Cost-Effectiveness of Organized Screening for Breast Cancer in Finland

A cost-utility analysis on the cost-effectiveness of the Finnish breast cancer screening program and its potential expansion to younger and older age groups

Carl F.W. Siegfrids

Student number: 632312 (UiO), 550289 (EUR)

Supervisor: Dr. Lars Asphaug

Department of Health Management and Health Economics, University of Oslo

Turku, Finland

June 30, 2022

“DECLARATION IN LIEU OF OATH

I hereby declare, under oath, that this master thesis has been my independent work and has not been aided with any prohibited means. I declare, to the best of my knowledge and belief, that all passages taken from published and unpublished sources or documents have been reproduced whether as original, slightly changed or in thought, have been mentioned as such at the corresponding places of the thesis, by citation, where the extent of the original quotes is indicated.

The paper has not been submitted for evaluation to another examination authority or has been published in this form or another.”

SIGNATURE	
June 29, 2022  date and signature of student	June 29, 2022 Lars Asphaug date and signature of supervisor

Abstract

The latest expansion of the Finnish Breast Cancer Screening Program (FBCSP) from covering women aged 50–59 to invite all women aged 50–69 to biennial screening, has not retrospectively been evaluated as to its cost-effectiveness. The debate on further expanding the FBCSP to younger and older age groups divides decision makers and the medical community alike. This study seeks to shed new light on the cost-effectiveness of the current FBCSP screening strategy with regard to age groups covered by it, as well as to inform the controversy surrounding the optimal screening strategy. A decision-analytic model was developed to perform a Cost-Utility Analysis (CUA), assessing the cost-effectiveness of the current FBCSP screening strategy compared to no screening, as well as that of expanding the program to all women aged 40–74 compared to the current FBCSP strategy, expressed in costs per incremental Quality-Adjusted Life Year (QALY). Applying a stage-shift approach, screening strategies are superimposed on a natural history model reflecting a scenario without screening, allowing for interruption of the natural progression of breast cancer through early detection of malignant tumors. Compared to no screening, the incremental cost-effectiveness ratio (ICER) of the current FBCSP strategy was estimated to €18 584 per QALY gained. The cost-effectiveness of the more extensive screening strategy was estimated to €21 580 per QALY gained, compared to the current FBCSP strategy. While mammography screening is likely to have health benefits to mortality and morbidity associated with breast cancer, the true effectiveness and cost-effectiveness of breast cancer screening through early detection is unclear. Further research to resolve uncertainty around the accuracy of mammography screening and to inform the controversy surrounding overdiagnosis is called for, in support of decision making regarding the optimal strategy for the FBCSP.

Table of Contents

List of Tables and Figures

Tables

Figures

Introduction and Background

Burden of Disease

As of 2020, breast cancer is the primary cause of female cancer incidence, with over 2.3 million new cases confirmed globally in 2020 (WHO, 2021). In Finland, 4 885 women were diagnosed with breast cancer in 2020, over a third of all registered forms of cancer among the female population (Pitkäniemi et al., 2022). It was also the most common cause of cancer mortality that year with a total of 968 confirmed deaths. The largest share of around a fifth of total cancer costs in Finland in 2014, approximately €927 million (of which ca. 80% were treatment costs alone), were related to breast cancer (Torkki et al., 2018). Mean annual treatment costs per breast cancer patient were estimated to €28 700 in a retrospective 10 year follow-up study (Lehtinen et al., 2019).

Progression and Classification of Breast Cancer

Breast cancer, or breast carcinoma, is defined as the uncontrolled growth and spread of malignant tumour cells emanating from different parts of the breast (Cancer Society of Finland, n.d.; Vehmanen, 2020). Most tumours, around 70%–80% of all diagnosed breast cancers, are referred to as ductal carcinomas and originate from the milk ducts. 10%–15% are lobular, originating from the breast's milk-producing glands called lobules, while less common types of breast cancers include tubular, medullary, papillary, inflammatory, mucinous, cribriform and Paget's disease. The risk of breast cancer increases with age, with over half of all newly diagnosed cases being discovered in women over 60 (Vehmanen, 2020). Only about 10% of all breast cancers are diagnosed in the 25–49 age group and, although uncommon, a few women under the age of 25 are diagnosed each year. Lifestyle choices cause roughly a third of all breast cancer cases, obesity and heavy alcohol consumption being strongly associated with an increased

risk of breast cancer. Genetic factors explain around 5%–10% of all breast cancers, mainly due to BRCA1 and BRCA2 gene mutations. Furthermore, hormonal factors such as early menstruation, late menopause, having your firstborn after the age of 30 or none at all, are also linked to a higher risk of breast cancer onset.

The Finnish Cancer Registry (FCR) separates between invasive carcinoma and carcinoma in situ, the latter often described as a precursor to invasive breast cancer (FCR, n.d.-b). Invasive breast cancer is further divided into localized breast cancer (LBC) and non-localized breast cancer (NLBC) based on tumor spread into surrounding tissue. While LBC is confined to the breast, NLBC has "metastasized to regional lymph nodes or further, grew into neighboring tissue or were known to be spread but not known how far" (Heinävaara, Sarkeala & Anttila, 2014).The stage classification used by the FCR is unique and cannot be translated to the internationally recognized Tumor, Node, Metastasis (TNM) classification standard.

