 70% of the population responds to Letrozole vs 30% non-responders
- Treating responders with Letrozole will decrease tumour size, according to time spent on therapy and probabilistic rate of change
- Non-responders' tumour size will increase, according to time spent on therapy and probabilistic rate of change
- Medical imaging or other means to measure changes in tumour volume after selecting treatment are not considered
- Letrozole is the only neoadjuvant systemic therapy course
- The threshold for selecting mastectomy or lumpectomy is 20mm
- Surgery failures or re-excision are not considered, as well no sequential surgeries
- Mastectomies are immediately followed by breast reconstruction
  Equivalent risk of recurrence between mastectomy and lumpectomy

6.4 Results

The probabilistic base case yields 7.89 QALY per patient for the Standard of Care, 7.93 for Neoadjuvant Endocrine for all, and 8.27 for EER4, at a respective average cost per patient of £24216, £24147, and £23755. The costs are inclusive of the cost of surgery, neoadjuvant therapy where appropriate, testing, and the cost of adjuvant therapy as estimated in Chapter 5. Similarly, the health outcomes are a combination of the QALY as calculated in Chapter 5 and the QALY estimated through the Discrete-Event Simulation. Strategy 2 (Neoadjuvant endocrine therapy for all) produces an ICER of -1079£/QALY over the Standard of Care, while EER4 an ICER of -1208£/QALY, as both strategies are more effective and less expensive than the Standard of Care. Assuming EER4 is a perfectly accurate test (i.e. sensitivity and specificity both equal to one), the ICER decreases to -1240£/QALY.

In terms of mastectomy displacement in favour of lumpectomy, Neoadjuvant endocrine therapy for all produces a delta of 8.7%, while EER4 a delta of 16.3%, with average savings per patient of £416 and £779, respectively. Using a cost-effectiveness threshold of 20,000£/QALY, EER4 has a 79.7% probability of being cost-effective against the Standard of Care, while neoadjuvant endocrine therapy for all 28.3%. Figure 24 shows a sample of 1000 simulations plotted on the incremental cost-effectiveness plane (in red, Neoadjuvant Endocrine Therapy for All, in blue, EER4).

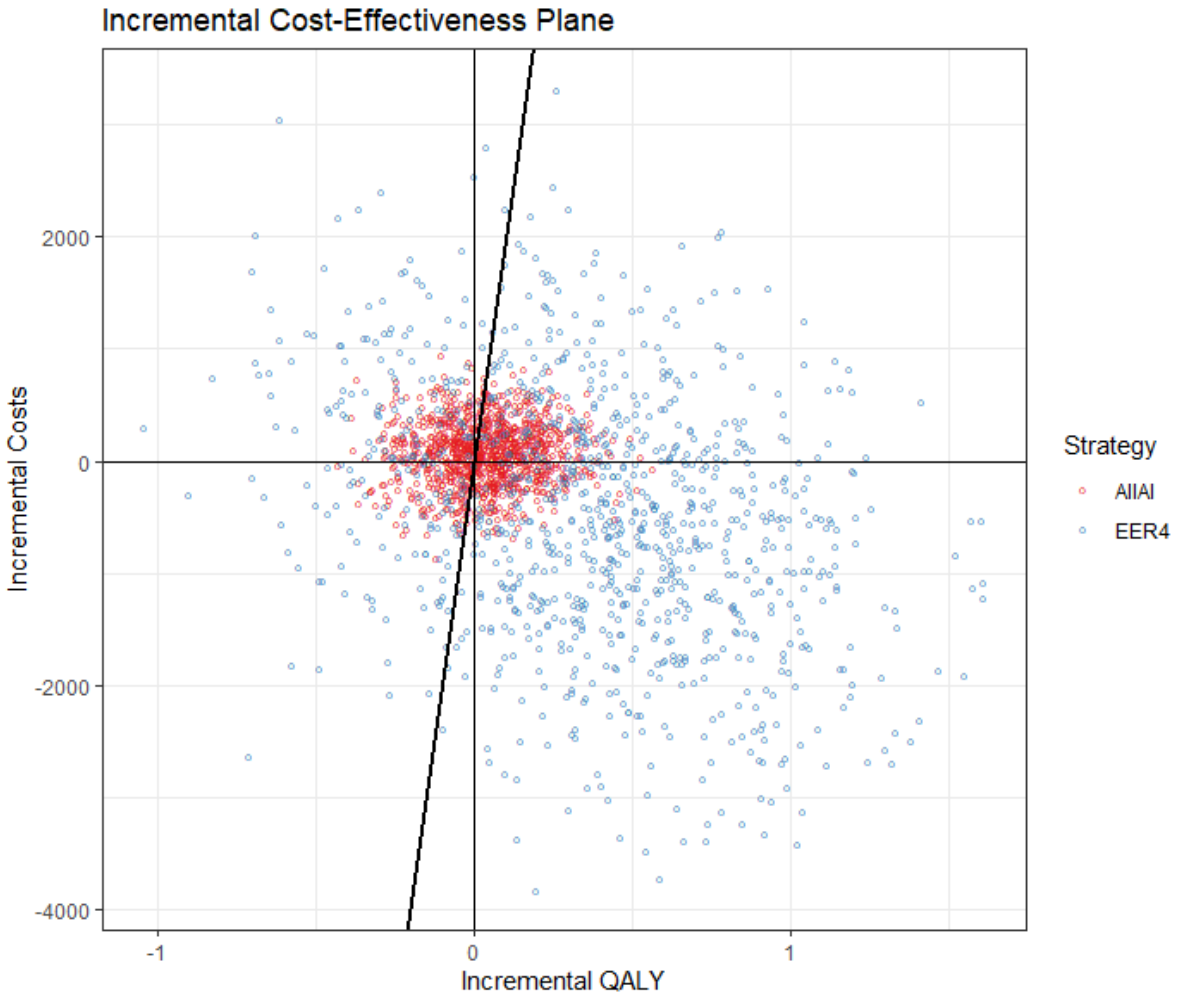


Figure 24: Incremental Cost-Effectiveness Plane

The deterministic base case, which utilises the mean of the probabilistic distributions of the input parameters, yields an ICER of -1521£/QALY and -1856£/QALY for Neoadjuvant endocrine therapy for All and EER4, respectively. Figure 25 displays the results of the deterministic one-way sensitivity analysis on the EER4 ICER: the extent to which the ICER is affected by varying key input parameters one at a time. Input parameters are varied by adding or subtracting 10% of their mean value. The size of the bar displays the extent to which the ICER is affected.

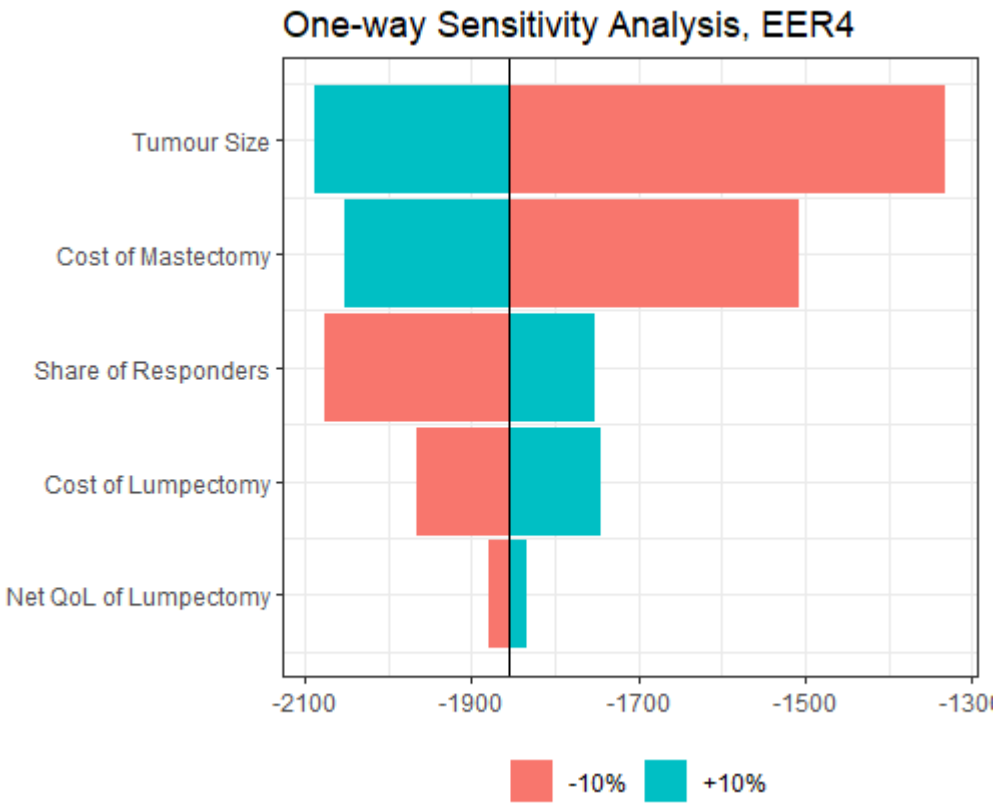


Figure 25: One-way Sensitivity Analysis, EER4

Tumour size and the cost of mastectomy are the input parameters that affect the deterministic base case ICER the most: by either reducing the size or the cost, the ICER moves towards the zero, making the potential advantages of EER4 over the Standard of Care less competitive.

6.5 Discussion

The results of the discrete-event simulation corroborate the results of the cost-effectiveness analysis presented in Chapter 5. Introducing EER4 in routine care has the potential for better health outcomes and reduced costs compared to the standard of care. Under the model assumptions, testing with EER4 reduces the overall number of mastectomies, thus favouring breast-conserving surgery, which in turn exhibits lower costs and marginally better quality of life. Furthermore, assuming a perfect test changes the probabilistic ICER by less than 3%, indicating that further

research to improve the accuracy of the test would not likely be a cost-effective use of resources.

The model assumes equivalent risk of recurrence between mastectomy and breast-conserving surgery. Historically, women undergoing breast-conserving surgery likely had a higher risk of recurrence compared to radical mastectomy (128), but recent evidence shows equivalent risks or even superiority in terms of recurrence, survival, and quality of life (129, 130).

Moreover, it appears that giving every patient neoadjuvant endocrine therapy might be a superior strategy compared to the Standard of Care. These results should be interpreted with caution, given the strong assumptions under which the DES model operates. Firstly, it should be noted that effectively there are virtually only two pathways that the patient can follow: either proceeding straight to surgery or being treated with Letrozole before surgery. Other types of treatment, such as neoadjuvant chemotherapy, or switching to other hormone medications have not been considered. Secondly, the model is mainly focused on the displacement of mastectomies in favour of breast conserving surgery, thus not taking into account other factors that might influence the cost-effectiveness of the testing strategies considered here; for example, some patients might achieve pCR with neoadjuvant therapy alone. Furthermore, the simulation has not examined strategies that include devices that can aide with treatment decision: in the case of neoadjuvant endocrine therapy for all, a non-responding patient might interrupt Letrozole earlier than expected because medical imaging might indicate that the tumour is not responding.

As seen in Figure 25, tumour size has the largest effect of all parameters on the ICER and this is due to the combined effect of two assumptions: firstly, the way tumour size changes based on response status and therapy, and secondly, the assumption that mastectomies are followed by breast reconstruction, and are thus more costly than breast conserving surgeries. For the first assumption, as long as the starting tumour size is relatively close to the surgery threshold of 20mm, the simulation will output a greater lumpectomy rate overall due to the higher proportion of responders and the fact that the tumour size of a treated responder can only decrease. A higher proportion of lumpectomies affects the overall costs, due to the cost differential between breast conserving surgery and mastectomy, with the latter

being more expensive as the model assumes 100% rate of breast reconstruction following surgery. This assumed rate is higher than observed rates: in Scotland the rate of breast reconstruction surgery following mastectomy is 30%, as of 2017 (131). Yet, the model assumes a 100% surgery rate for simplicity and for the limitations of HRQoL difference estimates available in the literature. As it can be seen in the asymmetry of the effect of tumour size on the ICER, starting values for size that are above the threshold have a smaller impact (as there are more responders, and tumour size will decrease for responders), compared to values below the threshold, where a higher proportion of the non-responders is more likely to receive breast conserving surgery, even if the tumour increases in size. The results of this interaction between assumptions might create a competitive advantage for EER4-guided therapy which is unlikely to be observed in reality.

While the primary focus of the DES model is on surgery displacement, the analysis did not include the potential for unsuccessful surgeries or re-excisions. As of 2017, Scotland's re-excision rate is on average across all health boards 14.5% (131). Moreover, the assumption of a fixed threshold for tumour size that determines whether a patient can be treated with breast conserving surgery or mastectomy is highly likely to oversimplify the reality of this kind of treatment decision. While breast conserving surgery for lesions greater than 40mm is unlikely, there is no clear and fixed cut-off value, as the decision depends on several factors, including size of the lesion relative to breast volume, position of the lesion, comorbidities and frailty of the patient. Given the experimental and explorative nature of this model, adding the effect of less-than-perfect surgeries might have concealed the pure effects from the testing strategies, but further and more detailed analysis on this type of biomarkers should reasonably include these effects.

The difference between the estimates of the probabilistic base case and the deterministic base case, while small in magnitude and inconsequential to strategy ranking, still reveals the inaccuracy and potential bias of not including the full range of values that an input parameter might take on, which in this case causes an overestimation of the saving generated by the alternatives. Nevertheless, identifying which input parameters affect the ICER the most and in which direction can assist in defining the most cost-effective role for the technology when introduced to routine care.

The model could benefit from further improvements: as it is, the rate at which the tumour volume decreases or increases due to therapy response is only time-dependant. Both tumour volume and tumour rate of change are drawn from probabilistic distributions, but the draws are independent; more realistically, the two distributions could be correlated, making the rate of volume change time-dependant and volume-dependant.

Moreover, using the results from the Markov simulation and combining them with the estimates of the DES presents some issues. PREDICT is used by the Markov model to simulate the current standard of care, by approximating all of the investigative procedures and the decisions that the patient care team would undertake to determine which treatment is most appropriate for the patient condition. Yet, PREDICT is not currently validated for patients treated neoadjuvantly, which is the specific population that this study focuses on. Efforts in trying to validate PREDICT for patients treated neoadjuvantly using data from the Scottish Cancer Registry data are currently underway, but accuracy rates close to PREDICT in the post-operative setting have yet to be achieved (132).

Overall, the simulation indicates that predictive tests of response to neoadjuvant therapy are likely to be a cost-effective alternative to either immediate surgery or unguided therapy for the treatment of ER+ early breast cancer. The results of the simulation, while robust, are to be viewed in the context of the assumptions described in the methods. Prospective data is needed to reduce uncertainty and accurately define the operational parameters of the molecular test, yet the early CEA indicates that potential savings and better health outcomes are likely to be expected from this kind of technology once introduced in routine operations.

7. Conclusion

This thesis had three aims: firstly, to provide an overview of the current research landscape for predictive biomarkers within the neoadjuvant context of breast cancer; secondly, to estimate the likely impact on the NHS of introducing this type of technology to routine care; lastly, to generate cost-effectiveness evidence through the use of real-world data, in the form of patient records.

7.1 Summary of Findings

This section provides a summary of the findings for each chapter:

- In Chapter 2, the literature review investigated the expanding research interest in novel biomarkers that can predict breast cancer response to neoadjuvant treatment. The review found a wide range of technologies and approaches, with varying degrees of accuracy and predictive capabilities. The review highlighted the need of conducting early cost-effectiveness analysis for this type of technologies to guide and select the most promising biomarkers as potential decision-making tools in clinical practice.
- Chapter 3 provided the context surrounding the use of real-world data in economic evaluation. The chapter presented the dataset used by the modelling platform and how the dataset was derived, along with a description of the data sources. The chapter also offered an explanation of the reasons why the dataset was divided in two cohorts, based on risk stratification and historical clinical practices. Details on the risk stratification are provided, which include the use of PREDICT and of the proxy score of OncotypeDX, GR-PR.
- In Chapter 4, the use of real-world data is demonstrated through a budget and decision impact analysis of OncotypeDX in NHS Lothian. The findings indicated that testing with OncotypeDX likely reduces the number of adjuvant chemotherapies given to breast cancer patients, but the current commercial cost of the 21-gene panel might be well above parity in terms of treatment displacement. While the results were in line with previous budget impact studies, the analysis was greatly affected by low-testing rates.
- In Chapter 5, the semi-Markov model provided the early cost-effectiveness estimates of adopting a biomarker, which is predictive of neoadjuvant therapy response, in NHS routine operations. The probabilistic estimates of the ICER

indicated potential net savings compared to the standard of care. The probabilistic one-way sensitivity analysis strongly suggested that the estimates for incremental net monetary benefits are likely to be robust for an appropriately wide range of values with respect to the cost of the biomarker and the benefit of identifying early which patients will respond to therapy, and which patients do not. The findings also included the results of the analysis when run with a higher-risk cohort.

- In Chapter 6, a Discrete Event Simulation compared the displacement of mastectomies in favour of breast conserving surgery under three different strategies: standard of care, neoadjuvant endocrine therapy for all, and neoadjuvant endocrine therapy guided by EER4; the results were then integrated with the cost-effectiveness estimates from the semi-Markov model. The findings corroborated the results of Chapter 5, by indicating the potential for net cost saving compared to the standard of care.