Screening as a Health Intervention

The purpose of screening is to identify and separate people with high risk of having a disease from those with a lower risk of having the disease, allowing for early intervention to prevent the disease from advancing (Wilson & Jungner, 1968). The nature of screening is investigative rather than diagnostic, as to the prevalence in an asymptomatic population. A screening test neither confirms nor falsifies the prevalence of disease, however, it can inform whether further assessment is warranted to confirm or rule out the disease. The rationale of cancer screening builds on the belief that cancer advances along a linear path consisting of increasingly progressive abnormalities (Croswell, Ransohoff, & Kramer, 2010; Holland & Stewart, 2005, pp. 1-4). Detection of these abnormalities in an early stage, where pathological change has occurred but has not yet been detected (i.e., due to clinical symptoms or opportunistic

discovery), would break the chain of the natural disease progression, which would reduce cancer mortality due to the favorable survival prospect of early-stage cancer.

The Finnish Breast Cancer Screening Program

Screening emerged from the wider routine of periodic health examination, which became an established general medical practice in the 1920s (Croswell et al., 2010; Smith, Duffy & Tabar, 2012). In the decades after, major health campaigns, mainly by what is today known as the American Cancer Society, raised awareness and promoted early detection of cancer among women through pap smear and breast self-examination. Since the 1970s, clinical evidence from numerous randomized controlled trials (RCTs) of organized mammography screening emerged in support of its reductive effect on breast cancer mortality. Based on these findings, Finland was the first country to implement population-wide breast cancer screening for women aged 50–59 in 1987 (Hakama, Pukkala, Heikkila, & Kallio, 1997).

Today, all Finnish municipalities are required by the Ministry of Social Affairs and Health (339/2011) to offer breast cancer screening to at least all female citizens aged 50–69 by intervals of at least 26 months, being free to expand coverage to younger and older age groups as well. Since the launch of the FBCSP, however, the city of Turku in the Southwest Hospital District of Finland (VSSHP) is the only municipality to have seized this opportunity, by inviting women aged aged 40–84 to mammography screening (Parvinen, 2014). In 2009, the municipality decided to limit the coverage of its screening program in line with the rest of the country, which had gradually started expanding screening to women aged 60–69.

Effectiveness and Cost-Effectiveness of Breast Cancer Screening

Compared to a situation with no screening, the FBCSP is estimated to have decreased breast cancer related mortality among women by 33 percent during the time period 1992–2011

(Heinävaara, Sarkeala, & Anttila, 2016). The experimental screening program of Turku has been estimated to have decreased overall mortality by 36% (Parvinen et al., 2006). The average cost per screened patient in Finland under the current screening program is around 30 euros, which translates to approximately €10 million a year on a national level (FCR, n.d.-c).

While the effect of mammography screening on mortality through early detection is generally accepted in the literature (Broeders et al., 2018; Loberg, Lousdal, Bretthauer, & Kalager, 2015), mass-screening for breast cancer is highly contentious as to its true effectiveness. As breast cancer treatment has improved and awareness increased since the early RCT studies on screening and the launch of the FBCSP, the critical question is whether the advances made in treatment of the disease have eroded the effect of early detection on breast cancer survival (Jatoi, 2011). Furthermore, there are adverse effects to screening on patients' health-related quality of life (HRQoL) from false-positive tests and overdiagnosis, the latter representing cancers which, in the absence of screening, never would have emerged over their lifetime (Duffy et al., 2008).

Assuming that screening is still effective, despite its side effects and the improvements made in breast cancer treatment, there is also substantial variation in screening strategies (i.e., the combination of age groups included in the program and the frequency by which screening is offered) recommended in international organizations' guidelines and applied in screening programs for breast cancer (Esserman, Study, & Athena, 2017; Jayasekera & Mandelblatt, 2020; Schunemann et al., 2020).

The decision on which screening strategy to apply and whether to provide a screening program in the first place is made not based on its effectiveness or, as it is uncertain, its expected impact on survival and HRQoL alone (Goldie, 2003). Because screening programs are publicly

funded health interventions, they are subject to spending budgets and, therefore, need to prove sustainability and affordability as well, often expressed in terms of cost-effectiveness.

A screening strategy is said to be cost-effective when expected to produce more health gains in terms of improving either or both the length and quality of life to the population affected by it, either to lower, equal or acceptably higher costs compared to another strategy, or to the absence of screening (Drummond, Sculpher, Claxton, Stoddart, & Torrance, 2015, pp. 5-7). What constitutes an acceptable cost depends on the preferences of the population affected by the intervention which, in the case of publicly funded health interventions, often is society as a whole (Drummond et al., 2015, p. 25).

The estimated cost-effectiveness of organized breast cancer screening varies considerably in the literature, both between different and similar strategies. A systematic literature review on the cost-effectiveness of various breast cancer screening strategies by Jayasekera and Mandelblatt et al. (2020) found an average increase in health gains per person of 0.033 QALYs. As noted in the review, comparison between costs per person for the increase in health gains due to screening requires similarity between the studies under evaluation regarding perspectives, units of health gains and population settings, to mention a few. A literature review conducted in the chapter on the theoretical framework of this analysis, including studies using the societal perspective and QALYs as units of health gain, found reported estimates between €21 500 and €85 000 per QALY gained from different screening strategies compared to no screening.

Relevance of Study

In 2022 the Finnish population-wide cervix cancer screening program will expand from covering the age groups 35–60 to include all women aged 30–65, while a potential expansion of the FBSCP will need to wait for sufficient support by scientific evidence (Lehtinen, 2021). The

cost-effectiveness of the FBCSP has been assessed twice since its implementation, both in its original form covering women aged 50–59 only (Leivo et al., 1999) and for the potential expansion to include women aged 60–69 before its introduction, by the Finnish Office for Health Care Technology Assessment (FinOHTA, 2001). Neither of the studies considered the impact on HRQoL and no cost-effectiveness analysis has been conducted on the potential expansion of the FBCSP to include other age groups as well.

The evidence on the effectiveness of screening women under 50 is limited (Lauby-Secretan et al., 2015), some advocating against it as women under 50 generally have higher breast density, which negatively affects the accuracy of mammography screening (van den Ende, Oordt-Speets, Vroling, & van Agt, 2017). However, a recent study by Duffy et al. (2020) suggests potential benefits of screening for this age group and over half of Finnish women under the age of 50 already undertake opportunistic screening in private care (Holmberg, 2017). The European Commission Initiatives on Breast and Colorectal Cancer (ECIBC, 2022) recommends biennial or triennial screening for women aged 45–59, as well as triennial screening for women aged 70–74, while the Swedish breast cancer screening program has invited all women aged 40–74 to biennial screening for over 30 years. Expanding the FBCSP to women aged 40–74 has engaged the Finnish political sphere and medical community alike, both sides calling for evaluation of its potential harms and benefits.

Objective and Research Questions of the Study

To the background of the uncertainty regarding the effectiveness and cost-effectiveness of organized breast cancer screening, the objective of this study is to evaluate the cost-effectiveness of the current FBCSP strategy. Furthermore, it seeks to inform the controversy surrounding the optimal screening strategy, by comparing biennial screening of women aged 40–

74 to current practice under the FBCSP. Using the PICO criteria (Aslam & Emmanuel, 2010) the research question is broken down into four segments: Population, Intervention, Comparator and Outcomes. The PICO criteria of this analysis are:

- Population: Finnish women aged 40 and over
- Interventions: Biennial screening in ages 50–69, biennial screening in ages 40–74
- Comparators: No screening, biennial screening in ages 50–69
- Outcomes: Incremental cost-effectiveness ratio (EUR per QALY gained)

Based on the criteria and the objective of this study, the research questions are:

1. Is the FBCSP's current strategy (biennial screening of women aged 50–69) cost-effective compared to no screening?
2. Is an expansion of the FBCSP strategy to invite all women aged 40–74 to biennial screening cost-effective compared to the current screening strategy?

Theoretical Framework

Accuracy of Mammography Screening

The nature of screening as a non-diagnostic intervention reflects the fact that the accuracy of the screening test is never 100% correct in identifying the prevalence of a disease. Although double reading mammography is considered the gold standard instrument in screening for breast cancer (Weber et al., 2015), it is inevitably subject to human errors of radiologists in assessing the results of the mammogram, as well as technological limitations of X-ray imaging (Ekpo, Alakhras, & Brennan, 2018). Importantly, breast density and the speed of tumor growth, both of which are highly correlated with age, have a major impact on the ability of mammography screening to correctly identify malignant tumors (Saarenmaa et al., 2001).

Due to these limitations, a screening test can result in four different outcomes (Parikh, Mathai, Parikh, Sekhar, & Thomas, 2008; Wilson & Jungner, 1968):

1) Cancer is correctly identified and confirmed through diagnostic follow-up examinations.
2) Cancer is incorrectly identified and ruled out through diagnostic follow-up examinations.
3) Cancer is correctly ruled out.
4) Cancer is incorrectly ruled out.

The first outcome represents a true-positive test result, when a de facto malignant cancer tumor is identified and confirmed, resulting in the cancer patient being appropriately directed toward necessary medical attention. The ability of mammography screening to correctly identify subjects with malignant breast cancer is determined by its sensitivity, the inverse of which represents the probability of the fourth outcome, where a de facto malignant tumor is overlooked or incorrectly interpreted as benign, resulting in the cancer patient being rejected necessary

medical attention (i.e., a false-negative test result). The second outcome represents a false-positive test result, the probability of which is given by the inverse of the specificity (or true-positive value) of the screening test, which determines the screening test's ability to correctly rule out the prevalence of malignant breast cancer (i.e., a true-negative test result).

Biases of Breast Cancer Screening

Despite the internationally widespread implementation of organized population-wide screening programs for some major cancer groups, including breast, cervical and prostate cancer, they remain highly contested as to their true effectiveness (Smith, Mettlin & Eyre, 2003). Generally, the lack of consensus regarding the effectiveness of cancer screening can be attributed to a number of biases present in evaluation of screening programs, including lead-time bias, length bias and overdiagnosis.

Lead-time bias arises from the artificial improvement in breast cancer patient survival created by lead time (i.e., the reduction in time between cancer onset and time of clinical diagnosis) when lead time is counted towards survival and the time of death remains unchanged (Duffy et al., 2008; Smith, Mettlin & Eyre, 2003). In other words, lead-time bias occurs when screening only advances diagnosis to an earlier point in time, with any survival benefit of screening being a direct result of lead time rather than "breaking the chain" of the natural history of disease through early detection.

Length bias occurs due to variation in the speed of growth between tumors. Slower growing tumors are susceptible for detection through screening for a longer period of time than faster growing tumors and, thus, are more likely to be detected by a screening test (Duffy et al., 2008; Smith, Mettlin & Eyre, 2003). Generally, slower growing tumors also have a more

favorable survival prospect compared to faster growing tumors, which may overestimate the survival of screen-detected breast cancers compared to that of clinically detected cancers.

Overdiagnosis, which can be defined as an extreme case of length bias, occurs when a cancer that would never have displayed clinical symptoms in one's lifetime is detected by screening (Duffy et al., 2008; Smith, Mettlin & Eyre, 2003). Because all cancers which are discovered also are treated, over diagnosed cancers are said to be overtreated.