The analytical framework of the semi-Markov model and of the Discrete-Event Simulation presented in this thesis both highlighted the potential benefits of introducing to NHS routine practices a novel biomarker that can predict therapy response for early-stage breast cancer patients. The cost-effectiveness estimates indicate that adopting a technology as the EER4 biomarker has the strong potential for improved health outcomes for ER-positive breast cancer patient, and for cost savings for the NHS.

7.2 Strengths and Limitations

The literature review of current molecular markers predicting neoadjuvant therapy response in early breast cancer has examined the extent of the research field of these technologies still under development. While the review was not fully comprehensive in its scope, narrowing the search to molecular markers was conducive to the overall purpose of this thesis. The markers examined used similar technologies (such gene-sequencing, IHC, FISH) as the one employed by EER4 and potentially occupy a similar clinical context. It could be argued that with the appropriate assumptions and adaptations, any one of the markers presented in the review could be evaluated by the models in Chapter 5 and 6. This would have not

been the case for markers that are not molecular, for example imaging markers. The review is arguably timely and novel with respect to the growing research interest in this topic; to the author's knowledge, only one other literature review has examined the field: Tae et al. (2018) adopted a broader scope by including marker types other than molecular, and included commercially available adjuvant assays repurposed for the preoperative setting (133). Finally, even though this thesis mainly focuses on hormone-positive early breast cancer, the review also highlighted the research gap and clinical need with respect to other breast cancer subtypes. In particular, triple negative breast cancer, characterized by a worse prognosis in general compared to other subtypes, has still limited treatment options, and even fewer predictive markers.

The use of routine patient data in a cost-effectiveness analysis places this study among the increasing number of economic evaluations that opt to integrate real-world data in their estimates. The large number of patients, paired with the appropriate inclusion criteria, produced a study cohort that is representative of Scottish breast cancer patient population. In turn, this allowed to produce credible estimates of cost-effectiveness of a predictive test, which could potentially be more reflective of clinical practice, compared to the use of literature sources alone or using RCT data. Concerns remain regarding the risk of bias: either from the lack of randomization, or from the selection criteria.

Moreover, access to patient data presented several challenges: due to disclosure risks, the breadth and depth of the data requested had to be scaled down from the original proposal of the study. Patient privacy and disclosure risk consideration are of the utmost importance in any data request application. This, paired with processing times, has limited the amount of information that could be used for this study. In particular, three key pieces of information were not present on the final dataset: information on recurrence (time to recurrence event, type of recurrence), information on chemotherapy assignment, and information on neoadjuvant treatment. Furthermore, other data sources of interest would have been beneficial to the analysis if linked to the cancer patient cohort. In particular, detailed information on chemotherapy prescriptions and administration, and the Prescribing Information System, which stores data on all community prescriptions in Scotland, including anti-cancer hormone therapy. The consequences of the lack of this information is

discussed in the paragraph concerning the strength and limitations of the semi-Markov model.

The budget and decision impact analysis of OncotypeDX served as preliminary retrospective demonstration of the use of RWD. It revealed that the actual use of the 21-gene assay differed from the guidelines, and that despite the analytic limitations due to low testing rates, the chemotherapy savings were not enough to achieve parity with the cost of testing with OncotypeDX. As mentioned in Chapter 4, the analysis would have benefitted from higher testing rates, and an investigation of health outcomes would have been possible with enough follow-up and data on recurrence rates. A potential revision to the analysis is discussed in section 7.3, Future Research.

Throughout the study, the selection for intermediate values of OncotypeDX recurrence score do not reflect current clinical research practice. In section 3.4, which presents details surrounding risk stratification and the use of GR-PR as proxy for the OncotypeDX recurrence score, a value of 18 on the recurrence score was selected for separating low and high risk, and as a consequence the chemotherapy recommendation. At the time of the analysis, OncotypeDX recurrence score intervals of less than 18 for low-risk, between 18-30 for intermediate risk, and greater than 30 for high risk was still commonly used (also called “commercial cut-offs”), but intervals of 11 to 25 were already being established as the new standard, as reflected by the specifications of the TAILORx trial and the RxPONDER trials (134, 135). The decision to use the cut-off value of 18 was made for consistency with the assumptions of the analysis performed in Chapter 4, and due to the performance of the GR-PR scoring system. Nonetheless, the results of these trials and different recurrence score cut-off values might affect the results presented in this study.

The TAILORx study was a large, randomized, phase III clinical trial that aimed to determine the benefit of adjuvant chemotherapy for breast cancer patient fitting the eligibility criteria of OncotypeDX who had intermediate recurrence score values (with range 11-25). The study found that women with an intermediate recurrence score can be safely spared from chemotherapy and treated with endocrine therapy alone, while those with low or high recurrence scores should receive tailored treatment based on their individual risk profile.

The RxPONDER study was a large, randomized, phase III clinical trial that aimed to determine the benefit of adjuvant chemotherapy for women with hormone receptor-positive, HER2-negative, node-positive breast cancer who had low OncotypeDX recurrence score values (≤25). The study found that women in this group who received endocrine therapy alone had similar rates of invasive disease-free survival at 5 years compared to those who received both endocrine therapy and chemotherapy. However, in women with a recurrence score of 26-100, the addition of chemotherapy to endocrine therapy was associated with improved invasive disease-free survival at 5 years. The RxPONDER study's findings have important implications for the treatment of node-positive breast cancer patients with low OncotypeDX recurrence scores. These patients can now be spared the toxicity and costs of chemotherapy and treated with endocrine therapy alone. However, in those with high recurrence scores, the addition of chemotherapy may still be necessary to reduce the risk of recurrence.

The results and implications from these trials can potentially affect the output of the semi-Markov model: different OncotypeDX intermediate values affect the decision to treat with endocrine therapy alone or endocrine plus chemotherapy, thus affecting chemotherapy displacement, which in turn influences both the long-term recurrence rates and treatment costs. Given the evidence from the TAILORx and RxPONDER trials, and the performance of GR-PR for the 11-25 recurrence score interval, potential future analysis should incorporate new threshold values. Yet, given the overall small role that OncotypeDX plays in the economic model presented in this thesis, the change in threshold values would foreseeably have a negligible impact on the INMB produced by the strategy PREDICT + OncotypeDX, and thus would not change the ranking of the other strategies.

The model presented in Chapter 5 displays a robust framework, informed by a consolidated model structure, that takes account of the different outcomes between local and distant recurrence, and it reflects the potential adverse effects of chemotherapy, in terms health effects (as congestive heart failure and acute myeloid leukaemia), and cost effects (as chemotherapy toxicity). The use of PREDICT as a driver for recurrence rates has the advantage of providing the possibility to replicate the analysis in other population contexts, needing only the key variables for the calculation of PREDICT. Furthermore, the inclusion of the Secondary Cohort in the

analysis shows how the cost-effectiveness estimates are robust with respect to a higher-risk population, where the chemotherapy treatment rates are supposedly higher, and thus the potential benefits of identifying responders early become comparatively smaller (reducing the estimated net cost-savings of the predictive markers).

At the time of writing, this thesis is the first study to perform a practical application of the conditional net benefit frontier as described in section 5.3.5 with the POSA methods in McCabe et al. 2020 (119). This variation of POSA utilizes a two-stage Monte Carlo simulation that, while computationally demanding and often requiring coding the model around it, allows for examining the likelihood of extreme and decision-switching values and provides an intuitive visualization of the results. In this specific example, the conditional net benefit frontier strongly suggested that even at unlikely values of the probabilistic distribution of EER4 costs and effects (reflecting the uncertainty surrounding a developing technology), the strategy would still deliver the highest net monetary benefits.

The literature sources used to inform the measures of Health-Related Quality of Life might present a point of weakness for the analytic model: the baseline utility values date back to a 1998 EQ-5D survey in the UK and these values might not reflect present baseline utility values anymore (102). The estimated utility decrements for breast cancer-related health states might suffer from the same problem: Campbell et al 2011 (103) derived the HRQoL from three trials, other economic evaluations, and patient preferences studies (106, 136-140); all dated pre-2006. In this timespan, clinical pathways and treatment practices have greatly changed for early breast cancer patients, and these utility values might not be an appropriate reflection of patient experience in the present anymore. There exists a large degree of heterogeneity (both in methods and results) among utility estimates for early breast cancer, making it difficult to select appropriate and up-to-date values in an economic evaluation (141). The rationale for choosing the values used in Campbell et al 2011 in this study lies mainly in the population context and the focus on the consequences of chemotherapy and its displacement.

As mentioned previously, the lack of access to recurrence data presented a challenge for the analysis: not being able to derive internal time-to-event estimates,

along with the type of recurrence, weakens the internal validity of the transition probabilities used in the semi-Markov model. The use of PREDICT mitigates this problem, given its established validity, but internally consistent event rates would have been preferable. On the subject of access to data, the analysis had no information on which patients were treated in the neoadjuvant setting, whether with endocrine or chemotherapy; this resulted in adopting several underlying assumptions that increase the uncertainty of the estimates: in particular, the model takes the PREDICT estimates at face value, regardless of the true neoadjuvant treatment status (which cannot be derived with the data available). PREDICT is currently not validated for patients treated with neoadjuvant therapy: holding other things equal, the algorithm assign the same risk profile for tumours of the same size at surgery, whether it was the original size or it was shrunk through neoadjuvant chemotherapy. Having access to individual-level data on neoadjuvant treatment would have improved costing estimates as well. The model currently uses the estimates from the OPTIMA trial to assign the type of chemotherapy that patients receive: access to individual-level chemotherapy data could have improved the estimation of costs and toxicity event rates, producing a more accurate estimate of cost-effectiveness. Access to all three sources of missing information would have meant that the model could have directly compared the difference in recurrence (and other events) rates between patients treated with neoadjuvant therapy versus patients treated in the adjuvant setting only, with or without adjuvant chemotherapy. Furthermore, given the lack of patient-level neoadjuvant treatment information, the semi-Markov model could not simulate accurately the preoperative setting and instead relies on either aggregate estimates or subsequent input from the DES model in Chapter 6.

The semi-Markov model was developed in base R language, with the explicit intention of adding as few dependencies as possible and avoiding using a Markov modelling-specific software package. This decision was taken for two reasons: the code would be transparent and reproducible regardless of updated or deprecated dependencies, and the model would not need to rely on “black box” processes, where black box refers to the inability or difficulty for the user to understand how the inputs are transformed in the final outputs. Markov modelling-specific software packages certainly have their merits, but coding the model from the start in base language had great training and learning value.

A similar choice could not be made for the Discrete Event Simulation: running this type of simulations in R, while certainly possible, comes with its own specific issues. DES can be either event-oriented or trajectory-oriented: event-oriented simulations are optimized for simulating several different types of events and maintaining a time-stamped event-list; trajectory-oriented simulations focus on a smaller pool of possible events and are optimized for measuring time-to-events accompanied by a known number of final branches ("arrivals"). Base R language is efficient at matrix multiplication, which can be leveraged for event-oriented simulations, but unfortunately the modelling of the neoadjuvant setting for early breast cancer is better represented by trajectory-oriented simulations (which base R language is not efficient at or practical for). For this reason, the "simmer" package was selected, which enables efficient trajectory-oriented simulations in R by integrating C++ code in the background. This meant a trade-off between accepting a black box process for the convenience of running code in a familiar and known programming language.

Beyond software considerations, the simulation itself is presented as a novel proof-of-concept aimed at modelling neoadjuvant treatment in early breast cancer: a conceptually intuitive platform that can be further adapted for the inclusion of other events and adverse effects associated with therapy. As covered in the discussion section of Chapter 6, several simplifying assumptions were made, and as such the model is not comprehensive of all the consequences of treating in the neoadjuvant setting. The presented modelling platform could have been improved by the seamless integration of a two-step modelling approach (between the DES and the semi-Markov model) that is fully internally consistent. However, the lack of access to patient-level data regarding treatment prevented the derivation of input parameters from RWD, which would have ensured consistency between the DES model and the concepts presented in the thesis. Furthermore, this would have enabled a smooth integration between the DES model and the semi-Markov model.

To achieve a seamless two-step modelling platform, an early breast cancer patient cohort treated neoadjuvantly with aromatase inhibitors, along with the associated follow-up data, outcome data, and surgery information, would have been necessary to populate the models fully. With timely delivery of such data, several of the simplifying assumptions made in the DES model could have been avoided, such as the assumptions surrounding re-excision and breast reconstruction rates. Overall,

the output of the DES model would have been more representative of the current experience and outcomes of breast cancer patients being treated in Scotland.

On the semi-Markov model side, the availability of outcome data (recurrence, death, and other relevant endpoints) would have eliminated the need for reliance on PREDICT, prevented potential spill-over effects between states, and ensured consistency with the DES model. The models could have been run in a more intuitive chronological order, with outcomes and events flowing from one model to the other seamlessly.

Future iterations of the model should aim to use the appropriate type and quantity of data to achieve a seamless and fully internally consistent two-step modelling approach.

Overall, the results of the two models are reflective of the assumptions under which the early cost-effectiveness analysis was carried out, and they characterize the uncertainty surrounding a technology under development and its likely impact on NHS routine practices. Whenever underlying assumptions were made, appropriate justification was provided. The choices related to these assumptions, along with the issues associated with structural uncertainty, remain a challenge in general for any research design that aims to accurately model any clinical pathway. Yet, these choices are of extreme importance for early CEA in particular: model structures for novel technologies have a near-infinite number of permutations, and thus the results of a simulation are always dependent on the analyst choices and open to bias. To mitigate this, extensive sensitivity analysis has been implemented to ensure a reasonable degree of confidence in the results, along with a pragmatic approach to interpretation.

In conclusion, this thesis not only presents the results of an early cost-effectiveness analysis, but it is a reflection of a path of learning and training from the start of the study until thesis submission, represented in modelling decision, the use of data, and programming choices.

7.3 Future Research

Work is underway to adapt EER4, which measures the level of expression of 4 genes before and on-treatment, into a more clinically accessible biomarker: EA2Clin (142). This novel test is immunohistochemistry-based and integrates the measurement of two protein, before and on-treatment, with clinically observable variables. As EA2Clin is IHC-based, it would bypass the need of expensive gene-level measurement equipment and allow for it to be performed in local pathology laboratories. In light of these developments, and once EA2Clin is clinically validated, the cost-effectiveness estimates from this thesis should be revaluated, and the model adapted for assessing EA2Clin. The comparison between the early cost-effectiveness estimates of EER4 and those of EA2Clin would potentially serve as an example of the impact of a technology transitioning to a more accessible method, with the potential for maturing into a clinical decision-making tool.

Patients that were initially tested in NHS Lothian with OncotypeDX since its introduction have now accumulated at least five years of follow-up. In conjunction with an updated version of PREDICT, this presents the potential for updating and refining the budget and decision impact analysis presented in Chapter 4. With the inclusion of health outcomes data, this potential future analysis could deliver better estimates of the chemotherapy displacement effect of the 21-gene signature. A comparison of recurrence rates of the tested patients against the rates of historical patients with similar clinical attributes would provide a more accurate and informative estimate of decision impact.

The modelling platform in Chapter 5 was designed to be flexible and adaptable. While in its current state the model measures the impact of before- and on-treatment neoadjuvant biomarkers for ER-positive HER2-negative breast cancer, it can be modified for markers that use different timings and provide different benefits. Moreover, by using PREDICT as the main driver of the transition probabilities, including potential treatment benefits, the model can be further adapted for the other breast cancer subtypes, namely HER2-positive and Triple Negative breast cancer. Alternatively, acquiring patient-level data on recurrences and neoadjuvant treatment would allow for correcting the model's shortcoming described in section 7.2.

Further developments of the modelling platform would include modifications to the Discrete Event Simulation in Chapter 6. These modifications would include integrating additional events in the trajectory, such as margin re-excisions, unsuccessful surgeries, and adverse events. The model could be further improved by adding additional strategies in the model structure: predicted non-responders are immediately assigned to surgery, but other options could reasonably be available. This could include a change in endocrine therapy (e.g. from Letrozole to tamoxifen), neoadjuvant chemotherapy, or even the introduction in a basket trial. Other changes would reasonably encompass the choice of software, with the intention of moving away from depending on third-party software packages, and coding the model directly in a base language without the black box processes described in section 7.2.

The current standard method for reporting costings across the NHS is reference costs, which take financial data from the general ledger and generate an average cost per patient. This average costs per patient were used in this study to inform the cost associated with chemotherapy assignment, delivery, and toxicity. Patient Level Information and Costing System (PLICS) has been used in this study in the single instance of calculating per-episode costs of Febrile Neutropoenia. PLICS uses the same data sources as the reference costs, but it provides costs by individual patient using patient identifiers (Community Health Index in Scotland, NHS Number in England and Wales), and calculating the cost for each part of the care provided for each individual. Future analyses should make use of PLICS where available, granting a higher level of accuracy on the costing side of an economic evaluation. Wider use of this approach would contribute to the growing library of methods available for using Real-World Data and Evidence within economic evaluations.