Health Economic Evaluation

At the core of any economic evaluation lies the reality of scarce resources and the need to make choices between competing alternatives for deployment of those resources. The purpose of health economic evaluation is, on the one hand, to inform decision-making by providing a framework for using clinical evidence in an optimal way through comparison of relevant alternatives' future costs and consequences (Drummond et al., 2015, pp. 2-4; Edlin, McCabe, Hulme, Hall, & Wright, 2015, p. 2). On the other hand, as decisions in a publicly funded healthcare system are taken on people in a society who bare some or all costs but realize only some or none of the benefits, health economic evaluation seeks to identify and explicate the social values serving as criteria guiding those decisions. The criteria aim to answer questions about how to maximize the health of the individual patient and the rest of the population, as well as what constitutes an acceptable cost for improved population health.

Health economic evaluation is strictly of comparative nature, requiring a suitable approach for identification, measurement and valuation of future expenditures and benefits associated with competing policies, services or interventions under comparison, as well as for taking the uncertainty surrounding these exercises into account (Briggs, Sculpher, & Claxton, 2006, pp. 1-2; Drummond et al., 2015, p. 4). Benefits associated with the alternative under

comparison are mainly concerned with individual health benefits (e.g., survival benefits of a particular health intervention), however, may also include other consequences of value such as reassurance in times of uncertainty regarding your own or loved ones' health, or societal gains in terms of productivity. The costs consist not only of the health care system's resources to be allocated in favor of the competing alternatives, but also patients' and their families' time.

While costs, in terms of identification, measurement and valuation of resources used on a health intervention are similar across health economic evaluations, the consequences they bring about may differ as to their very nature, depending on the type of evaluation applied (Drummond et al., 2015, p. 5; Edlin et al., 2015, p. 3). There are three major ways of conducting full health economic evaluations (i.e., evaluations including assessment of both costs and consequences) including cost-effectiveness, cost-utility and cost-benefit analysis, unique as to the way they identify, measure and value relevant consequences of the health interventions considered.

Cost-effectiveness analysis (CEA) is commonly used to describe both pure cost-effectiveness analyses as well as cost-utility analyses, the latter sometimes seen as a special case of CEA (Edlin et al., 2015, p. 3). CEA uses natural units to measure the benefit of health interventions, such as life-years gained (LYG) or number of cancers detected, aiming to either maximize health benefits for a given cost (e.g., a healthcare budget) or minimize the costs for a given health benefit (Edlin et al., 2015, pp. 3-4). Because CEA only considers a single outcome measure of effect, which needs to be common to the alternatives under comparison, only interventions with the same objective can be compared (Drummond et al., 2015, pp. 5-7; Edlin et al., 2015, pp. 3-5). A cancer screening program, for example, cannot be compared to appendectomy as cancer deaths averted cannot be compared to burst appendices averted.

CUA, however, makes comparison between health interventions with different objectives across different disease areas possible, as it adopts a generic measure of health benefits taking into account not only interventions' life-prolonging benefits, but also the impact they have on the quality of life (QoL) of patients (Drummond et al., 2015, pp. 8-9; Edlin et al., 2015, pp. 5-6). This measure is known as the QALY, which captures health gains from an intervention in terms of both reduced mortality and morbidity. As disutilites of health interventions are not taken into account by CEA, CUA is often preferred when comparing interventions with consequences that impact the QoL of the patients or their loved ones.

Health state values rely on techniques to measure individuals' preferences regarding health outcomes, the most common of which include scaling and choice methods such as Visual Analog Scale (VAS), Standard Gamble (SG) and Time Trade-Off (TTO) (Drummond et al., 2015, pp. 133-137). All methods rely on presenting hypothetical situations to individuals asked to value changes to their health, which would happen with certainty or by a stated probability.

The cost-effectiveness of an intervention is often expressed in terms of incremental costs per unit of effect over the comparator, in a so-called incremental cost-effectiveness ratio (ICER):

$$ICER = \frac{C_1 - C_0}{E_1 - E_0}$$

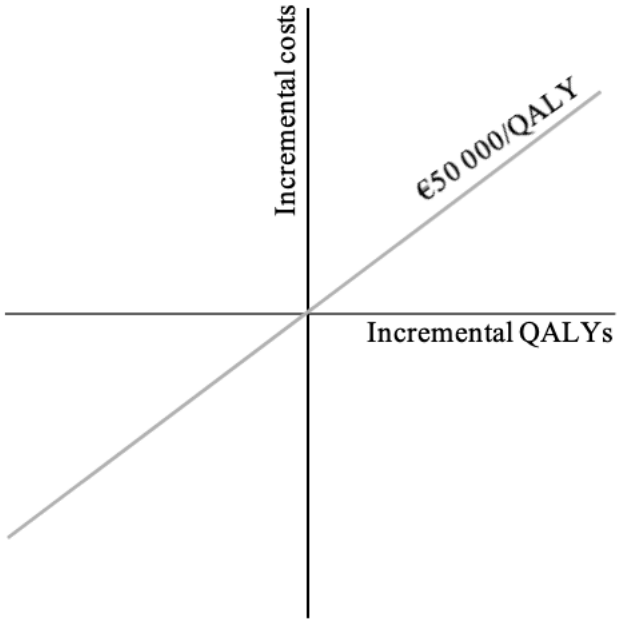
The incremental cost of the intervention per unit of effect is then compared to a cost-effectiveness threshold, representing the societal willingness-to-pay (WTP) per incremental LYG or QALY (Briggs et al., 2006, pp. 79-82). Another way of looking at the WTP-threshold is in terms of the acceptable opportunity cost for investing into a health intervention. A threshold of €50 000 per QALY gained, for example, implies that the society is willing to accept one QALY to be displaced elsewhere in the healthcare sector for every €50 000 spent on a health

intervention, which produces at least one QALY per €50 000 invested into it. In other words, it is not enough for an intervention to have a positive ICER (either due to lower costs or higher benefits than the comparator, or both) to be cost-effective, rather, the ICER must also be lower or equal to the WTP-threshold.