References

1. Breast cancer risk | Cancer Research UK. 2015 – 2017.
2. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A. 2001 Sep 11 ;98(19):10869–74.
3. Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Veizer J, McMichael JF, et al. Comprehensive molecular portraits of human breast tumours. Nature 2012 490:7418. 2012 Sep 23;490(7418):61–70.
4. Sotiriou C, Neo SY, Mcshane LM, Korn EL, Long PM, Jazaeri A, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study.
5. Harris LN, Ismaila N, McShane LM, Andre F, Collyar DE, Gonzalez-Angulo AM, et al. Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline. J Clin Oncol. 2016 Apr 1;34(10):1134–50.
6. Thompson AM, Moulder-Thompson SL. Neoadjuvant treatment of breast cancer. Annals of Oncology. 2012;23(Suppl 10):x231.
7. Recommendations | Early and locally advanced breast cancer: diagnosis and management | Guidance | NICE.
8. Haybittle JL, Blamey RW, Elston CW, Johnson J, Doyle PJ, Campbell FC, et al. A prognostic index in primary breast cancer. Br J Cancer. 1982;45(3):361–6.
9. Galea MH, Blamey RW, Elston CE, Ellis IO. The Nottingham Prognostic Index in primary breast cancer. Breast Cancer Res Treat. 1992 Oct;22(3):207–19.
10. Balslev I, Axelsson CK, Zedeler K, Rasmussen BB, Carstensen B, Mouridsen HT. The Nottingham Prognostic Index applied to 9,149 patients from the studies of the Danish Breast Cancer Cooperative Group (DBCG). Breast Cancer Res Treat. 1994 Jan;32(3):281–90.
11. Fong Y, Evans J, Brook D, Kenkre J, Jarvis P, Gower-Thomas K. The Nottingham Prognostic Index: five- and ten-year data for all-cause Survival within a Screened Population. Ann R Coll Surg Engl. 2015;97(2):137.
12. Ravdin PM, Siminoff LA, Davis GJ, Mercer MB, Hewlett J, Gerson N, et al. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. Journal of Clinical Oncology. 2001 Feb 15;19(4):980–91.
13. Wishart GC, Azzato EM, Greenberg DC, Rashbass J, Kearins O, Lawrence G, et al. PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. Breast Cancer Res. 2010 Jan 6;12(1).
14. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2020.
15. Tramonti G. nhs.predict: Breast Cancer Survival and Therapy Benefits. Edinburgh, United Kingdom; 2020.

16. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med. 2004 Dec 30 ;351(27):2817–26.

17. Bernard PS, Parker JS, Mullins M, Cheung MCU, Leung S, Voduc D, et al. Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. Journal of Clinical Oncology. 2009 Mar 3 ;27(8):1160.

18. van de Vijver MJ, He YD, van 't Veer LJ, Dai H, Hart AAM, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med. 2002 Dec 19 ;347(25):1999–2009.

19. Filipits M, Rudas M, Jakesz R, Dubsky P, Fitzal F, Singer CF, et al. A new molecular predictor of distant recurrence in ER-positive, HER2-negative breast cancer adds independent information to conventional clinical risk factors. Clin Cancer Res. 2011 Sep 15;17(18):6012–20.

20. Cuzick J, Dowsett M, Pineda S, Wale C, Salter J, Quinn E, et al. Prognostic value of a combined estrogen receptor, progesterone receptor, Ki-67, and human epidermal growth factor receptor 2 immunohistochemical score and comparison with the Genomic Health recurrence score in early breast cancer. J Clin Oncol. 2011 Nov 10;29(32):4273–8.

21. Ikeda T, Jinno H, Matsui A, Masamura S, Kitajima M. The Role of Neoadjuvant Chemotherapy for Breast Cancer Treatment. Vol. 9, Breast Cancer. 2002.

22. Chia YH, Ellis MJ, Ma CX. Neoadjuvant endocrine therapy in primary breast cancer: indications and use as a research tool. Br J Cancer. 2010;103:759–64.

23. Eiermann W, Paepke S, Appfelstaedt J, Llombart-Cussac A, Eremin J, Vinholes J, et al. Preoperative treatment of postmenopausal breast cancer patients with letrozole: A randomized double-blind multicenter study. Ann Oncol. 2001;12(11):1527–32.

24. Dowsett M, Smith IE, Ebbs SR, Dixon JM, Skene A, Griffith C, et al. Short-Term Changes in Ki-67 during Neoadjuvant Treatment of Primary Breast Cancer with Anastrozole or Tamoxifen Alone or Combined Correlate with Recurrence-Free Survival. Clinical Cancer Research. 2005 Jan 15;11(2):951s–8s.

25. Yeo B, Dowsett M. Neoadjuvant endocrine therapy: Patient selection, treatment duration and surrogate endpoints. Breast. 2015 Nov 1;24 Suppl 2:S78–83.

26. Ling DI, Lynd LD, Harrison M, Anis AH, Bansback N. Early cost–effectiveness modeling for better decisions in public research investment of personalized medicine technologies. https://doi.org/102217/cer-2018-0033. 2018 Dec 10;8(1):7–19.

27. Buisman LR, Rutten-van Mölken M,P.M.H., Postmus D, Luime JJ, Uyl-de Groot C,A., Redekop WK. THE EARLY BIRD CATCHES THE WORM: EARLY COST-EFFECTIVENESS ANALYSIS OF NEW MEDICAL TESTS. Int J Technol Assess Health Care 2016;32(1-2):46-53.

28. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. Health Econ. 2006 Dec;15(12):1295–310.

29. Santa-Maria CA, Camp M, Cimino-Mathews A, Harvey S, Wright J, Stearns V. Neoadjuvant Therapy for Early-Stage Breast Cancer: Current Practice, Controversies, and Future Directions. Oncology. 2015;29(11).

30. Demichele A, Yee D, Berry DA, Albain KS, Benz CC, Boughey J, et al. The Neoadjuvant Model Is Still the Future for Drug Development in Breast Cancer. 2015.

31. Gerdes SJ, Lemke H, Baisch H, Wacker HH, Schwab U. Ki-67. antigen defined by the monoclonal antibody proliferation-associated human nuclear Cell cycle analysis of a cell. 1984.

32. Yerushalmi R, Woods R, Ravdin PM, Hayes MM, Gelmon KA. Ki67 in breast cancer: prognostic and predictive potential. Vol. 11, The Lancet Oncology. 2010. p. 174–83.

33. Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, et al. Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. J Clin Oncol. 2009;27:1160–7.

34. Cheang MCU, Chia SK, Voduc D, Gao D, Leung S, Snider J, et al. Ki67 Index, HER2 Status, and Prognosis of Patients With Luminal B Breast Cancer. Articles | JNCI. 2009;101:736.

35. Fasching PA, Heusinger K, Haeberle L, Niklos M, Hein A, Bayer CM, et al. Ki67, chemotherapy response, and prognosis in breast cancer patients receiving neoadjuvant treatment. 2011.

36. Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, Taube S, et al. American Society of Clinical Oncology 2007 Update of Recommendations for the Use of Tumor Markers in Breast Cancer. Journal of Clinical Oncology. 2007 Nov 20;25(33).

37. Dowsett M, Nielsen TO, A'Hern R, Bartlett J, Coombes RC, Cuzick J, et al. Assessment of Ki67 in Breast Cancer: Recommendations from the International Ki67 in Breast Cancer Working Group. JNCI Journal of the National Cancer Institute. 2011 Nov 16;103(22).

38. Urruticoechea A, Smith IE, Dowsett M. Proliferation Marker Ki-67 in Early Breast Cancer. Journal of Clinical Oncology. 2005 Oct 1;23(28).

39. Brown JR, Digiovanna MP, Killelea B, Lannin DR, Rimm DL. Quantitative assessment Ki-67 score for prediction of response to neoadjuvant chemotherapy in breast cancer. Laboratory Investigation. 2014 Jan;94(1):98–106.

40. Yoshioka T, Hosoda M, Yamamoto M, Taguchi K, Hatanaka KC, Takakuwa E, et al. Prognostic significance of pathologic complete response and Ki67 expression after neoadjuvant chemotherapy in breast cancer. Breast Cancer. 2015 Mar 1;22(2):185–91.

41. Tokuda E, Horimoto Y, Arakawa A, Himuro T, Senuma K, Nakai K, et al. Differences in Ki67 expressions between pre- and post-neoadjuvant chemotherapy specimens might predict early recurrence of breast cancer. Hum Pathol. 2017 May 1;63:40–5.

42. Tan QX, Qin QH, Yang WP, Mo QG, Wei CY. Prognostic value of Ki67 expression in HR-negative breast cancer before and after neoadjuvant chemotherapy. Vol. 7, Int J Clin Exp Pathol. 2014.

43. Ellis MJ, Suman VJ, Hoog J, Goncalves R, Sanati S, Creighton CJ, et al. Ki67 proliferation index as a tool for chemotherapy decisions during and after neoadjuvant aromatase inhibitor treatment of breast cancer: Results from the American college of surgeons oncology group Z1031 trial (alliance). Journal of Clinical Oncology. 2017 Apr 1;35(10):1061–9.

44. Ellis MJ, Tao Y, Luo J, A'Hern R, Evans DB, Bhatnagar AS, et al. Outcome prediction for estrogen receptor-positive breast cancer based on postneoadjuvant endocrine therapy tumor characteristics. J Natl Cancer Inst. 2008 Oct;100(19):1380–8.

45. Goncalves R, DeSchryver K, Ma C, Tao Y, Hoog J, Cheang M, et al. Development of a Ki-67-based clinical trial assay for neoadjuvant endocrine therapy response monitoring in breast cancer. Breast Cancer Res Treat. 2017 Sep 1;165(2):355–64.

46. Leung SCY, Nielsen TO, Zabaglo LA, Arun I, Badve SS, Bane AL, et al. Analytical validation of a standardised scoring protocol for Ki67 immunohistochemistry on breast cancer excision whole sections: an international multicentre collaboration. Histopathology. 2019 Aug 8;75(2).

47. Polley MYC, Leung SCY, McShane LM, Gao D, Hugh JC, Mastropasqua MG, et al. An International Ki67 Reproducibility Study. JNCI: Journal of the National Cancer Institute. 2013 Dec 18;105(24).

48. Haibe-Kains B, Desmedt C, Loi S, Culhane AC, Bontempi G, Quackenbush J, et al. A Three-Gene Model to Robustly Identify Breast Cancer Molecular Subtypes. JNCI: Journal of the National Cancer Institute. 2012 Feb 22;104(4).

49. Smith I, Robertson J, Kilburn L, Wilcox M, Evans A, Holcombe C, Horgan K, Kirwan C, Mallon E, Sibbering M, Skene A, Vidya R, Cheang M, Banerji J, Morden J, Sidhu K, Dodson A, Bliss JM, Dowsett M. Long-term outcome and prognostic value of Ki67 after perioperative endocrine therapy in postmenopausal women with hormone-sensitive early breast cancer (POETIC): an open-label, multicentre, parallel-group, randomised, phase 3 trial. Lancet Oncol. 2020 Nov;21(11):1443-1454.

50. Gentles AJ, Newman AM, Liu CL, Bratman S v, Feng W, Kim D, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat Med. 2015 Aug 20;21(8).

51. Andre F, Dieci M v, Dubsky P, Sotiriou C, Curigliano G, Denkert C, et al. Molecular Pathways: Involvement of Immune Pathways in the Therapeutic Response and Outcome in Breast Cancer. 2013;

52. Syn NL, Teng MWL, Mok TSK, Soo RA. De-novo and acquired resistance to immune checkpoint targeting. Lancet Oncol. 2017 Dec;18(12).

53. Miyashita M, Sasano H, Tamaki K, Hirakawa H, Takahashi Y, Nakagawa S, et al. Prognostic significance of tumor-infiltrating CD8+ and FOXP3+ lymphocytes in residual tumors and alterations in these parameters after neoadjuvant chemotherapy in triple-negative breast cancer: A retrospective multicenter study. Breast Cancer Research. 2015 Sep 4;17(1).

54. Herman Fridman W, erôme Galon J, Pag es F, Tartour E, Saut es-Fridman C, Kroemer G. Prognostic and Predictive Impact of Intra-and Peritumoral Immune Infiltrates. 2011;

55. Jeschke J, Bizet M, Desmedt C, Calonne E, Dedeurwaerder S, Garaud S, et al. DNA methylation-based immune response signature improves patient diagnosis in multiple cancers. Journal of Clinical Investigation. 2017 Aug 1;127(8):3090–102.

56. Sota Y, Naoi Y, Tsunashima R, Kagara N, Shimazu K, Maruyama N, et al. Construction of novel immune-related signature for prediction of pathological complete response to neoadjuvant chemotherapy in human breast cancer. Annals of Oncology. 2014 Jan 1;25(1):100–6.

57. Bertucci F, Ueno NT, Finetti P, Vermeulen P, Lucci A, Robertson FM, et al. Gene expression profiles of inflammatory breast cancer: Correlation with response to neoadjuvant chemotherapy and metastasis-free survival. Annals of Oncology. 2014 Feb;25(2):358–65.

58. Harris LN, You F, Schnitt SJ, Witkiewicz A, Lu X, Sgroi D, et al. Predictors of Resistance to Preoperative Trastuzumab and Vinorelbine for HER2-Positive Early Breast Cancer. Clinical Cancer Research. 2007 Feb 15;13(4).

59. Salgado R, Denkert C, Campbell C, Savas P, Nuciforo P, Aura C, de Azambuja E, Eidtmann H, Ellis CE, Baselga J, Piccart-Gebhart MJ, Michiels S, Bradbury I, Sotiriou C, Loi S. Tumor-Infiltrating Lymphocytes and Associations With Pathological Complete Response and Event-Free Survival in

HER2-Positive Early-Stage Breast Cancer Treated With Lapatinib and Trastuzumab: A Secondary Analysis of the NeoALTTO Trial. JAMA Oncol. 2015 Jul;1(4):448-54.

60. Denkert C, von Minckwitz G, Darb-Esfahani S, Lederer B, Heppner BI, Weber KE, Budczies J, Huober J, Klauschen F. Tumour-infiltrating lymphocytes and prognosis in different subtypes of breast cancer: a pooled analysis of 3771 patients treated with neoadjuvant therapy. Lancet Oncol. 2018 Jan;19(1):40-50.

61. Varadan V, Gilmore H, Miskimen KLS, Tuck D, Parsai S, Awadallah A, et al. Immune signatures following single dose trastuzumab predict pathologic response to preoperative trastuzumab and chemotherapy in HER2-positive early breast cancer. Clinical Cancer Research. 2016 Jul 1;22(13):3249–59.

62. Perez EA, Romond EH, Suman VJ, Jeong JH, Sledge G, Geyer CE, et al. Trastuzumab Plus Adjuvant Chemotherapy for Human Epidermal Growth Factor Receptor 2–Positive Breast Cancer: Planned Joint Analysis of Overall Survival From NSABP B-31 and NCCTG N9831. Journal of Clinical Oncology. 2014 Nov 20;32(33).

63. The National Institute for Health and Care Excellence. Pertuzumab for adjuvant treatment of HER2-positive early stage breast cancer [Internet]. 2019. Available from: www.nice.org.uk/guidance/ta569

64. Harris LN, Varadan V, Sandoval · M, Varadan V, Sandoval M, Harris LN. Novel Biomarkers in the Continuum of Breast Cancer. Adv Exp Med Biol. 2016;882.

65. Prat A, Guarneri V, Pascual T, Brasó-Maristany F, Sanfeliu E, Paré L, Schettini F, Martínez D, Jares P, Griguolo G, Dieci MV. Development and validation of the new HER2DX assay for predicting pathological response and survival outcome in early-stage HER2-positive breast cancer. EBioMedicine. 2022 Jan;75:103801.

66. Bertucci F, Finetti P, Viens P, Birnbaum D. EndoPredict predicts for the response to neoadjuvant chemotherapy in ER-positive, HER2-negative breast cancer. Cancer Lett. 2014 Dec 1;355(1):70–5.

67. Prat A, Lluch A, Turnbull AK, Dunbier AK, Calvo L, Albanell J, et al. A PAM50-based chemoendocrine score for hormone receptor-positive breast cancer with an intermediate risk of relapse. Clinical Cancer Research. 2017 Jun 15;23(12):3035–44.

68. Turnbull AK, Arthur LM, Renshaw L, Larionov AA, Kay C, Dunbier AK, et al. Accurate prediction and validation of response to endocrine therapy in breast cancer. Journal of Clinical Oncology. 2015 Jul 10;33(20):2270–8.

69. Magbanua MJM, Wolf DM, Yau C, Davis SE, Crothers J, Au A, et al. Serial expression analysis of breast tumors during neoadjuvant chemotherapy reveals changes in cell cycle and immune pathways associated with recurrence and response. Breast Cancer Research. 2015 May 29;17(1).

70. Bownes RJ, Turnbull AK, Martinez-Perez C, Cameron DA, Sims AH, Oikonomidou O. On-treatment biomarkers can improve prediction of response to neoadjuvant chemotherapy in breast cancer. Available from: https://doi.org/10.1186/s13058-019-1159-3

71. Selli C, Turnbull AK, Pearce DA, Li A, Fernando A, Wills J, et al. Molecular changes during extended neoadjuvant letrozole treatment of breast cancer: Distinguishing acquired resistance from dormant tumours 11 Medical and Health Sciences 1112 Oncology and Carcinogenesis 11 Medical and Health Sciences 1103 Clinical Sciences. Breast Cancer Research. 2019 Jan 7;21(1).

72. Lips EH, Michaut M, Hoogstraat M, Mulder L, Besselink NJM, Koudijs MJ, et al. Next generation sequencing of triple negative breast cancer to find predictors for chemotherapy response. Breast Cancer Research. 2015 Oct 3;17(1).

73. Bertucci F, Ng CKY, Patsouris A, Droin N, Piscuoglio S, Carbuccia N, et al. Genomic characterization of metastatic breast cancers. Nature. 2019 May 23;569(7757):560–4.