The relationship between the ICER and the WTP-threshold is best explained in a cost-effectiveness plane (CE-plane), as illustrated in Figure 1, where the WTP-threshold is represented by the diagonal line running through the origin. Whenever the ICER is in the north-east quadrant of the CE-plane, to be considered cost-effective, it also needs to be below the threshold diagonal (or above the threshold diagonal in the south-west quadrant).

Figure 1

Illustration of a cost-effectiveness plane.



Due to its influence on the ICER, the choice of comparator is critical for the reliability of an intervention's estimated cost-effectiveness (Drummond et al., 2015, p. 24). Usually, the

comparator is standard of care or current practice at the time of evaluation, however, if there is doubt regarding the cost-effectiveness of the standard of care itself, it might not be a suitable comparator. Importantly, in the face of a large number of potentially relevant alternatives to consider, health economic evaluation seeks to minimize the possibility of overlooking appropriate alternatives from the analysis. The absence of any intervention may also be an appropriate comparator, such as in the case of cancer screening.

Cost-benefit analysis (CBA) is another, although rarely used, way to conduct full health economic evaluations, distinguishable from CEA and CUA in that it values benefits in monetary terms (Edlin et al., 2015, p. 3). The theoretical foundation of CBA is the idea that social welfare can be maximized by investing resources into interventions with greater monetary value of the benefits they produce compared to the incremental costs of providing them. Valuation of benefits usually relies on methods of eliciting individuals' WTP for the expected benefits of health interventions in hypothetical situations, which is difficult to apply at scale to represent a population's valuation (Briggs et al., 2006, p. 3).

The perspective chosen for the economic evaluation, which depends on the decision-makers within the healthcare system intended to be informed by the study, guides the range of costs and consequences to be included in the analysis (Drummond et al., 2015, pp. 24-25). A healthcare payer perspective, for example, is limited to costs incurred by the healthcare payer (e.g., the NHS in the UK or the Social Insurance Institution in Finland) while a healthcare sector perspective is extended to include all healthcare costs regardless of who pays for them. Importantly, the healthcare sector perspective accounts for patient out-of-pocket expenses. A limited societal perspective includes costs incurred outside of the healthcare sector (e.g., patient

and informal caregiver time productivity lost), however, unlike the full societal perspective, it excludes consequences affecting other than the healthcare sector (Kim et al., 2020).

Uncertainty and Decision Modelling

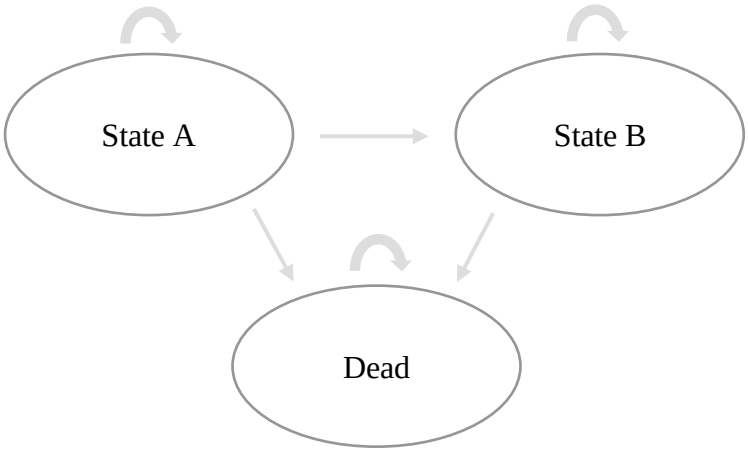
The need for a decision model. RCTs are essential in generating evidence for evaluation of health interventions, policies and programs. While being the gold standard source of evidence, they are inappropriate to serve as the only source for resource use and health benefits in health economic evaluations, unless they capture all relevant outcomes over an appropriate time horizon for the interventions being evaluated (Briggs et al., 2006, pp. 6-8; Drummond et al., 2015, p. 311). Usually, however, RCTs suffer limitations related to short follow-up times and few and/or irrelevant number of comparisons. Furthermore, most RCTs in breast cancer screening were conducted between the early 1960s and 1980s, over 40 years ago from today (de Koning, 2003). Therefore, changes and improvements in breast cancer treatment have rendered data from the pre-screening era and evidence from the RCTs unsuitable for directly inferring the impact of screening on health benefits associated with breast cancer survival (Glasziou & Houssami, 2011). Uncertainty surrounding the evidence based on which health economic evaluations are conducted often makes decision-making reliant on decision modelling, capable of addressing a specific health care system's decision problem at a given point in time by applying a full range of evidence combined from different sources (Briggs et al., 2006, pp. 6-8; Drummond et al., 2015, p. 311). Kielhorn and von der Schulenburg (2000) define decision modelling as "a systematic approach to decision making under conditions of uncertainty, in which the probability of each possible event, along with the consequences of those events, is explicitly stated".

Decision models. Models used in health economic evaluation can broadly be categorized into cohort and patient-level models, depending on whether the experience of the average patient from a homogenous population, or that of the individual patient with unique characteristics is considered (Briggs et al., 2006, pp. 23-29; Drummond et al., 2015, pp. 328-331). The simplest form of a decision model is the decision tree, in which patients move along pathways of branches, representing event possibilities. The branches emanate from chance nodes, representing uncertainty as to the probability of a patient to experience certain events and their associated costs and outcomes. While intuitive and easy to implement, major limitations of decision trees include their lack of explicity with regard to elapse of time and complexity with long time horizons.