74. Jiang YZ, Ma D, Suo C, Shi J, Xue M, Hu X, et al. Genomic and Transcriptomic Landscape of Triple-Negative Breast Cancers: Subtypes and Treatment Strategies. Cancer Cell. 2019 Mar 18;35(3):428-440.e5.

75. Whitworth P, Stork-Sloots L, de Snoo FA, Richards P, Rotkis M, Beatty J, et al. Chemosensitivity Predicted by BluePrint 80-Gene Functional Subtype and MammaPrint in the Prospective Neoadjuvant Breast Registry Symphony Trial (NBRST). Ann Surg Oncol. 2014;21(10):3261–7.

76. Sørlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A. 2003 Jul 8;100(14):8418-23.

77. Abkevich V, Timms K, Hennessy B, Potter J, Carey M, Meyer L, et al. Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. Br J Cancer. 2012;107:1776–82.

78. Birkbak NJ, Wang ZC, Kim JY, Eklund AC, Li Q, Tian R, et al. Cancer Research. Cancer Research on. 2021;

79. Popova T, Mani E, Rieunier G, Caux-Moncoutier V, Tirapo C, Dubois T, et al. Ploidy and Large-Scale Genomic Instability Consistently Identify Basal-like Breast Carcinomas with BRCA1/2 Inactivation. 2012;

80. Telli ML, Timms KM, Reid J, Hennessy B, Mills GB, Jensen KC, et al. Cancer Therapy: Clinical Homologous Recombination Deficiency (HRD) Score Predicts Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer. 2016;

81. Buchanan J, Wordsworth S, Schuh A. Issues surrounding the health economic evaluation of genomic technologies Europe PMC Funders Group. Pharmacogenomics. 2013;14(15):1833–47.

82. Miquel-Cases A, Schouten PC, Steuten LMG, Retèl VP, Linn SC, van Harten WH. (Very) Early technology assessment and translation of predictive biomarkers in breast cancer. Cancer Treat Rev. 2017 Jan;52.

83. Blok EJ, Bastiaannet E, van den Hout WB, Liefers GJ, Smit VTHBM, Kroep JR, et al. Systematic review of the clinical and economic value of gene expression profiles for invasive early breast cancer available in Europe. Cancer Treat Rev. 2018 Jan;62.

84. Hall PS, Smith A, Hulme C, Vargas-Palacios A, Makris A, Hughes-Davies L, et al. Value of Information Analysis of Multiparameter Tests for Chemotherapy in Early Breast Cancer: The OPTIMA Prelim Trial. Value in Health. 2017 Dec 1;20(10):1311–8.

85. Almario CV, Noah BD, Jusufagic A, Lew D, Spiegel BMR. Cost Effectiveness of Biomarker Tests for Irritable Bowel Syndrome With Diarrhea: A Framework for Payers. Clin Gastroenterol Hepatol. 2018 Sep;16(9):1434–1441.

86. Campos NG, Tsu V, Jeronimo J, Mvundura M, Kim JJ. Estimating the value of point-of-care HPV testing in three low- and middle-income countries: a modeling study. BMC Cancer. 2017 Nov;17(1):791.

87. Critselis E, Vlahou A, Stel VS, Morton RL. Cost-effectiveness of screening type 2 diabetes patients for chronic kidney disease progression with the CKD273 urinary peptide classifier as compared to urinary albumin excretion. Nephrol Dial Transplant. 2018 Mar;33(3):441–449.

88. Degeling K, Schivo S, Mehra N, Koffijberg H, Langerak R, de Bono JS, IJzerman MJ. Comparison of Timed Automata with Discrete Event Simulation for Modeling of Biomarker-Based Treatment Decisions: An Illustration for Metastatic Castration-Resistant Prostate Cancer. Value Health. 2017 Dec;20(10):1411–1419.

89. Doble B, John T, Thomas D, Fellowes A, Fox S, Lorgelly P. Cost-effectiveness of precision medicine in the fourth-line treatment of metastatic lung adenocarcinoma: An early decision analytic model of multiplex targeted sequencing. Lung Cancer. 2017 May;107:22–35.

90. Khoudigian-Sinani S, Blackhouse G, Levine M, Thabane L, O'Reilly D. The premarket assessment of the cost-effectiveness of a predictive technology "Straticyte™" for the early detection of oral cancer: a decision analytic model. Health Econ Rev. 2017 Oct;7(1):35.

91. Lansdorp-Vogelaar I, Goede SL, Bosch LJW, Melotte V, Carvalho B, van Engeland M, Meijer GA, de Koning HJ, van Ballegooijen M. Cost-effectiveness of High-performance Biomarker Tests vs Fecal Immunochemical Test for Noninvasive Colorectal Cancer Screening. Clin Gastroenterol Hepatol. 2018 Apr;16(4):504–512.

92. Lotan Y, Woldu SL, Sanli O, Black P, Milowsky MI. Modelling cost-effectiveness of a biomarker-based approach to neoadjuvant chemotherapy for muscle-invasive bladder cancer. BJU Int. 2018 Sep;122(3):434–440.

93. Mitchell D, Guertin JR, Dubois A, Dubé MP, Tardif JC, Iliza AC, Fanton-Aita F, Matteau A, LeLorier J. A Discrete Event Simulation Model to Assess the Economic Value of a Hypothetical Pharmacogenomics Test for Statin-Induced Myopathy in Patients Initiating a Statin in Secondary Cardiovascular Prevention. Mol Diagn Ther. 2018 Apr;22(2):241–254.

94. Terjesen CL, Kovaleva J, Ehlers L. Early Assessment of the Likely Cost Effectiveness of Single-Use Flexible Video Bronchoscopes. Pharmacoecon Open. 2017 Jun;1(2):133–141.

95. Weaver DT, Raphel TJ, Melamed A, Rauh-Hain JA, Schorge JO, Knudsen AB, Pandharipande PV. Modeling treatment outcomes for patients with advanced ovarian cancer: Projected benefits of a test to optimize treatment selection. Gynecol Oncol. 2018 May;149(2):256–262.

96. Willemsen RTA, Kip MMA, Koffijberg H, Kusters R, Buntinx F, Glatz JFC, Dinant GJ; 'RAPIDA' – Study Team ('RAPIDA': RAPid Test for Investigating Complaints Possibly Due to Acute Coronary Syndrome). Early health technology assessment of future clinical decision rule aided triage of patients presenting with acute chest pain in primary care. Prim Health Care Res Dev. 2018 Mar;19(2):176-188

97. Luime JJ, Buisman LR, Oppe M, Hazes JM, Rutten-van Mölken MP. Cost-Effectiveness Model for Evaluating New Diagnostic Tests in the Evaluation of Patients With Inflammatory Arthritis at Risk of Having Rheumatoid Arthritis. Arthritis Care Res (Hoboken). 2016 Jul;68(7):927-35.

98. Buisman LR, Luime JJ, Oppe M, Hazes JM, Rutten-van Mölken MP. A five-year model to assess the early cost-effectiveness of new diagnostic tests in the early diagnosis of rheumatoid arthritis. Arthritis Res Ther. 2016 Jun 10;18(1):135.

99. Chen IT, Aung T, Thant HN, Sudhinaraset M, Kahn JG. Cost-effectiveness analysis of malaria rapid diagnostic test incentive schemes for informal private healthcare providers in Myanmar. Malar J. 2015 Feb 5;14:55.

100. IJzerman MJ, Koffijberg H, Fenwick E, Krahn M. Emerging Use of Early Health Technology Assessment in Medical Product Development: A Scoping Review of the Literature. Pharmacoeconomics. 2017 Jul;35(7):727-740.

101. Frempong SN, Sutton AJ, Davenport C, Barton P. Economic evaluation of medical tests at the early phases of development: a systematic review of empirical studies. Expert Rev Pharmacoecon Outcomes Res. 2018 Feb;18(1):13-23.

102. Abel L, Shinkins B, Smith A, Sutton AJ, Sagoo GS, Uchegbu I, Allen AJ, Graziadio S, Moloney E, Yang Y, Hall P. Early Economic Evaluation of Diagnostic Technologies: Experiences of the NIHR Diagnostic Evidence Co-operatives. Med Decis Making. 2019 Oct;39(7):857-866.

103. The National Institute for Health and Care Excellence. Tumour profiling tests to guide adjuvant chemotherapy decisions in early breast cancer Diagnostics guidance. 2018.

104. Prasad V, Mailankody S. Research and Development Spending to Bring a Single Cancer Drug to Market and Revenues After Approval. JAMA Intern Med. 2017 Nov 1;177(11).

105. Bowrin K, Briere JB, Levy P, Millier A, Clay E, Toumi M. Cost-effectiveness analyses using real-world data: an overview of the literature. J Med Econ. 2019 Jun 3;22(6).

106. Barnish M, Turner S. The value of pragmatic and observational studies in health care and public health. Pragmat Obs Res. 2017 May;Volume 8.

107. Dhruva S, Ross J, Desai N. Real-World Evidence: Promise and Peril For Medical Product Evaluation. 2018 Aug;

108. Bell H, Wailoo AJ, Hernandez M, Faria R. he use of real world data for the estimation of treatment effects in NICE decision making. NICE Decision Support Unit. 2016;

109. Hampson G, Towse A, Dreitlein WB, Henshall C, Pearson SD. Real-world evidence for coverage decisions: opportunities and challenges. J Comp Eff Res. 2018 Dec;7(12).

110. Bullement A, Podkonjak T, Robinson MJ, Benson E, Selby R, Hatswell AJ, et al. Real-world evidence use in assessments of cancer drugs by NICE. Int J Technol Assess Health Care. 2020 Aug 1;36(4):388–94.

111. Khosla S, White R, Medina J, Ouwens M, Emmas C, Koder T, et al. Real world evidence (RWE) – a disruptive innovation or the quiet evolution of medical evidence generation? F1000Res. 2018 Aug 29;7.

112. Thibodeau S, Voutsadakis IA. Prediction of Oncotype Dx recurrence score using clinical parameters: A comparison of available tools and a simple predictor based on grade and progesterone receptor. Hematol Oncol Stem Cell Ther. 2019 Jun;12(2).

113. Holt S, Bertelli G, Humphreys I, Valentine W, Durrani S, Pudney D, et al. A decision impact, decision conflict and economic assessment of routine Oncotype DX testing of 146 women with node-negative or pNImi, ER-positive breast cancer in the UK. 2013;

114. Loncaster J, Armstrong A, Howell S, Wilson G, Welch R, Chittalia A, et al. Impact of Oncotype DX breast Recurrence Score testing on adjuvant chemotherapy use in early breast cancer: Real world experience in Greater Manchester, UK. European Journal of Surgical Oncology. 2017 May 1;43(5):931–7.

115. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. BMJ. 2013 Mar 25 346.

116 Hall PS, McCabe C, Stein RC, Cameron D. Economic evaluation of genomic test-directed chemotherapy for early-stage lymph node-positive breast cancer. J Natl Cancer Inst. 2012 Jan 4;104(1):56–66.

117. Joint Formulary Committee. British National Formulary (online). London: BMJ Group and Pharmaceutical Press.

118. Department of Health and Social Care. NHS Reference Costs from 2015 to 2016. 2016;

119. Stein RC, Dunn JA, Bartlett JM, Campbell AF, Marshall A, Hall P, et al. OPTIMA prelim: a randomised feasibility study of personalised care in the treatment of women with early breast cancer. Health Technol Assess (Rockv). 2016 Feb;20(10).

120. Kind P, Dolan P, Gudex C, Williams A. Variations in population health status: results from a United Kingdom national questionnaire survey. BMJ. 1998 Mar 7;316(7133):736-41.

121. Campbell HE, Epstein D, Bloomfield D, Griffin S, Manca A, Yarnold J, et al. The cost-effectiveness of adjuvant chemotherapy for early breast cancer: A comparison of no chemotherapy and first, second, and third generation regimens for patients with differing prognoses. Eur J Cancer. 2011 Nov;47(17).

122. Kirsch J, McGuire A. Establishing health state valuations for disease specific states: an example from heart disease. Health Econ. 2000 Mar;9(2).

123. Castejón N, Cappelleri JC, Cuervo J, Lang K, Mehta P, Mokgokong R, et al. Social preferences for health states associated with acute myeloid leukemia for patients undergoing treatment in the United Kingdom. Health Qual Life Outcomes. 2018 Apr 18;16(1).

124. Baum M, Buzdar A, Cuzick J, Forbes J, Houghton J, Howell A, et al. Anastrozole alone or in combination with tamoxifen versus tamoxifen alone for adjuvant treatment of postmenopausal women with early-stage breast cancer: results of the ATAC (Arimidex, Tamoxifen Alone or in Combination) trial efficacy and safety update analyses. Cancer. 2003 Nov 1;98(9):1802–10.

125. Walkington L, Newsham A, Deverede L. Patterns of breast cancer recurrence and associated health care costs of 1000 patients: a longitudinal study. ncri.co.uk. 2012;

126. Albain KS, Barlow WE, Ravdin PM, Farrar WB, Burton G v., Ketchel SJ, et al. Adjuvant chemotherapy and timing of tamoxifen in postmenopausal patients with endocrine-responsive, node-positive breast cancer: a phase 3, open-label, randomised controlled trial. Lancet. 2009 ;374(9707):2055–63.

127. Office of National Statistics. Interim Life Tables. 2009;

128. Albain K, Anderson S, Arriagada R, Barlow W, Bergh J, Bliss J, et al. Comparisons between different polychemotherapy regimens for early breast cancer: Meta-analyses of long-term outcome among 100 000 women in 123 randomised trials. The Lancet. 2012 Feb 4;379(9814):432–44.

129. Townsend N. Coronary heart disease statistics: A compendium of health statistics. bhf.org.uk. 2012;

130. Cowie MR, Wood DA, Coats AJS, Thompson SG, Suresh V, Poole-Wilson PA, et al. Survival of patients with a new diagnosis of heart failure: a population based study. Heart. 2000;83(5):505.

131. Bhayat F, Das-Gupta E, Smith C, McKeever T, Hubbard R. The incidence of and mortality from leukaemias in the UK: A general population-based study. BMC Cancer. 2009 Jul 26;9(1):1–6.

132. Wolff AC, Blackford AL, Visvanathan K, Rugo HS, Moy B, Goldstein LJ, et al. Risk of Marrow Neoplasms After Adjuvant Breast Cancer Therapy: The National Comprehensive Cancer Network Experience. J Clin Oncol. 2014 Dec 22;1–10.

133. Oliver S. Haematological malignancies in England Cancer Diagnosed 2001 - 2008. nycris.nhs.uk. 2013;

134. Briggs A. Economics notes: Handling uncertainty in economic evaluation. BMJ : British Medical Journal. 1999 Jul 7;319(7202):120.

135. Claxton K, Sculpher M, McCabe C, Briggs A, Akehurst R, Buxton M, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. Health Econ. 2005 Apr;14(4):339–47.

136. McCabe C, Paulden M, Awotwe I, Sutton A, Hall P. One-Way Sensitivity Analysis for Probabilistic Cost-Effectiveness Analysis: Conditional Expected Incremental Net Benefit. Pharmacoeconomics. 2020 Feb 1;38(2):135–41.

137. McCabe C, Tramonti G, Sutton A, Hall P, Paulden M. Probabilistic One-Way Sensitivity Analysis with Multiple Comparators: The Conditional Net Benefit Frontier. Pharmacoeconomics. 2021 Jan;39(1):19-24. doi: 10.1007/s40273-020-00983-7. Epub 2020 Nov 23.

138. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. Health Econ. 2006 Dec;15(12).

139. Zhang X. Application of discrete event simulation in health care: a systematic review. BMC Health Serv Res. 2018 Dec 4;18(1).

140. Ucar I, Smeets B, Azcorra A. simmer: Discrete-Event Simulation for R. 2019.

141. Redaniel MT, Martin RM, Cawthorn S, Wade J, Jeffreys M. The association of waiting times from diagnosis to surgery with survival in women with localised breast cancer in England. 2013.

142. Thompson AM, Moulder-Thompson SL. Neoadjuvant treatment of breast cancer. Annals of Oncology. 2012 Sep;23.

143. Selli C, Sims AH. Neoadjuvant Therapy for Breast Cancer as a Model for Translational Research. Breast Cancer (Auckl). 2019 Jan 19;13.

144. Grant Y, Al-Khudairi R, St John E, Barschkett M, Cunningham D, Al-Mufti R, et al. Patient-level costs in margin re-excision for breast-conserving surgery. 2018;

145. Norum J, Olsen JA, Wist EA. Lumpectomy or mastectomy? Is breast conserving surgery too expensive? Vol. 45, Breast Cancer Research and Treatment. 1997.

146. Johns N, Dixon JM. Should patients with early breast cancer still be offered the choice of breast conserving surgery or mastectomy? European Journal of Surgical Oncology (EJSO). 2016 Nov;42(11).

147. Hwang ES, Lichtensztajn DY, Gomez SL, Fowble B, Clarke CA. Survival after lumpectomy and mastectomy for early stage invasive breast cancer. Cancer. 2013 Apr 1;119(7).

148. Chen K, Liu J, Zhu L, Su F, Song E, Jacobs LK. Comparative effectiveness study of breast-conserving surgery and mastectomy in the general population: A NCDB analysis. Oncotarget. 2015 Nov 24;6(37).

149. NHS National Services Scotland. Breast Cancer Quality Performance Indicators, Patients diagnosed between January 2015 and December 2017. Information Services Division, 2019 Apr 9.