State-transition models (STMs), based on mutually exclusive and collectively exhaustive health states are flexible and can handle multiple event possibilities (Drummond et al., 2015, pp. 331-335; Edlin et al., 2015, pp. 79-80). The most widely used cohort STMs is the Markov model, which is illustrated in Figure 2. Transition possibilities between health states “A”, “B” and “Dead”, represented by the straight arrows, occur within a specified cycle length with a certain probability. The probability to stay in a health state, represented by the curved arrows above each state, is simply the inverse of the sum of transition probabilities from the state. Each health state is associated with state values, such as QALYs and costs, the expected value of which are calculated by summing up the average proportion of the population residing in that health state each cycle over the time horizon of the model (i.e., the total number of cycles over which outcomes are predicted). As Markov models assume are cohort models, patients move through health states according to average transition probabilities.

Figure 2

Illustration of the structure of a state-transition Markov model.



A drawback of Markov models is that, once having transitioned to a new health state, the model has no memory of the timing of the transition or the health state from which the patient transitioned (Drummond et al., 2015, pp. 336-337). This feature is called the Markov assumption, and essentially means that probabilities of future events cannot depend on experience, which is often the case with cancer treatment. Patient-level or micro-simulation is another type of STM adopted by more sophisticated breast cancer screening models such as the MISCAN series. The main advantage of microsimulation over cohort models is that patients can be tracked, and costs and outcomes estimated based on unique accumulated disease histories. They are, however, more demanding of specific data to inform event probabilities conditional upon patient characteristics, as well as of computational power.

Assessing uncertainty. Uncertainty can originate from various sources and contribute to decision uncertainty in economic evaluation, the ultimate goal of which is to inform decision-making. Because economic evaluation is concerned with estimating future expected benefits and

costs of alternative courses of action, uncertainty is inherent in decision-making informed by the evaluation in so far as the future is uncertain (Briggs et al., 2006, pp. 77-78; Edlin et al., 2015, pp. 59-60). Understanding sources of uncertainty and their impact on the outcomes of health economic evaluation and decision-making is important in order to minimize the risk of making a suboptimal decision. Where there is value in delaying a decision and collect more information, making a suboptimal decision can be costly, especially where decision reversal is problematic, as is often the case with policy changes in the healthcare sector.

Broadly speaking, uncertainty in health economic modelling can be defined as first- and second order uncertainty, dealing with uncertainty arising from the patient population and uncertainty attributable to the decision-analytic model, respectively (Briggs et al., 2006, pp. 82-83; Drummond et al., 2015, pp. 392-393). For cohort-level decision-analytic modelling, dealing with average expected costs and benefits in a population, second-order uncertainty is the main concern. Second-order uncertainty can be split into parameter uncertainty, on the one hand, representing uncertainty introduced by the fact that input parameters (i.e., transition probabilities, utilities and costs etc.) are mere estimations, carrying a degree of uncertainty as to their true value. Structural uncertainty, on the other hand, relates to assumptions applied in the conceptualization and construction of the model.

A major advantage of decision-analytic models is their ability to handle uncertainty. By means of sensitivity analysis, the impact on model outputs of allowing parameter input values to vary can be assessed (Drummond et al., 2015, pp. 393-398; Edlin et al., 2015, pp. 65-66). Deterministic sensitivity analysis (DSA) involves analyzing how the ICER changes when varying single and multiple parameters' values, called one-way (OWSA) and multi-way (MWSA) sensitivity analysis, holding all other parameters constant. While not informative for

evaluating decision uncertainty, OWSA and MWSA are straight-forward to implement and allow for easy assessment of individual parameters' impact on the sensitivity of model outcomes.

DSA also fails to account for the fact that all parameters, simultaneously, are uncertain (Drummond et al., 2015, p. 395). A method that resolves this limitation, increasingly regarded as the standard approach in cost-effectiveness modelling to assess the level of decision-uncertainty, is probabilistic sensitivity analysis (PSA). PSA is based on the practice of repeatedly drawing random values from the input parameters' distributions (i.e., Monte Carlo simulation) and recording model outputs in terms of the ICER associated with each iteration (Drummond et al., 2015, p. 399 & 403; Edlin et al., 2015, pp. 68-74). By plotting each simulated ICER on a CE-plane, the uncertainty as to the true value of an alternative's ICER is visualized by the dispersion of the data points on the CE-plane. A clustered set of data points represents less uncertainty than a more dispersed one. While impossible to compare ICERs in the north-east and south-west quadrants by looking at the CE-plane, it can be resolved by converting ICERs into net benefits.

Net benefits can be expressed in terms of net health benefits (NHB) or net monetary benefits (NMB), representing the expected value of the incremental benefit of an intervention after netting out its incremental costs. Using Monte Carlo simulation and recording the number of times each competing alternative offers the highest net benefit, their respective probability to be cost-effective can be calculated as the proportion of times they record the highest net benefit (Drummond et al., 2015, pp. 405-406; Edlin et al., 2015, p. 74). When calculating these probabilities for a range of different cost-effectiveness thresholds, a cost-effectiveness acceptability curve (CEAC) can be plotted, visualizing which intervention has the highest probability to be cost-effective at a given threshold. Importantly, the intervention with the highest probability to be cost-effective at a given threshold is not necessarily the optimal choice

in terms of maximizing the expected net benefit. The probability that the intervention with the highest expected net benefit expected to be cost-effective at a given threshold, actually is cost-effective, is given by the cost-effectiveness acceptability frontier (CEAF).