150. Gray E, Marti J, Brewster DH, Wyatt JC, Hall PS; SATURNE Advisory Group. Independent validation of the PREDICT breast cancer prognosis prediction tool in 45,789 patients using Scottish Cancer Registry data. Br J Cancer. 2018 Oct;119(7):808-814.

151. Tan W, Yang M, Yang H, Zhou F, Shen W. Predicting the response to neoadjuvant therapy for early-stage breast cancer: tumour-, blood-, and imaging-related biomarkers. Cancer Manag Res. 2018 Oct 9;10:4333-4347.

152. Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, Geyer CE Jr, Dees EC, Goetz MP, Olson JA Jr, Lively T. Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer. N Engl J Med. 2018 Jul 12;379(2):111-121.

153. Kalinsky K, Barlow WE, Gralow JR, Meric-Bernstam F, Albain KS, Hayes DF, Lin NU, Perez EA, Goldstein LJ. 21-Gene Assay to Inform Chemotherapy Benefit in Node-Positive Breast Cancer. N Engl J Med. 2021 Dec 16;385(25):2336-2347.

154. The Adjuvant Breast Cancer Trials Collaborative Group. Polychemotherapy for early breast cancer: results from the international adjuvant breast cancer chemotherapy randomized trial. J Nat Cancer Inst, 99 (7) (2007), pp. 506-515.

155. Ellis P, Barrett-Lee P, Johnson L. Sequential docetaxel as adjuvant chemotherapy for early breast cancer (TACT): an open-label, phase III, randomised controlled trial. Lancet, 373 (9676) (2009), pp. 1681-1692.

156. Suh W, Hillner BE, Pierce LJ, Hayman JA. Cost-effectiveness of radiation therapy following conservative surgery for ductal carcinoma in situ of the breast. Int J Radiat Oncol Biol Phys, 61 (4) (2005), pp. 1054-1061.

157. Hayman JA, Fairclough DL, Harris JR, Weeks JC. Patient preferences concerning the trade-off between the risks and benefits of routine radiation therapy after conservative surgery for early stage breast cancer. J Clin Oncol, 15 (1997), pp. 1252-1260.

158. Sorensen S, Brown R, Benedict A, Flood E, Revicki D. Patient-rated utilities in postmenopausal early breast cancer: a cross country comparison. ISPOR Seventh Annual European Congress, Hamburg (2004)

159. Kaur MN, Yan J, Klassen AF, David JP, Pieris D, Sharma M, Bordeleau L, Xie F. A Systematic Literature Review of Health Utility Values in Breast Cancer. Med Decis Making. 2022 Jul;42(5):704-719.

160. Turnbull A, Fernando A, Renshaw L, Keys J, Thomas JS, Sims AH, Dixon JM. EA2Clin: A novel immunohistochemical prognostic and predictive test for patients with estrogen receptor-positive breast cancer. Cancer Res (2018) 78 (4_Supplement): P4-08-03.

Appendix

Appendix 1:
Code scripts for the semi-Markov model presented in Chapter 5. The scripts are organized as follows: the "**master.R**" script calls "**SOCdraw.R**" for generating the standard of care parameter draws, and "**model.R**" for generating the function producing the Markov traces. Then it calls for "**OTDXdraw.R**", "**RIGIDdraw.R**", and "**FLEXdraw.R**" to generate the parameter draw associated with OncotypeDX + PREDICT, EER4, and EER4 + PREDICT. Then, "**master.R**" uses the Markov traces to generate the appropriate ICERs, INMB, and CEAF.

master.R:

```
rm(list=ls(all=TRUE))

library(gtools)
library(tidyverse)

{ #Set Global variables

seed <- 19062020      #set seed for random number generator
S <- 7        #number of health states
disc.b <- 0.035 #discount rate for benefits
disc.c <- 0.035 #discount rate for costs
Nsim <- 10000    #Number of simulations
lambda <- 20000 #Threshold
comp <- 0.95    #Compliance rate

source("SOCdraw.R")
source("model.R")

#draw parameters Nsim times

SOCdraw()

#### Montecarlo simulation for therapy guided by Predict ####

{

sim.Pred.high      <- array(c(NA,NA),c(Nsim,2))
sim.Pred.low       <- array(c(NA,NA),c(Nsim,2))
sim.Pred.high.chemo <- array(c(NA,NA),c(Nsim,2))
sim.Pred.low.chemo  <- array(c(NA,NA),c(Nsim,2))

}

for (i in 1:Nsim) {
  sim.Pred.high.chemo[i,] <- model(i,pRec.high.chemo,1,0,)
  sim.Pred.low.chemo[i,]  <- model(i,pRec.low.chemo,1,0,)
  sim.Pred.high[i,]      <- model(i,pRec.high,0,0,)
  sim.Pred.low[i,]       <- model(i,pRec.low,0,0,)

}

costs.Pred <- sim.Pred.high.chemo[,2]*propHigh*comp + sim.Pred.high[,2]*propHigh*(1-comp)+
                sim.Pred.low.chemo[,2]*propLow*(1-comp)+sim.Pred.low[,2]*propLow*(comp)
```

```
QALYs.Pred <- sim.Pred.high.chemo[,1]*propHigh*comp + sim.Pred.high[,1]*propHigh*(1-comp)+
                sim.Pred.low.chemo[,1]*propLow*(1-comp)+sim.Pred.low[,1]*propLow*(comp)


#### Montecarlo simulation for therapy guided by Predict and OncotypeDX (in uncertain cases)

source("OTDXdraw.R")

OTDXdraw()

{

sim.Pred.highB      <- array(c(NA,NA),c(Nsim,2))
sim.Pred.lowB       <- array(c(NA,NA),c(Nsim,2))
sim.Pred.high.chemoB <- array(c(NA,NA),c(Nsim,2))
sim.Pred.low.chemoB  <- array(c(NA,NA),c(Nsim,2))
sim.Pred.med        <- array(c(NA,NA),c(Nsim,2))
sim.Pred.med.chemo   <- array(c(NA,NA),c(Nsim,2))

}

for (i in 1:Nsim) {
  sim.Pred.high.chemoB[i,] <- model(i,pRec.high.chemo,1,0,)
  sim.Pred.low.chemoB[i,]  <- model(i,pRec.low.chemo,1,0,)
  sim.Pred.highB[i,]       <- model(i,pRec.high,0,0,)
  sim.Pred.lowB[i,]        <- model(i,pRec.low,0,0,)
  sim.Pred.med.chemo[i,]   <- model(i,pRec.med.chemo,1,1,ctestDX)
  sim.Pred.med[i,]         <- model(i,pRec.med.chemo,0,1,ctestDX)

}

costs.OTDX <- sim.Pred.high.chemoB[,2]*CTBdistro[,3]*comp +
                sim.Pred.highB[,2]*CTBdistro[,3]*(1-comp)+
                sim.Pred.low.chemoB[,2]*CTBdistro[,1]*(1-comp)+
                sim.Pred.lowB[,2]*CTBdistro[,1]*(comp)+
                sim.Pred.med.chemo[,2]*CTBdistro[,2]*propHighDX +
                sim.Pred.med[,2]*CTBdistro[,2]*propLowDX

QALYs.OTDX <- sim.Pred.high.chemoB[,1]*CTBdistro[,3]*comp +
                  sim.Pred.highB[,1]*CTBdistro[,3]*(1-comp)+
                  sim.Pred.low.chemoB[,1]*CTBdistro[,1]*(1-comp)+
                  sim.Pred.lowB[,1]*CTBdistro[,1]*(comp)+
                  sim.Pred.med.chemo[,1]*CTBdistro[,2]*propHighDX +
                  sim.Pred.med[,1]*CTBdistro[,2]*propLowDX

##### Montecarlo simulation of therapy guided by EER4

source("RIGIDdraw.R")

RIGIDdraw()

{
sim.Rigid.high      <- array(c(NA,NA),c(Nsim,2))
sim.Rigid.low       <- array(c(NA,NA),c(Nsim,2))
sim.Rigid.high.chemo <- array(c(NA,NA),c(Nsim,2))
sim.Rigid.low.chemo  <- array(c(NA,NA),c(Nsim,2))
}
```

```
for (i in 1:Nsim) {
  sim.Rigid.high.chemo[i,] <- model(i,pRec.high.chemo,1,1,cEER4)
  sim.Rigid.low.chemo[i,]  <- model(i,pRec.low.chemo,1,1,cEER4)
  sim.Rigid.high[i,]       <- model(i,pRec.high,0,1,cEER4)
  sim.Rigid.low[i,]        <- model(i,pRec.low,0,1,cEER4)
}

costs.Rigid <- sim.Rigid.high.chemo[,2]*nres.rate*comp + sim.Rigid.high[,2]*nres.rate*(1-comp)+
                  sim.Rigid.low.chemo[,2]*res.rate*(1-comp)+sim.Rigid.low[,2]*res.rate*(comp)

QALYs.Rigid <- sim.Rigid.high.chemo[,1]*nres.rate*comp + sim.Rigid.high[,1]*nres.rate*(1-comp)+
                   sim.Rigid.low.chemo[,1]*res.rate*(1-comp)+sim.Rigid.low[,1]*res.rate*(comp)

#### Montecralo simulation of therapy guided by Predict in with EER4

source("FLEXdraw.R")

FLEXdraw()

{
sim.Flex.R.low        <- array(c(NA,NA),c(Nsim,2))
sim.Flex.R.low.chemo   <- array(c(NA,NA),c(Nsim,2))
sim.Flex.R.med        <- array(c(NA,NA),c(Nsim,2))
sim.Flex.R.med.chemo   <- array(c(NA,NA),c(Nsim,2))
sim.Flex.R.high       <- array(c(NA,NA),c(Nsim,2))
sim.Flex.R.high.chemo  <- array(c(NA,NA),c(Nsim,2))
sim.Flex.NR.low       <- array(c(NA,NA),c(Nsim,2))
sim.Flex.NR.low.chemo  <- array(c(NA,NA),c(Nsim,2))
sim.Flex.NR.med       <- array(c(NA,NA),c(Nsim,2))
sim.Flex.NR.med.chemo  <- array(c(NA,NA),c(Nsim,2))
sim.Flex.NR.high      <- array(c(NA,NA),c(Nsim,2))
sim.Flex.NR.high.chemo <- array(c(NA,NA),c(Nsim,2))

}

for (i in 1:Nsim) {
  sim.Flex.R.low[i,]        <- model(i,pRec.R.low,0,1,cEER4)
  sim.Flex.R.low.chemo[i,]   <- model(i,pRec.R.low.chemo,1,1,cEER4)
  sim.Flex.R.med[i,]        <- model(i,pRec.R.med,0,1,cEER4)
  sim.Flex.R.med.chemo[i,]   <- model(i,pRec.R.med.chemo,1,1,cEER4)
  sim.Flex.R.high[i,]       <- model(i,pRec.R.high,0,1,cEER4)
  sim.Flex.R.high.chemo[i,]  <- model(i,pRec.R.high.chemo,1,1,cEER4)
  sim.Flex.NR.low[i,]       <- model(i,pRec.NR.low,0,1,cEER4)
  sim.Flex.NR.low.chemo[i,]  <- model(i,pRec.NR.low.chemo,1,1,cEER4)
  sim.Flex.NR.med[i,]       <- model(i,pRec.NR.med,0,1,cEER4)
  sim.Flex.NR.med.chemo[i,]  <- model(i,pRec.NR.med.chemo,1,1,cEER4)
  sim.Flex.NR.high[i,]      <- model(i,pRec.NR.high,0,1,cEER4)
  sim.Flex.NR.high.chemo[i,] <- model(i,pRec.NR.high.chemo,1,1,cEER4)

  }

costs.Flex <- sim.Flex.R.low[,2]*res.rate*CTBdistro[,1]*comp +
                  sim.Flex.R.low.chemo[,2]*res.rate*CTBdistro[,1]*(1-comp) +
                  sim.Flex.R.med[,2]*res.rate*CTBdistro[,2]*comp +
                  sim.Flex.R.med.chemo[,2]*res.rate*CTBdistro[,2]*(1-comp) +
                  sim.Flex.R.high[,2]*res.rate*CTBdistro[,3]*(1-comp) +
```

```
        sim.Flex.R.high.chemo[,2]*res.rate*CTBdistro[,3]*comp +
        sim.Flex.NR.low[,2]*nres.rate*CTBdistro[,1]*comp +
        sim.Flex.NR.low.chemo[,2]*nres.rate*CTBdistro[,1]*(1-comp) +
        sim.Flex.NR.med[,2]*nres.rate*CTBdistro[,2]*(1-comp) +
        sim.Flex.NR.med.chemo[,2]*nres.rate*CTBdistro[,2]*comp +
        sim.Flex.NR.high[,2]*nres.rate*CTBdistro[,3]*(1-comp) +
        sim.Flex.NR.high.chemo[,2]*nres.rate*CTBdistro[,3]*comp

QALYs.Flex <- sim.Flex.R.low[,1]*res.rate*CTBdistro[,1]*comp +
                sim.Flex.R.low.chemo[,1]*res.rate*CTBdistro[,1]*(1-comp) +
                sim.Flex.R.med[,1]*res.rate*CTBdistro[,2]*comp +
                sim.Flex.R.med.chemo[,1]*res.rate*CTBdistro[,2]*(1-comp) +
                sim.Flex.R.high[,1]*res.rate*CTBdistro[,3]*(1-comp) +
                sim.Flex.R.high.chemo[,1]*res.rate*CTBdistro[,3]*comp +
                sim.Flex.NR.low[,1]*nres.rate*CTBdistro[,1]*comp +
                sim.Flex.NR.low.chemo[,1]*nres.rate*CTBdistro[,1]*(1-comp) +
                sim.Flex.NR.med[,1]*nres.rate*CTBdistro[,2]*(1-comp) +
                sim.Flex.NR.med.chemo[,1]*nres.rate*CTBdistro[,2]*comp +
                sim.Flex.NR.high[,1]*nres.rate*CTBdistro[,3]*(1-comp) +
                sim.Flex.NR.high.chemo[,1]*nres.rate*CTBdistro[,3]*comp

}


ICER.OTDX <- mean(costs.OTDX - costs.Pred) / mean(QALYs.OTDX - QALYs.Pred)

ICER.Rigid<- mean(costs.Rigid - costs.Pred)/ mean(QALYs.Rigid - QALYs.Pred)

ICER.Flex <- mean(costs.Flex - costs.Pred) / mean(QALYs.Flex - QALYs.Pred)


#Net benefit analysis

NB.Pred  <- QALYs.Pred*lambda - costs.Pred
NB.OTDX  <- QALYs.OTDX*lambda - costs.OTDX
NB.RIGID <- QALYs.Rigid*lambda - costs.Rigid
NB.FLEX  <- QALYs.Flex*lambda - costs.Flex

maxNB <- ifelse(NB.Pred >= NB.OTDX,NB.Pred,NB.OTDX)
maxNB <- ifelse(maxNB >= NB.RIGID,maxNB,NB.RIGID)
maxNB <- ifelse(maxNB >= NB.FLEX,maxNB,NB.FLEX)

CE.Pred  <- ifelse(NB.Pred == maxNB,1,0)
CE.OTDX  <- ifelse(NB.OTDX == maxNB,1,0)
CE.RIGID <- ifelse(NB.RIGID == maxNB,1,0)
CE.FLEX  <- ifelse(NB.FLEX == maxNB,1,0)

prob.CE.Pred  <- mean(CE.Pred, na.rm = TRUE)
prob.CE.OTDX  <- mean(CE.OTDX, na.rm = TRUE)
prob.CE.RIGID <- mean(CE.RIGID, na.rm = TRUE)
prob.CE.FLEX  <- mean(CE.FLEX, na.rm = TRUE)
```

```
##EVPI

EVPI <- mean(maxNB, na.rm = TRUE) - max(mean(NB.Pred, na.rm = TRUE),mean(NB.OTDX, na.rm =
        TRUE),mean(NB.RIGID, na.rm = TRUE),mean(NB.FLEX, na.rm = TRUE))

I <- NA

for (t in 1:10){
  I[t] <- 3700/1.035^t
}

pop.EVPI <- EVPI*sum(I)

## CEAC

lamb <- seq(1000,100000,200)

OUT <- array(NA,c(length(lamb),3))


for (i in 1:length(lamb)) {
  NB.Pred      <- QALYs.Pred*lamb[i] - costs.Pred
  NB.OTDX      <- QALYs.OTDX*lamb[i] - costs.OTDX
  NB.RIGID     <- QALYs.Rigid*lamb[i] - costs.Rigid
  NB.FLEX      <- QALYs.Flex*lamb[i] - costs.Flex
  CE.OTDX      <- ifelse(NB.OTDX>NB.Pred,1,0)
  CE.RIGID     <- ifelse(NB.RIGID>NB.Pred,1,0)
  CE.FLEX      <- ifelse(NB.FLEX>NB.Pred,1,0)
  prob.CE.OTDX  <- mean(CE.OTDX)
  prob.CE.RIGID <- mean(CE.RIGID)
  prob.CE.FLEX  <- mean(CE.FLEX)

  OUT[i,] <- c(prob.CE.OTDX,prob.CE.RIGID,prob.CE.FLEX)
}

###CEAF

OUT <- list()

lamb <- seq(1000,200000,200)

for (i in 1:length(lamb)){
  NB.Pred <- QALYs.Pred*lamb[i] - costs.Pred
  NB.OTDX <- QALYs.OTDX*lamb[i] - costs.OTDX
  NB.RIGID     <- QALYs.Rigid*lamb[i] - costs.Rigid
  NB.FLEX      <- QALYs.Flex*lamb[i] - costs.Flex
  maxNB <- ifelse(NB.Pred >= NB.OTDX,NB.Pred,NB.OTDX)
  maxNB <- ifelse(maxNB >= NB.RIGID,maxNB,NB.RIGID)
  maxNB <- ifelse(maxNB >= NB.FLEX,maxNB,NB.FLEX)
  CE.Pred  <- ifelse(NB.Pred == maxNB,1,0)
  CE.OTDX  <- ifelse(NB.OTDX == maxNB,1,0)
  CE.RIGID <- ifelse(NB.RIGID == maxNB,1,0)
  CE.FLEX  <- ifelse(NB.FLEX == maxNB,1,0)
  prob.CE.Pred  <- mean(CE.Pred)
  prob.CE.OTDX  <- mean(CE.OTDX)
  prob.CE.RIGID <- mean(CE.RIGID)
```

```
  prob.CE.FLEX  <- mean(CE.FLEX)
  EVPI <- mean(maxNB) - max(mean(NB.Pred),mean(NB.OTDX),mean(NB.RIGID),mean(NB.FLEX))

  mean.NB.Pred <- mean(NB.Pred)
  mean.NB.OTDX <- mean(NB.OTDX)
  mean.NB.RIGID <- mean(NB.RIGID)
  mean.NB.FLEX <- mean(NB.FLEX)
  mean.NB.max <- max(mean.NB.Pred,mean.NB.OTDX,mean.NB.RIGID,mean.NB.FLEX)
  ceaf <- ifelse(mean.NB.max==mean.NB.Pred,prob.CE.Pred,0)
  ceaf <- ifelse(mean.NB.max==mean.NB.OTDX,prob.CE.OTDX,ceaf)
  ceaf <- ifelse(mean.NB.max==mean.NB.RIGID,prob.CE.RIGID,ceaf)
  ceaf <- ifelse(mean.NB.max==mean.NB.FLEX,prob.CE.FLEX,ceaf)
  ceaf.test <- ifelse(mean.NB.max==mean.NB.Pred,1,0)
  ceaf.test <- ifelse(mean.NB.max==mean.NB.OTDX,2,ceaf)
  ceaf.test <- ifelse(mean.NB.max==mean.NB.RIGID,3,ceaf)
  ceaf.test <- ifelse(mean.NB.max==mean.NB.FLEX,4,ceaf)

  OUT[[i]] <-

list(NB.Pred=NB.Pred,NB.OTDX=NB.OTDX,NB.RIGID=NB.RIGID,NB.FLEX=NB.FLEX,maxNB=maxNB,

          CE.Pred=CE.Pred,CE.OTDX=CE.OTDX,CE.RIGID=CE.RIGID,CE.FLEX=CE.FLEX,EVPI=EVPI,

          prob.CE.Pred=prob.CE.Pred,prob.CE.OTDX=prob.CE.OTDX,prob.CE.RIGID=prob.CE.RIGID,

          prob.CE.FLEX=prob.CE.FLEX,mean.NB.Pred=mean.NB.Pred,mean.NB.OTDX=mean.NB.OTDX,

         mean.NB.RIGID=mean.NB.RIGID,mean.NB.FLEX=mean.NB.FLEX,ceaf=ceaf,ceaf.test=ceaf.test)
}


CEAF <- data.frame(lambda=lamb,Probability.CE = NA, test=NA)

for (i in 1:length(lamb)){
  CEAF$Probability.CE[i] <- OUT[[i]]$ceaf
  CEAF$CE.Pred[i] <- OUT[[i]]$prob.CE.Pred
  CEAF$CE.OTDX[i] <- OUT[[i]]$prob.CE.OTDX
  CEAF$CE.RIGID[i] <- OUT[[i]]$prob.CE.RIGID
  CEAF$CE.FLEX[i] <- OUT[[i]]$prob.CE.FLEX
  CEAF$test[i] <- OUT[[i]]$ceaf.test
  CEAF$EVPI[i] <- OUT[[i]]$EVPI*sum(I)
  CEAF$Probability.CE[i] <- OUT[[i]]$ceaf

}
```

SOCdraw.R:

```
resp1 <- runif(num.pat,0,1)
resp <- NA
resp[resp1<=0.7]<-1
resp[resp1>0.7]<-0
dat$Response <- resp

startage <- round(mean(dat$Age,rm.na = TRUE))
H <- 100 - startage #set time horizon
T <- H+2 #set length of Markov trace

SOCdraw <- function() {

 set.seed(seed)

 propLow <<- rbeta(Nsim,sum(dat$Chemo.Benefit..10..years <= 3, na.rm = TRUE),
                   sum(dat$Chemo.Benefit..10..years > 3, na.rm = TRUE))
 propHigh <<- 1 - propLow

 RC <<- dat$GRPR
 RSP <<- dat$Response

 # Chemotherapy assignments (OPTIMA trial)

 chemo.type <- rdirichlet(Nsim,c(0*num.pat,
                    round(0.2*num.pat),
                    round(0.1*num.pat),
                    round(0.7*num.pat),
                    0*num.pat,
                    0*num.pat))

 pFEC    <<- chemo.type[,1]
 pFECT   <<- chemo.type[,2]
 pTC     <<- chemo.type[,3]
 pFEC75  <<- chemo.type[,4]
 pEpiCMF <<- chemo.type[,5]
 pFECpw  <<- chemo.type[,6]

#### Transition probabilities
propLR <<- rbeta(Nsim,292,663) #proportion of recurrences that are local ## Disease recurrence

rfs10  <<- (dat$Relapse.free..10..years+dat$Horm.Benefit..10..years.1)/100

CTB    <<- dat$Chemo.Benefit..10..years #Chemotherpay benefit to OS


 mu   <- mean(rfs10) #mean of rfs
 vrnc <- var(rfs10) #variance of rfs
 alpha.h <- ((1-mu)/vrnc-1/mu)*mu^2 #shape parameter for beta distribution
 beta.h  <- alpha.h*(1/mu-1) #shape parameter for beta distribution

 rfs.sim <- array(NA,c(Nsim,(length(rfs10))))

 for (z in 1:length(rfs10)) {rfs.sim[,z] <- rbeta(Nsim,alpha.h,beta.h)}
```

```
rfs.sim <<- rfs.sim #array of estimated rfs under hormone therapy

h.horm.sim <<- -log(rfs.sim)/10 #annual hazard under horm therapy

#chemotherapy effect

alpha <- 71.2115 #shape paramater 1
beta  <- 26.3385 #shape parameter 2

HRchemo <- rbeta(Nsim,alpha,beta) # Hazard Ratio distribution for chemotherapy (mean 0.73)

h.chemo.sim <- array(NA,c(Nsim,length(rfs10)))

for (j in 1:Nsim) {

  h.chemo.sim[j,] <- h.horm.sim[j,]*HRchemo[j]

}

h.chemo.sim <<- h.chemo.sim

predict.high.chemo <<- apply(h.chemo.sim[,CTB>3],1,mean)

predict.high       <<- apply(h.horm.sim[,CTB>3],1,mean)

predict.low        <<- apply(h.horm.sim[,CTB<=3],1,mean)

predict.low.chemo  <<- apply(h.chemo.sim[,CTB<=3],1,mean)

 #Predict High Risk, Chemo

rRec.5  <- predict.high.chemo #annual event rate, to year 5
rRec.10 <- predict.high #annual event rate, from 5 to year 10

# Vector pRec for probability of recurrence by cycle (high risk with chemo)

pRec.high.chemo <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))

for (cycle in 0:T){

pRec.high.chemo[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)

}

pRec.high.chemo <<- pRec.high.chemo

#Predict High Risk, NO Chemo
rRec.5  <- predict.high #annual event rate, to year 5
rRec.10 <- predict.high #annual event rate, from 5 to year 10

# Vector pRec for probability of recurrence by cycle (high risk with NO chemo)

pRec.high <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))

for (cycle in 0:T){

  pRec.high[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
  else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)

}

pRec.high <<- pRec.high
```

```
#Predict Low Risk, NO Chemo

rRec.5  <- predict.low #annual event rate, to year 5
rRec.10 <- predict.low #annual event rate, from 5 to year 10

 # Vector pRec for probability of recurrence by cycle (low risk with NO chemo)

pRec.low <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))

for (cycle in 0:T){

  pRec.low[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
  else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)

}

pRec.low <<- pRec.low

#Predict Low Risk, Chemo

rRec.5  <- predict.low.chemo #annual event rate, to year 5
rRec.10 <- predict.low #annual event rate, from 5 to year 10

# Vector pRec for probability of recurrence by cycle (low risk group with chemo)

pRec.low.chemo <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))

for (cycle in 0:T){

  pRec.low.chemo[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
  else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)

}

pRec.low.chemo <<- pRec.low.chemo

## Constant post-recurrence survival

pDeath_DR <<- rbeta(Nsim,1.00,2.35)

pDeath_DR_low <<- rbeta(Nsim,0.4201557,2.580957) #SWOG 8814 no chemo


## Congestive Heart Failure

pop_CHF <- rep(NA,102)
b1 <- -12.9273605
b2 <- 0.09409354
for (a in 1:102){
 pop_CHF[a] <- exp(b1 + b2*a) #annual probability of developing CHF
}
pdeath_CHF <<- rbeta(Nsim,136,84) #annual probability of death from CHF

hr_CHF_anthra <- exp(rnorm(Nsim,0.458,0.191)) #HR for CHF after anthracycline treatment

prob_CHF <<- array(c(0,0),c(102,Nsim))

r_CHF <- array(c(0,0),c(102,Nsim))
```

```
for (c in 1:102){
  r_CHF[c,] <- (-log(1-pop_CHF[c])*hr_CHF_anthra - (-log(1-pop_CHF[c])))*(pFEC+pFECT)
  prob_CHF[c,] <- 1-exp(-r_CHF[c,])
  prob_CHF[c,] <- ifelse(prob_CHF[c,] >= 0, prob_CHF[c,],0)
}
prob_CHF <<- prob_CHF

## AML

pop_AML <- rep(296/100000,Nsim) #annual rate of AML
rr_AML <- rlnorm(Nsim,1.71,0.79) #rate ratio AML with chemo
prob_AML <<- 1-exp(-pop_AML*rr_AML) #annual excess rate of AML with chemo

rmr_AML <<- -log(1-rbeta(Nsim,1524,60))/5 #excess annual hazard after AML


# lifetable, annual probability of death at given age (minus breast cancer specific mortality)
mr <<- c(0.000353, 0.000193, 0.000161, 0.000117, 0.000096, 0.000098, 0.000082, 0.000090,
0.000078, 0.000093, 0.000097, 0.000100, 0.000118, 0.000119, 0.000158, 0.000173, 0.000245,
0.000271, 0.000257, 0.000241, 0.000264, 0.000263, 0.000247, 0.000294, 0.000287, 0.000337,
0.000311, 0.000358, 0.000381, 0.000416, 0.000406, 0.000487, 0.000539, 0.000575, 0.000598,
0.000644, 0.000727, 0.000795, 0.000897, 0.000980, 0.001056, 0.001151, 0.001267, 0.001341,
0.001530, 0.001648, 0.001828, 0.002067, 0.002151, 0.002559, 0.002692, 0.002861, 0.003158,
0.003537, 0.003755, 0.004141, 0.004390, 0.004717, 0.005303, 0.005696, 0.006452, 0.006907,
0.007798, 0.008516, 0.009178, 0.010084, 0.011251, 0.012330, 0.013535, 0.015072, 0.016561,
0.018374, 0.020832, 0.023483, 0.025871, 0.029214, 0.032712, 0.036780, 0.041825, 0.047053,
0.052661, 0.058476, 0.066223, 0.074507, 0.083203, 0.092440, 0.101085, 0.114035, 0.124557,
0.140443, 0.160738, 0.179041, 0.197859, 0.215442, 0.234414, 0.255550, 0.271716, 0.300529,
0.314942, 0.5,0.5,0.5)
#excess mortality during first year of chemo

mort_chemo <<- rbeta(Nsim,1.6,677.6)

#### Utility Parameters

# female age-group specific norms

U <<- NA
U[1:24] <<- 0.94
U[25:34] <<- 0.93
U[35:44] <<- 0.91
U[45:54] <<- 0.85
U[55:64] <<- 0.81
U[65:74] <<- 0.78
U[75:102] <<- 0.71

#utility decrements

uDFdec     <<- rlnorm(Nsim,-8.117,2.148) #disease-free
uDFdec.chemo<<- rlnorm(Nsim,-2.365,0.325) #disease-free first year of chemo
uLRdec     <<- rlnorm(Nsim,-2.290,0.359) #local recurrence
uDRdec     <<- rlnorm(Nsim,-1.317,0.496) #distant recurrence
uCHF <<- rbeta(Nsim,103.2988905,92.34294761) #CHF utility
uAML <<- rbeta(Nsim,2,2) #estimate of AML utility
```

```
#### Cost parameters

# Recurrence states

cDF     <<- rlnorm(Nsim,6.91,0.004)*1.2
cLR     <<- rlnorm(Nsim,8.72,0.08)*1.2
cDFaLR  <<- rlnorm(Nsim,7.20,0.111)*1.2
cDR     <<- rlnorm(Nsim,7.43,0.019)*1.2
cTerm3  <<- rlnorm(Nsim,7.63,0.003)*1.2
cTerm   <<- cTerm3 - (cDR/4)

# Follow-up

cMedOnc1 <- 197 #first visit consultant-led medical oncology
cMedOnc2 <- 163 # follow-up visit consultant-led
cMammo   <- exp(rnorm(Nsim,3.695585469,0.192924247)) #mammogram
cSpN     <- rnorm(Nsim,100,17.8) #specialist nurse visit

  # Drug treatment cost

cBloods <- 10.1 #cost of blood test
cDeliver <- exp(rnorm(1000,5.53,0.17)) #cost of chemo delivery
cLine <- 23.95 #cost of central line

cFEC <<- 6*(69.24 + 207.07 + cBloods + cDeliver + cSpN)+ cMedOnc1 + cMedOnc2 + cLine
cFEC75 <<- 6*(69.24 + 72.07 + cBloods + cDeliver + cSpN)+ cMedOnc1 + cMedOnc2 + cLine
cTC <<- 4*(75.89 + 150.16+ cBloods + cDeliver + cSpN)+ cMedOnc1 + cMedOnc2 + cLine
cFECT <<- 3*(69.24 + 207.07 + cBloods + cDeliver + cSpN) + 3*(112.2 + 150.16 + cBloods + cDeliver +
cSpN)+ cMedOnc1 + cMedOnc2*2 + cLine
cFECpw <<- 3*(69.24 + 207.07 + cBloods + cDeliver + cSpN)+ 3*(37.87 + 0.14 + cBloods*3 +
cDeliver*3 + cSpN)+ cMedOnc1 + cMedOnc2*2 + cLine
cEpiCMF <<- 4*(28.7 + 0.05 + cBloods + cDeliver + cSpN) + 4*(88.99+3.17 + cBloods*2 + cDeliver*2
+ cSpN)+ cMedOnc1 + cMedOnc2*2 + cLine
cTreat <<- pFEC*cFEC + pFEC75*cFEC75 + pTC*cTC + pFECT*cFECT + pFECpw*cFECpw +
cEpiCMF*pEpiCMF

  #Heart Failure

cCHF <<- exp(rnorm(Nsim,log(2338.71)-(2.5^2)/2,2.5))

#AML

cAML <<- exp(rnorm(Nsim,8.401479,0.85))

  #### Toxicity Parameters ####
# Febrile Neutropoenia

toxNeut.FEC    <- rbeta(Nsim,84,911)

toxNeut.FECT   <- rbeta(Nsim,112,889)

toxNeut.TC     <- rbeta(Nsim,23,483)

toxNeut.EpiCMF <- rbeta(Nsim,137,892)

ctoxNeut.short <- rlnorm(Nsim,6.67,0.46974)

ctoxNeut.long  <- rlnorm(Nsim,8.156,0.03800)
```

```
ctoxNeut      <- ctoxNeut.short*0.5 + ctoxNeut.long*0.5
 # Allergic reaction
toxAll <- rbeta(Nsim,1.46,363.54)
ctoxAll.short <- rlnorm(Nsim,6.19,0.055)
ctoxAll.long  <- rlnorm(Nsim,7.4951,0.0966)
ctoxAll      <- ctoxAll.short*0.5 + ctoxAll.long*0.5
#Nausea
toxNau.FEC    <- rbeta(Nsim,204,791)
toxNau.FECT   <- rbeta(Nsim, 112,889)
toxNau.TC     <- rbeta(Nsim,15,491)
toxNau.EpiCMF <- rbeta(Nsim,24,1005)
ctoxNau.short <- rlnorm(Nsim,5.83,0.018)
ctoxNau.long  <- rlnorm(Nsim,6.749,0.080)
ctoxNau       <- ctoxNau.short*0.5 + ctoxNau.long*0.5
#Diarrhoea
toxDiarr.FEC    <- rbeta(Nsim,1,996)
toxDiarr.FECT   <- rbeta(Nsim,1,1002)
toxDiarr.TC     <- rbeta(Nsim,12,494)
toxDiarr.EpiCMF <- rbeta(Nsim,46,983)
ctoxDiarr.short <- exp(rnorm(Nsim,5.87,0.018))
ctoxDiarr.long  <- exp(rnorm(Nsim,7.009,0.039))
ctoxDiarr       <- ctoxDiarr.short*0.5 + ctoxDiarr.long*0.5
#Anaemia
toxAn.FEC    <- rbeta(Nsim,14,981)
toxAn.FECT   <- rbeta(Nsim,7,994)
toxAn.TC     <- rbeta(Nsim,5,501)
toxAn.EpiCMF <- rbeta(Nsim,31,998)
ctoxAn.short <- rlnorm(Nsim,6.47,0.043)
ctoxAn.long  <- rlnorm(Nsim,6.994,0.129)
ctoxAn       <- ctoxAn.short*0.5 + ctoxAn.long*0.5
#Thrombocytopoenia
```

```
 toxThrom.FEC    <- rbeta(Nsim,3,992)

 toxThrom.FECT   <- rbeta(Nsim,4,997)

 toxThrom.TC     <- rbeta(Nsim,2,504)

 toxThrom.EpiCMF <- rbeta(Nsim,10,1019)

 ctoxThrom.short <- rlnorm(Nsim,6.29,0.0001)

 ctoxThrom.long  <- rlnorm(Nsim,7.141,0.274)

 ctoxThrom       <- ctoxThrom.short*0.5 + ctoxThrom.long*0.5

 #Stomatitis

 toxStom.FEC    <- rbeta(Nsim,40,995)

 toxStom.FECT   <- rbeta(Nsim,59,942)

 toxStom.TC     <- rbeta(Nsim,4,502)

 toxStom.EpiCMF <- rbeta(Nsim,1,1030)

 ctoxStom.short <- rlnorm(Nsim,5.95,0.144)

 ctoxStom.long  <- rlnorm(Nsim,7.33,0.20116)

 ctoxStom       <- ctoxStom.short*0.5 + ctoxStom.long*0.5

 cTox.FEC <- toxNeut.FEC*ctoxNeut + toxNau.FEC*ctoxNau + toxDiarr.FEC*ctoxDiarr +
toxAn.FEC*ctoxAn + toxThrom.FEC*ctoxThrom + toxStom.FEC*ctoxStom

 cTox.FEC75 <- cTox.FEC*0.66666

 cTox.FECT <- toxNeut.FECT*ctoxNeut + toxNau.FECT*ctoxNau + toxDiarr.FECT*ctoxDiarr +
toxAn.FECT*ctoxAn + toxThrom.FECT*ctoxThrom + toxStom.FECT*ctoxStom

 cTox.TC <- toxNeut.TC*ctoxNeut + toxNau.TC*ctoxNau + toxDiarr.TC*ctoxDiarr + toxAn.TC*ctoxAn
+ toxThrom.TC*ctoxThrom + toxStom.TC*ctoxStom

 cTox.FECpw <- cTox.FECT

 cTox.EpiCMF <- toxNeut.EpiCMF*ctoxNeut + toxNau.EpiCMF*ctoxNau + toxDiarr.EpiCMF*ctoxDiarr
+ toxAn.EpiCMF*ctoxAn + toxThrom.EpiCMF*ctoxThrom + toxStom.EpiCMF*ctoxStom

 cTox <<- (cTox.FEC75*pFEC75 + cTox.FEC*pFEC + cTox.FECT*pFECT + cTox.TC*pTC +
cTox.FECpw*pFECpw + cTox.EpiCMF*pEpiCMF)*1.2

 }
```