While sensitivity analysis is a good way to assess the scale of parameter uncertainty and the probability of an intervention's cost-effectiveness given this uncertainty, it cannot identify the drivers of decision uncertainty (Drummond et al., 2015, pp. 411-415; Fenwick et al., 2020). Value-of-information (VoI) analysis provides an assessment of the value of research in terms of how the evidence it generates can reduce decision uncertainty and, therein, maximize the payoffs associated with it. Expected Value of Perfect Information (EVPI) represents the value of eliminating all uncertainty, while Expected Value of Partial Perfect Information (EVPPI) represents the value of perfect information for specific parameters. On the individual level, EVPI and EVPPI are calculated as the difference between the expected value of a decision (i.e., the expected net benefit) if complete information on all or some parameters were available, respectively, compared to the expected value of a decision based on existing evidence.

Modelling Breast Cancer Screening for Health Economic Evaluation

According to a systematic review of cost-effectiveness models for general population breast cancer screening by Schiller-Fruehwirth et al. (2017), state-transition models were by far the most frequently used modelling approach, used in 27 of 35 studies included in the review. Around half of the STMs included in the review were individual-level microsimulation models, the other half being Markov models adopting a cohort approach.

Natural history of disease. The essence of modelling the effectiveness of breast cancer screening using a decision-analytic model is to first model the natural history of breast cancer, defined as the natural progression of breast cancer in the absence of screening (Abrahamsson,

Isheden, Czene, & Humphreys, 2020; International Agency for Research on Cancer, 2016). The disease progression is then allowed to be disrupted by various screening regimes introduced to the natural history model, enabling tumors to be discovered in an earlier stage than they would have been in the absence of screening. This "breaking the chain" of the natural disease progression is driven by the ability of screening to detect tumors before they turn symptomatic and are clinically diagnosed upon examination. Three- and five-state models, illustrated in Figures 3 and 4, are frequently used to model the natural history of breast cancer for the evaluation of screening effectiveness, breaking the disease progress down into disease-free, preclinical and clinical phases (Tan et al., 2013). The five-state model separates between tumors, based on whether or not they have progressed into surrounding tissue, for example.

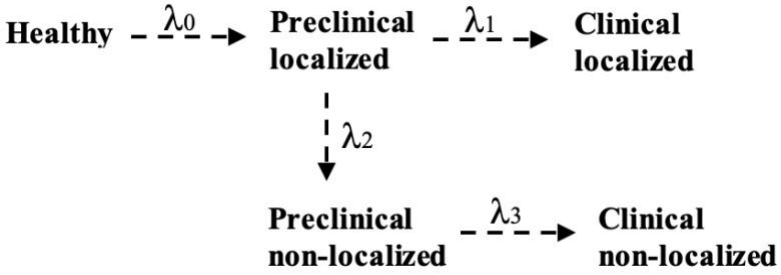
Figure 3

Illustration of a three-state natural history model.

Note. Lambda λ = transition probability between two health states.

Figure 4

Illustration of a five-state natural history model.



Note. Lambda λ = transition probability between two health states.

The preclinical phase includes tumors which have not yet turned symptomatic but are detectable by screening, which allows for evaluating the impact of screening in terms of early detection. The clinical phase, on the other hand, includes tumors that have progressed to a symptomatic stage. These natural history models assume screening to be effective only in so far as it can detect tumors in the preclinical phase (Cheung, Hutton, & Brettschneider, 2017).

In contrast to the clinical phase in which breast cancer tumors are diagnosed upon examination of clinical symptoms, the preclinical phase is unobservable (Cheung et al., 2017). This is because any tumor detected before showing clinical symptoms, either through routine or opportunistic screening, or concurrently with examination or treatment of some other ailment, is immediately treated. As it would be unethical not to treat, the timing of preclinical disease onset (i.e., the underlying incidence rate) and the time from tumor detection until it would show clinical symptoms remain unknown. However, methods to approximate the mean sojourn time (MST), which is the mean duration of the preclinical phase, have been developed to determine the optimal round length of population screening programs. As MST is the inverse of the transition rate from the preclinical to the clinical phase, it is imperative to understand the rate by which tumors exit the preclinical phase in estimating the MST (Cheung et al., 2017; Wu et al., 2010).

According to the systematic review by Schiller-Fruehwirth et al. (2017), the number of stages applied to model invasive breast cancer vary considerably between CEAs of breast cancer screening, as do assumptions regarding the role of Ductal Carcinoma In Situ (DCIS) as a precursor to invasive breast cancer (IBC). The wide range of different applications of DCIS and its modelled relationship with IBC reflect a poorly understood natural history (Heller, Plaunova, & Gao, 2021; Schiller-Fruhwirth et al., 2017). However, excluding DCIS from the analysis could

underestimate both the benefit of screening, because only invasive breast cancer could be detected, as well as overdiagnosis due to screening, under the assumption that not all DCIS lesions progress to IBC (S. Heinävaara, personal communication, 20.01.2022; Schiller-Fruhwirth et al., 2017). The CEAs included in Schiller-Fruehwirth et al. (2017) that modelled DCIS from which progression to IBC was allowed applied different assumptions regarding the proportion of IBC preceded by DCIS, varying between 5% and 40%.

Parameter calibration. Because the preclinical phase is unobservable, parameters needed to model the natural history of breast cancer (e.g., the underlying incidence rate representing transitions into the preclinical phase) often need to be calibrated (Hunink et al., 2014, pp. 344-350; Vanni et al., 2011). Calibration is an estimation process of uncertain or unknown parameters, with the objective to identify parameter values that achieve a good fit in terms of consistency between model predictions and calibration targets (i.e., empirical data such as observed prevalence or mortality). It is also a tool for adjusting all epidemiological parameters to increase model credibility. Calibration is not to be confused with validation, a method to assess the credibility of model predictions and transparency of its informing parameters.