model.R:

```
#### Markov transition matrix and Trace ####

tps <- array(NA,c(S,S,T))
trace <- matrix(nrow = T, ncol = S)
qtime <- rep(NA,len=T)
cost <- rep(NA,len=T)

model <- function(i,pRec,chemo,test,ctest=0){

  T <- H+2
## Transition Matrix

 ## States: 1=Disease-free, 2=Local Recurrence, 3=Disease-free after lcoal recurrence
 ##       4=Distant Recurrence, 5=Congestive Heart Failure, 6=AML, 7=Death

 for (t in 1:T){
tps[1,1,t] <- 1-(1-exp(- mr[startage + t]-ifelse(chemo==1 & t==1,log(1-mort_chemo[i]),0))) - pRec[t,i]
               - prob_CHF[startage + t,i] - prob_AML[i]
tps[1,2,t] <- pRec[t,i]*propLR[i]
tps[1,3,t] <- 0
tps[1,4,t] <- pRec[t,i]*(1-propLR[i])
tps[1,5,t] <- prob_CHF[startage + t,i]
tps[1,6,t] <- prob_AML[i]
tps[1,7,t] <- 1-exp( - mr[startage+t]-(ifelse(chemo==1 & t==1,log(1-mort_chemo[i]),0)))

tps[2,1,t] <- 0
tps[2,2,t] <- 0
tps[2,3,t] <- 1-mr[startage + t]- pRec[t,i]*(1-propLR[i])
tps[2,4,t] <- pRec[t,i]*(1-propLR[i])
tps[2,5,t] <- 0
tps[2,6,t] <- 0
tps[2,7,t] <- mr[startage + t]

tps[3,1,t] <- 0
tps[3,2,t] <- 0
tps[3,3,t] <- 1-mr[startage + t] - pRec[t,i]*(1-propLR[i]) - ifelse(chemo==1,prob_CHF[startage+ t,i],0)
              - ifelse(chemo==1, prob_AML[i],0)
tps[3,4,t] <- pRec[t,i]*(1-propLR[i])
tps[3,5,t] <- ifelse(chemo==1, prob_CHF[startage + t,i],0)
tps[3,6,t] <- ifelse(chemo==1, prob_AML[i],0)
tps[3,7,t] <- mr[startage + t]

tps[4,1,t] <- 0
tps[4,2,t] <- 0
tps[4,3,t] <- 0
tps[4,4,t] <- 1-ifelse(chemo==0,pDeath_DR_low[i],pDeath_DR[i])
tps[4,5,t] <- 0
tps[4,6,t] <- 0
tps[4,7,t] <- ifelse(chemo==0,pDeath_DR_low[i], pDeath_DR[i])

tps[5,1,t] <- 0
tps[5,2,t] <- 0
tps[5,3,t] <- 0
tps[5,4,t] <- 0
```

```
 tps[5,5,t] <- 1- pdeath_CHF[i]
 tps[5,6,t] <- 0
 tps[5,7,t] <- pdeath_CHF[i]

 tps[6,1,t] <- 0
 tps[6,2,t] <- 0
 tps[6,3,t] <- 0
 tps[6,4,t] <- 0
 tps[6,5,t] <- 0
 tps[6,6,t] <- 1-(1-exp(-(mr[startage+t]+rmr_AML[i])))
 tps[6,7,t] <- 1-exp(-(mr[startage+t]+rmr_AML[i]))

 tps[7,1,t] <- 0
 tps[7,2,t] <- 0
 tps[7,3,t] <- 0
 tps[7,4,t] <- 0
 tps[7,5,t] <- 0
 tps[7,6,t] <- 0
 tps[7,7,t] <- 1

 }

 #### Markov Trace ####

 # trace for t = 1 [all start in state 1]
 trace[1,1] <- 1
 trace[1,-1] <- 0
 # trace for t=> 2
 for (t in 2:T) {
  trace[t,] <- trace[t-1,] %*% tps[,,t]
 }
 #### Output ####

  # QALY

 for (t in 2:T) {

  qtime[t] <- ((if (t==2) trace[t,1]*(ifelse(chemo==1,U[startage+t]-uDFdec.chemo[i],U[startage+t]-
uDFdec[i])) else trace[t,1]*(U[startage+t]-uDFdec[i])) + trace[t,2]*(U[startage+t]-uLRdec[i]) +
trace[t,3]*(U[startage + t]-uDFdec[i]) + trace[t,4]*(U[startage + t]-uDRdec[i]) + trace[t,5]*uCHF[i] +
trace[t,6]*uAML[i])/((1+disc.b)^(t-1))  }

 QALYs <- qtime[2]/2+sum(qtime[3:(T-1)])+qtime[T]/2

  # Costs

 for (t in 2:T) {

  cost[t] <- (  ((if (t==2) trace[t,1]*ifelse(chemo==1,cTreat[i]+cTox[i]+cDF[i]/2,cDF[i]/2) else
trace[t,1]*ifelse(t<=10,cDF[i],0))  + trace[t,2]*cLR[i] + trace[t,3]*cDF[i]  + trace[t,4]*cDR[i]  +
trace[t,5]*cCHF[i] + trace[t,6]*cAML[i] + if(t==1) 0 else trace[t-1,4]*tps[4,6,t-1]*cTerm[i] )
/((1+disc.c)^(t-1)))  }

 COSTs <- sum(cost[2:(T-1)])+cost[T]/2+ifelse(test==1,ctest,0)
 #Model return:
 return(c(QALYs,COSTs))
}
```

OTDXdraw.R:

```
#Therapy guided by Predict in combination with OncotypeDX in uncertain cases

set.seed(seed)

CTB1 <- NA
CTB1[CTB<3]<-0
CTB1[CTB>=3 & CTB<5]<-1
CTB1[CTB>=5]<-2
CTB1 <- as.integer(CTB1)
RT <- table(CTB1,RC)

CTBdistro <- rdirichlet(Nsim,c(sum(CTB1==0),sum(CTB1==1),sum(CTB1==2)))
propLowDX <<- rbeta(Nsim, RT[2,1],RT[2,2]+RT[2,3]) #Proportion of low risk vs high risk
propHighDX <<- 1-propLowDX
ctestDX <<- rep(1250,Nsim) #cost of OncotypeDX

OTDXdraw <- function() {
 #Recurrence rates
 predict.high.chemo <<- apply(h.chemo.sim[,CTB>=5],1,mean)
 predict.high      <<- apply(h.horm.sim[,CTB>=5],1,mean)
 predict.low.chemo  <<- apply(h.chemo.sim[,CTB<3],1,mean)
 predict.low       <<- apply(h.horm.sim[,CTB<3],1,mean)
 predict.med.chemo  <<- apply(h.chemo.sim[,CTB>3 & RC>=1],1,mean)
 predict.med       <<- apply(h.horm.sim[,CTB>3 & RC<1],1,mean)

 #Predict High Risk, Chemo
 rRec.5  <- predict.high.chemo #annual event rate, to year 5
 rRec.10 <- predict.high #annual event rate, from 5 to year 10

 # Vector pRec for probability of recurrence by cycle (high risk with chemo)
 pRec.high.chemo <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))
 for (cycle in 0:T){
 pRec.high.chemo[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
                            else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)

 }

 pRec.high.chemo <<- pRec.high.chemo

 #Predict High Risk, NO Chemo
 rRec.5  <- predict.high #annual event rate, to year 5
 rRec.10 <- predict.high #annual event rate, from 5 to year 10



 # Vector pRec for probability of recurrence by cycle (high risk with NO chemo)
 pRec.high <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))
 for (cycle in 0:T){
  pRec.high[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
                       else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)
 }

 pRec.high <<- pRec.high
```

```
#Predict Low Risk, NO Chemo
rRec.5  <- predict.low #annual event rate, to year 5
rRec.10 <- predict.low #annual event rate, from 5 to year 10

# Vector pRec for probability of recurrence by cycle (low risk with NO chemo)
pRec.low <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))
for (cycle in 0:T){
 pRec.low[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
            else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)
}

pRec.low <<- pRec.low

#Predict Low Risk, Chemo
rRec.5  <- predict.low.chemo #annual event rate, to year 5
rRec.10 <- predict.low #annual event rate, from 5 to year 10


# Vector pRec for probability of recurrence by cycle (low risk group with chemo)
pRec.low.chemo <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))
for (cycle in 0:T){
  pRec.low.chemo[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
                      else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)
}

pRec.low.chemo <<- pRec.low.chemo

#Predict High Risk, Chemo
rRec.5  <- predict.high.chemo #annual event rate, to year 5
rRec.10 <- predict.high #annual event rate, from 5 to year 10

# Vector pRec for probability of recurrence by cycle (high risk with chemo)
pRec.high.chemo <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))
for (cycle in 0:T){
  pRec.high.chemo[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
                        else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)
}

pRec.high.chemo <<- pRec.high.chemo

#Predict Uncertain Risk, OTDX low risk, no chemo
rRec.5  <- predict.med #annual event rate, to year 5
rRec.10 <- predict.med #annual event rate, from 5 to year 10

# Vector pRec for probability of recurrence by cycle (uncertain risk with NO chemo
pRec.med <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))

for (cycle in 0:T){
  pRec.med[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
                    else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)
}
pRec.med <<- pRec.med

#Predict High Risk, Chemo
rRec.5  <- predict.high.chemo #annual event rate, to year 5
rRec.10 <- predict.high #annual event rate, from 5 to year 10
```

```
 # Vector pRec for probability of recurrence by cycle (high risk with chemo)
 pRec.high.chemo <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))
 for (cycle in 0:T){
   pRec.high.chemo[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
          else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)
 }
 pRec.high.chemo <<- pRec.high.chemo


 #Predict Uncertain Risk, OTDX high, Chemo
 rRec.5  <- predict.med.chemo #annual event rate, to year 5
 rRec.10 <- predict.high #annual event rate, from 5 to year 10

  # Vector pRec for probability of recurrence by cycle (high risk with NO chemo)
 pRec.med.chemo <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))
 for (cycle in 0:T){
   pRec.med.chemo[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
                  else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)
 }

 pRec.med.chemo <<- pRec.med.chemo

}
```

RIGIDdraw.R:

```
# EER4 Rigid

res.rate  <<- rbeta(Nsim,209,89) #average response rate to endocrine therapy
nres.rate <<- 1-res.rate

cEER4 <<- (rlnorm(Nsim,7.422, 0.030)) #cost of test

RIGIDdraw <- function() {

 #endocrine effect on Responders

 HR_resp <- rbeta(length(rfs10),397,738)
 e.horm.sim <- array(NA,c(Nsim,length(rfs10)))
 for (z in 1:length(rfs10)) {
   e.horm.sim[,z] <- h.horm.sim[,z]*HR_resp[z]
 }

 e.horm.sim <<- e.horm.sim


 #chemotherapy effect on Non-Responders
 alpha <- 71.2115 #shape paramater 1
 beta  <- 26.3385 #shape parameter 2

 HRchemo <- rbeta(Nsim,alpha,beta) # Hazard Ratio distribution for chemotherapy (mean 0.73)
 HRn_resp <- 1/HR_resp

 e.chemo.sim <- array(NA,c(Nsim,length(rfs10)))
```

```
for (j in 1:Nsim) {
  e.chemo.sim[j,] <- e.horm.sim[j,]*HRn_resp[j]*HRchemo[j]
}

e.chemo.sim <<- e.chemo.sim

#Recurrence rates

Rigid.high.chemo <<- apply(e.chemo.sim[,RSP==0],1,mean)
Rigid.high       <<- apply(e.horm.sim[,RSP==0],1,mean)
Rigid.low.chemo  <<- apply(e.chemo.sim[,RSP==1],1,mean)
Rigid.low        <<- apply(e.horm.sim[,RSP==1],1,mean)

#EER4 High Risk, Chemo
rRec.5  <- Rigid.high.chemo #annual event rate, to year 5
rRec.10 <- Rigid.high #annual event rate, from 5 to year 10

# Vector pRec for probability of recurrence by cycle (high risk with chemo)
pRec.high.chemo <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))
for (cycle in 0:T){
  pRec.high.chemo[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
                  else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)
}

pRec.high.chemo <<- pRec.high.chemo

#EER4 High Risk, NO Chemo
rRec.5  <- Rigid.high #annual event rate, to year 5
rRec.10 <- Rigid.high #annual event rate, from 5 to year 10

# Vector pRec for probability of recurrence by cycle (high risk with NO chemo)
pRec.high <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))
for (cycle in 0:T){
  pRec.high[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
                 else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)
}
pRec.high <<- pRec.high

#EER4 Low Risk, NO Chemo
rRec.5  <- Rigid.low #annual event rate, to year 5
rRec.10 <- Rigid.low #annual event rate, from 5 to year 10

# Vector pRec for probability of recurrence by cycle (low risk with NO chemo)
pRec.low <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))
for (cycle in 0:T){
  pRec.low[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
                  else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)
}

pRec.low <<- pRec.low

#EEr4 Low Risk, Chemo
rRec.5  <- Rigid.low.chemo #annual event rate, to year 5
rRec.10 <- Rigid.low #annual event rate, from 5 to year 10
```

```
 # Vector pRec for probability of recurrence by cycle (low risk group with chemo)
 pRec.low.chemo <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))
 for (cycle in 0:T){
   pRec.low.chemo[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
                        else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)
 }

 pRec.low.chemo <<- pRec.low.chemo

}
```

FLEXdraw.R:

```
#### EER4 Flexible ####

FLEXdraw <- function() {

   #Three Predict Risk Categories, Two EER4 risk categories, Chemo assigned based on Predict

 FLEX.R.low   <<- apply(e.horm.sim[,RSP==1 & CTB<3],1,mean)
 FLEX.R.med   <<- apply(e.horm.sim[,RSP==1 & CTB>=3 & CTB<5],1,mean)
 FLEX.R.high  <<- apply(e.horm.sim[,RSP==1 & CTB>=5],1,mean)
 FLEX.NR.low  <<- apply(e.horm.sim[,RSP==0 & CTB<3],1,mean)
 FLEX.NR.med  <<- apply(e.horm.sim[,RSP==0 & CTB>=3 & CTB<5],1,mean)
 FLEX.NR.high <<- apply(e.horm.sim[,RSP==0 & CTB>=5],1,mean)

 FLEX.R.low.chemo   <<- apply(e.chemo.sim[,RSP==1 & CTB<3],1,mean)
 FLEX.R.med.chemo   <<- apply(e.chemo.sim[,RSP==1 & CTB>=3 & CTB<5],1,mean)
 FLEX.R.high.chemo  <<- apply(e.chemo.sim[,RSP==1 & CTB>=5],1,mean)
 FLEX.NR.low.chemo  <<- apply(e.chemo.sim[,RSP==0 & CTB<3],1,mean)
 FLEX.NR.med.chemo  <<- apply(e.chemo.sim[,RSP==0 & CTB>=3 & CTB<5],1,mean)
 FLEX.NR.high.chemo <<- apply(e.chemo.sim[,RSP==0 & CTB>=5],1,mean)

 #Predict low risk, EER4 low risk, NO chemo
 rRec.5  <- FLEX.R.low #annual event rate, to year 5
 rRec.10 <- FLEX.R.low #annual event rate, from 5 to year 10

 # Vector pRec for probability of recurrence by cycle
 pRec.R.low <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))
 for (cycle in 0:T){
   pRec.R.low[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
         else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)
 }

 pRec.R.low <<- pRec.R.low

 #Predict low risk, EER4 low risk, with chemo
 rRec.5  <- FLEX.R.low.chemo #annual event rate, to year 5
 rRec.10 <- FLEX.R.low #annual event rate, from 5 to year 10

 # Vector pRec for probability of recurrence by cycle
 pRec.R.low.chemo <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))
 for (cycle in 0:T){
   pRec.R.low.chemo[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
```

```
            else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)
}

pRec.R.low.chemo <<- pRec.R.low.chemo

#Predict medium risk, EER4 low risk, NO chemo
rRec.5  <- FLEX.R.med #annual event rate, to year 5
rRec.10 <- FLEX.R.med #annual event rate, from 5 to year 10

# Vector pRec for probability of recurrence by cycle
pRec.R.med <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))
for (cycle in 0:T){
  pRec.R.med[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
                  else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)
}

pRec.R.med <<- pRec.R.med

#Predict medium risk, EER4 low risk, with chemo
rRec.5  <- FLEX.R.med.chemo #annual event rate, to year 5
rRec.10 <- FLEX.R.med #annual event rate, from 5 to year 10

# Vector pRec for probability of recurrence by cycle
pRec.R.med.chemo <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))
for (cycle in 0:T){
  pRec.R.med.chemo[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
                  else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)
}

pRec.R.med.chemo <<- pRec.R.med.chemo

#Predict high risk, EER4 low risk, NO chemo
rRec.5  <- FLEX.R.high #annual event rate, to year 5
rRec.10 <- FLEX.R.high #annual event rate, from 5 to year 10

# Vector pRec for probability of recurrence by cycle
pRec.R.high <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))
for (cycle in 0:T){
  pRec.R.high[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
               else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)
}

pRec.R.high <<- pRec.R.high

#Predict high risk, EER4 low risk, with chemo
rRec.5  <- FLEX.R.high.chemo #annual event rate, to year 5
rRec.10 <- FLEX.R.high #annual event rate, from 5 to year 10

# Vector pRec for probability of recurrence by cycle
pRec.R.high.chemo <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))
for (cycle in 0:T){
  pRec.R.high.chemo[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
                   else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)
}

pRec.R.high.chemo <<- pRec.R.high.chemo
```

```
### #Predict low risk, EER4 high risk, NO chemo
rRec.5  <- FLEX.NR.low #annual event rate, to year 5
rRec.10 <- FLEX.NR.low #annual event rate, from 5 to year 10

# Vector pRec for probability of recurrence by cycle
pRec.NR.low <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))
for (cycle in 0:T){
  pRec.NR.low[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
                else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)
}

pRec.NR.low <<- pRec.NR.low

#Predict low risk, EER4 high risk, with chemo
rRec.5  <- FLEX.NR.low.chemo #annual event rate, to year 5
rRec.10 <- FLEX.NR.low #annual event rate, from 5 to year 10

# Vector pRec for probability of recurrence by cycle
pRec.NR.low.chemo <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))
for (cycle in 0:T){
  pRec.NR.low.chemo[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
            else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)
}

pRec.NR.low.chemo <<- pRec.NR.low.chemo

#Predict medium risk, EER4 high risk, NO chemo
rRec.5  <- FLEX.NR.med #annual event rate, to year 5
rRec.10 <- FLEX.NR.med #annual event rate, from 5 to year 10

# Vector pRec for probability of recurrence by cycle
pRec.NR.med <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))
for (cycle in 0:T){
  pRec.NR.med[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)
}
pRec.NR.med <<- pRec.NR.med

  #Predict medium risk, EER4 high risk, with chemo
rRec.5  <- FLEX.NR.med.chemo #annual event rate, to year 5
rRec.10 <- FLEX.NR.med #annual event rate, from 5 to year 10

# Vector pRec for probability of recurrence by cycle
pRec.NR.med.chemo <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))
for (cycle in 0:T){
  pRec.NR.med.chemo[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
                else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)
}

pRec.NR.med.chemo <<- pRec.NR.med.chemo

#Predict high risk, EER4 high risk, NO chemo
rRec.5  <- FLEX.NR.high #annual event rate, to year 5
rRec.10 <- FLEX.NR.high #annual event rate, from 5 to year 10
```

```
 # Vector pRec for probability of recurrence by cycle
 pRec.NR.high <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))
 for (cycle in 0:T){
   pRec.NR.high[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
               else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)
 }

 pRec.NR.high <<- pRec.NR.high

 #Predict high risk, EER4 high risk, with chemo
 rRec.5  <- FLEX.NR.high.chemo #annual event rate, to year 5
 rRec.10 <- FLEX.NR.high #annual event rate, from 5 to year 10

 # Vector pRec for probability of recurrence by cycle
 pRec.NR.high.chemo <- array(c(c(0:T),rep(NA,Nsim)),c(T,Nsim))
 for (cycle in 0:T){
   pRec.NR.high.chemo[cycle,] <- if (cycle <=5) 1-exp(-rRec.5)
               else (if(cycle >5 & cycle <=10) 1-exp(-rRec.10) else 0)
 }

 pRec.NR.high.chemo <<- pRec.NR.high.chemo

}
```

Appendix 2:

Script for executing the Discrete Event Simulation model, using the “simmer” package.

```
#### Discrete Event Simulation ####

library(simmer)
library(tidyverse)
# set global variables
{
 set.seed(413)
 Nsim <- 1000
 s.distro <- rlnorm(Nsim,3,0.5) #initial size distribution
 s.neg.distro <- -rlnorm(Nsim,2.6317,0.5368) #decrease in tumour size for responders
 s.pos.distro <- rlnorm(Nsim,2.33,0.5768) #increase in tumour size for non-responders
 surg.wait <- rlnorm(Nsim,3.08,0.3) #diagnosis to surgery waiting time distribution
 neo.time <- rnorm(Nsim,150,30) #time spent in neoadjuvant treatment
 response <- 0.7 #share of patients responding to Letrozole
 p.gen <- runif(Nsim,0,1) #probability generator for random draws
 sensitivity <- 0.96 # of EER4
 specificity <- 0.94 # of EER4
 l.cost <- rlnorm(Nsim,8.699,0.2) #cost of lumpectomy
 m.cost <- rlnorm(Nsim,9.30,0.03) # cost of mastectomy + recon surgery
 QQ <- rlnorm(Nsim,-3.506,0.5)
 l.qaly <- QQ/0.045244
 cEER4M <- rlnorm(Nsim,9.52,0.05) #cost distro from Markov
 qEER4M <- rlnorm(Nsim,2.076,0.03) #QALY distro from Markov
 cSOCM <- rlnorm(Nsim,9.543,0.05) #cost distro from Markov
 qSOCM <- rlnorm(Nsim,2.039,0.03) #QALY distro from markov

}

# set simulation environment

clinic <- simmer()

# set SOC trajectory

SOC <-
 trajectory("SOC") %>%
 log_("Diagnosis") %>%
 set_attribute("size", function(){sample(s.distro,1)}) %>% #randomise size
 set_attribute("Treatment",0) %>% # no neoadjuvant treatment
 set_attribute("Surgery",function()ifelse(get_attribute(clinic,"size")>20,1,2)) %>% # type of surgery
 timeout(function(){sample(surg.wait,1)}) %>% # time from diagnosis to surgery
 log_("Surgery completed")

AllAI <-
 trajectory("AllAI") %>%
 log_("Diagnosis") %>%
 set_attribute("size", function(){sample(s.distro,1)}) %>% #randomise size
 set_attribute("Treatment",1) %>% # All receive neoadjuvant treatment
```

```
 set_attribute("Response",function()ifelse(sample(p.gen,1)<response,1,2)) %>% #randomise
response status, responder=1, non-responder=2
 log_("AI treatment") %>%
 branch(
  function() get_attribute(clinic,"Response"),continue=c(TRUE,TRUE),
   trajectory("Responder") %>%
    set_attribute("size",sample(s.neg.distro,1),mod="+") %>% #response to treatment
    log_("Responder"),   trajectory("Non-responder") %>%
    set_attribute("size",sample(s.pos.distro,1),mod="+") %>% #no response to treatment
    log_("Non-responder")) %>%
    set_attribute("Surgery",function()ifelse(get_attribute(clinic,"size")>20,1,2)) %>% # type of surgery
    timeout(function(){sample(surg.wait,1)}) %>% # time from diagnosis to surgery
    log_("Surgery completed")

# set trajectory for EER4

EER4 <-
 trajectory("EER4") %>%
 log_("Diagnosis") %>%
 set_attribute("size", function(){sample(s.distro,1)}) %>% #randomise size
 set_attribute("Response",function()ifelse(sample(p.gen,1)<response,1,2)) %>% #randomise
response status, responder=1, non-responder=2
 log_("AI treatment") %>%
timeout(14) %>%
 branch(
function() get_attribute(clinic,"Response"), continue=c(TRUE,TRUE),
trajectory("Predicted_Responder") %>%
set_attribute("Treatment",1) %>%
  log_("PR") %>%
  branch(
   function() ifelse(sample(p.gen,1)<sensitivity,1,2),continue=c(TRUE,TRUE),
              trajectory("True Responder") %>%
    set_attribute("size",sample(s.neg.distro,1),mod="+") %>%
   log_("True Responder"),  trajectory("False Responder") %>%
   set_attribute("size",sample(s.pos.distro,1),mod="+") %>%
  log_("False Responder")
),  trajectory("Predicted_Non-responder") %>%
   set_attribute("Treatment",0) %>%
 log_("PnR") %>%   branch(
    function() ifelse(sample(p.gen,1)<specificity,1,2),continue=c(TRUE,TRUE),
trajectory("True Non-responder") %>%
 log_("True Non-responder"),   trajectory("False Non-responder") %>%
   log_("False Non-responder")  )  ) %>%
set_attribute("Surgery",function()ifelse(get_attribute(clinic,"size")>20,1,2)) %>% # type of surgery
timeout(function(){sample(surg.wait,1)}) %>% # time from diagnosis to surgery
 log_("Surgery completed")
```

```
clinic <- simmer("Clinic") %>%
  add_generator("SOC ",SOC, from_to(0,1000, function(){1}),mon=2) %>%
  add_generator("AllAI ",AllAI, from_to(0,1000, function(){1}),mon=2) %>%
  add_generator("EER4 ",EER4, from_to(0,1000, function(){1}),mon=2)

attributes <- run(clinic,until = 1500) %>%
  get_mon_attributes()

arrivals   <- run(clinic,until = 1500) %>%
 get_mon_arrivals()

Results <- attributes %>%
  select(name,key,value) %>%
  group_by(name) %>%
  pivot_wider(names_from = key, values_from = value) %>%
  full_join(arrivals) %>%
  select(name,activity_time,size,Response,Treatment,Surgery)

Results$Surgery <- ifelse(Results$Surgery==1,"Mastectomy","Lumpectomy")

Results$Strategy <-
ifelse(substring(Results$name,1,3)=="SOC","SOC",ifelse(substring(Results$name,1,3)=="EER","EER4"
,"AllAI"))

ResultsSOC <- Results %>%
  filter(Strategy=="SOC") %>%
  mutate(cost.surg=ifelse(Surgery=="Mastectomy",m.cost,l.cost)) %>%
  mutate(cost.treat = as.numeric(Treatment)*activity_time*(2.75/28)) %>%
  mutate(extraQ = ifelse(Surgery=="Lumpectomy",l.qaly,0))

ResultsSOC$mkv.cost <- cSOCM
ResultsSOC$mkv.qaly <- qSOCM
ResultsSOC <- ResultsSOC %>%
  mutate(TQALY = mkv.qaly + extraQ) %>%
  mutate(Tcost = mkv.cost + cost.surg + cost.treat)

ResultsAllAI <- Results %>%
  filter(Strategy=="AllAI") %>%
  mutate(cost.surg=ifelse(Surgery=="Mastectomy",m.cost,l.cost)) %>%
  mutate(cost.treat = as.numeric(Treatment)*activity_time*(2.75/28)) %>%
  mutate(extraQ = ifelse(Surgery=="Lumpectomy",l.qaly,0))

ResultsAllAI$mkv.cost <- cSOCM
ResultsAllAI$mkv.qaly <- qSOCM
ResultsAllAI <- ResultsAllAI %>%
 mutate(TQALY = mkv.qaly + extraQ) %>%
  mutate(Tcost = mkv.cost + cost.surg + cost.treat)

ResultsEER4 <- Results %>%
  filter(Strategy=="EER4") %>%
  mutate(cost.surg=ifelse(Surgery=="Mastectomy",m.cost,l.cost)) %>%
```

```
  mutate(cost.treat = as.numeric(Treatment)*activity_time*(2.75/28)) %>%
  mutate(extraQ = ifelse(Surgery=="Lumpectomy",l.qaly,0))

ResultsEER4$mkv.cost <- cEER4M
ResultsEER4$mkv.qaly <- qEER4M
ResultsEER4 <- ResultsEER4 %>%
  mutate(TQALY = mkv.qaly + extraQ) %>%
  mutate(Tcost = mkv.cost + cost.surg + cost.treat)

{

ICERtb <- ResultsSOC %>%
  select(Strategy,TQALY,Tcost)

ICERtb1 <- ResultsAllAI %>%
  select(Strategy,TQALY,Tcost)

ICERtb2 <- ResultsEER4 %>%
  select(Strategy,TQALY,Tcost)

ICERtb <- ICERtb %>%
  rbind(ICERtb1,ICERtb2)

}

ICERtb$ICER <- (ICERtb$Tcost - ICERtb$Tcost[ICERtb$Strategy=="SOC"])/(ICERtb$TQALY -
ICERtb$TQALY[ICERtb$Strategy=="SOC"])

ICERplot <- ICERtb %>%
  filter(Strategy!="SOC")

mean(ResultsSOC$TQALY)
mean(ResultsAllAI$TQALY)
mean(ResultsEER4$TQALY)

mean(ResultsSOC$Tcost)
mean(ResultsAllAI$Tcost)
mean(ResultsEER4$Tcost)



(mean(ICERtb1$Tcost - ICERtb$Tcost))/(mean(ICERtb1$TQALY - ICERtb$TQALY))
(mean(ICERtb2$Tcost - ICERtb$Tcost))/(mean(ICERtb2$TQALY - ICERtb$TQALY))

surgcount <- Results %>%
  group_by(Strategy) %>%
  count(Surgery) %>%
  pivot_wider(names_from = Surgery, values_from = n)
surgcount$displacement <- surgcount$Lumpectomy - surgcount$Lumpectomy[3]
surgcount$savings <- surgcount$displacement*(mean(m.cost - l.cost))
surgcount$avgsavings <- surgcount$savings/1000

lambda <- seq(1000,100000,200)
CEAC <- array(NA,c(length(lambda),2))
```

```
for (i in 1:length(lambda)) {
  NB.SOC      <- ResultsSOC$TQALY*lambda[i]   - ResultsSOC$Tcost
  NB.AllAI    <- ResultsAllAI$TQALY*lambda[i] - ResultsAllAI$Tcost
  NB.EER4     <- ResultsEER4$TQALY*lambda[i]  - ResultsEER4$Tcost
  CE.AllAI    <- ifelse(NB.AllAI > NB.SOC, 1, 0)
  CE.EER4     <- ifelse(NB.EER4 > NB.SOC, 1,0)
  prob.CE.AllAI    <- mean(CE.AllAI)
  prob.CE.EER4      <- mean(CE.EER4)

  CEAC[i,] <- c(prob.CE.AllAI,prob.CE.EER4)
}

ICERtb1$IQALY <- ICERtb1$TQALY -ICERtb$TQALY
ICERtb2$IQALY <- ICERtb2$TQALY -ICERtb$TQALY
ICERtb1$Icost <- ICERtb1$Tcost -ICERtb$Tcost
ICERtb2$Icost <- ICERtb2$Tcost -ICERtb$Tcost
ICERtb1$IQALY <- rnorm(Nsim,0.04,0.1492)
ICERtb1$Icost <- rnorm(Nsim,69.4,300)
ICERplot2 <- rbind(ICERtb1,ICERtb2)
```