The effect of screening on breast cancer survival. There are two main approaches for modelling the benefit of screening on breast cancer survival; stage-shift modeling and screen-related mortality reduction (Schiller-Fruhwirth et al., 2017). The stage-shift approach relies on discrimination between different stages of breast cancer, according to tumor spread like the five-state natural history model in Figure 4, for example. The effect of screening is captured by a shift in diagnosed breast cancers toward a larger proportion being discovered in an earlier stage, with better survival prospects compared to later stage breast cancer (International Agency for Research on Cancer, 2016, pp. 294-295). The screen-related mortality reduction approach

utilizes published estimates on the reduction in breast cancer specific mortality associated with screening-detected cancer (Schiller-Fruhwirth et al., 2017). By applying mortality reductions to all screening-detected cancers in the model, the survival-benefit of screening compared to that of no screening is estimated.

Existing Literature and Study Contribution

Based on systematic literature reviews by Schiller-Fruehwirth et al. (2017) and Jayasekera and Mandelblatt (2020) as well as literature searches limited to publications after 2019, existing literature on the cost-effectiveness of organized breast cancer screening was assessed (see literature search strategy in Appendix A1). Relevant studies were selected based on the screening strategy and instrument being evaluated as well as the population for which they were conducted, according to how well they matched the strategies evaluated in this analysis (both of which assume mammography as the screening instrument) and resembled Finland's female population in terms of patient characteristics, respectively. Furthermore, studies using other than societal perspectives and QALY's as units of health effects were excluded, as these are preferred according to the Finnish Pharmaceutical Pricing Board (HILA, 2019).

Two studies estimating the cost-effectiveness of the FBCSP were found. One conducted by Leivo et al. (1999) for the screening strategy of the time including women aged 50–59 only, used nation-wide screening data from 1987–1992 to estimate cumulative mortality rates between screening and control groups. From a societal perspective compared to the absence of screening, the study estimated an ICER of USD 18 955 per LYG, ranging between USD 15 502 and USD 40 308 in sensitivity analyses, approximately €20 098 (€16 437– €42 738) in 2017 price levels*. FinOHTA (2001) evaluated the expected cost-effectiveness of expanding the FBCSP strategy to its current form, including women aged 50–69, using an unspecified modelling approach. The

study estimated ICERs between €2 227 and €42 316*. As these studies did not use QALY as the unit for health effects, they were excluded from Table 1 presenting relevant studies.

Table 1

Relevant studies on the cost-effectiveness of organized breast cancer screening.

AUTHORS AND STRATEGY	ICER (EUR/QALY)	PERSPECTIVE	POPULATION
Socialstyrelsen (2019)		Limited societal	Sweden
Biennial 50–69 vs no screening	€22 513*		
Biennial 40–69 vs biennial 50–69	€39 114*		
Biennial 50–74 vs biennial 50–69	€12 225*		
Van Luijt, Heijnsdijk & de Koning (2016)		Societal	Norway
Biennial 50–69 vs no screening	€21 500*		
Mittmann et al. (2015)		Societal	Canada
Biennial 50–69 vs no screening	€80 389*		
Biennial 50–74 vs no screening	€85 111*		
Rim et al. (2019)		Societal	USA
Biennial 40–64 vs no screening	€44 176*		

Note. *All estimates were converted to euros using World Bank (2020) and ECB (n.d.) exchange rates, as well as adjusted to the 2021 price level with Consumer Price Indices from Official Statistics Finland (OSF, 2021).

The study by the Swedish National Board of Health and Welfare (Socialstyrelsen, 2019) estimated relative risk-reductions to model the effect of screening on breast cancer mortality, while van Luijt, Heijnsdijk and de Koning (2017) used a MIcro-simulation SCreening ANalysis (MISCAN) model for the natural history of disease, allowing the natural stage-distribution to be disrupted by screening. The studies by Mittmann et al. (2015) and Rim et al. (2019) both used models of the Cancer Intervention and Surveillance Modeling Network (CISNET), adapted to relevant settings of their studies. The models are based on discrete-event simulation to predict cancer incidence and mortality under various screening strategies.

This study aims to contribute to the existing literature on cost-effectiveness of breast cancer screening by shedding new light on the cost-effectiveness of the FBCSP's current strategy compared to no screening, as well as of expanding it to younger and older age groups compared to the current screening strategy. The cost-effectiveness of the screening program is assessed in terms of cost per QALY gained, using updated evidence synthesized from multiple sources in a decision-analytic model.

Methods and Data

To assess the cost-effectiveness of screening, a Markov state-transition model was built using Microsoft Excel version 16.61. The model was constructed to reflect the natural history of breast cancer in the jurisdiction of the Finnish female population, simulating a hypothetical cohort of 100 000 women aged 40 years at average risk of developing breast cancer. The natural disease progression represents a hypothetical scenario in which screening is absent and breast cancers are diagnosed in the clinical phase only. To evaluate the effect of screening, strategies are superimposed on the natural history model, interrupting the natural disease progression which, in turn, impacts disease-specific survival according to the stage-shift approach.

The cycle length within which an event (i.e., a transition between two health states) can occur should reflect consistency with the clinical problem and the consequences of the intervention (Siebert et al., 2012). Accordingly, the cycle length is set to one year, as participation in screening takes place over the whole year of a screening round. Furthermore, best available evidence on model events, as well as costs and consequences associated with breast cancer, are estimated over the period of a calendar year. In line with national guidelines on economic evaluation, the time horizon over which the model predicts outcomes is set to lifetime (i.e., until the whole simulated cohort is dead either from breast cancer or from other causes) to account for all essential consequences of screening (EUnetHTA, 2015; HILA, 2019).

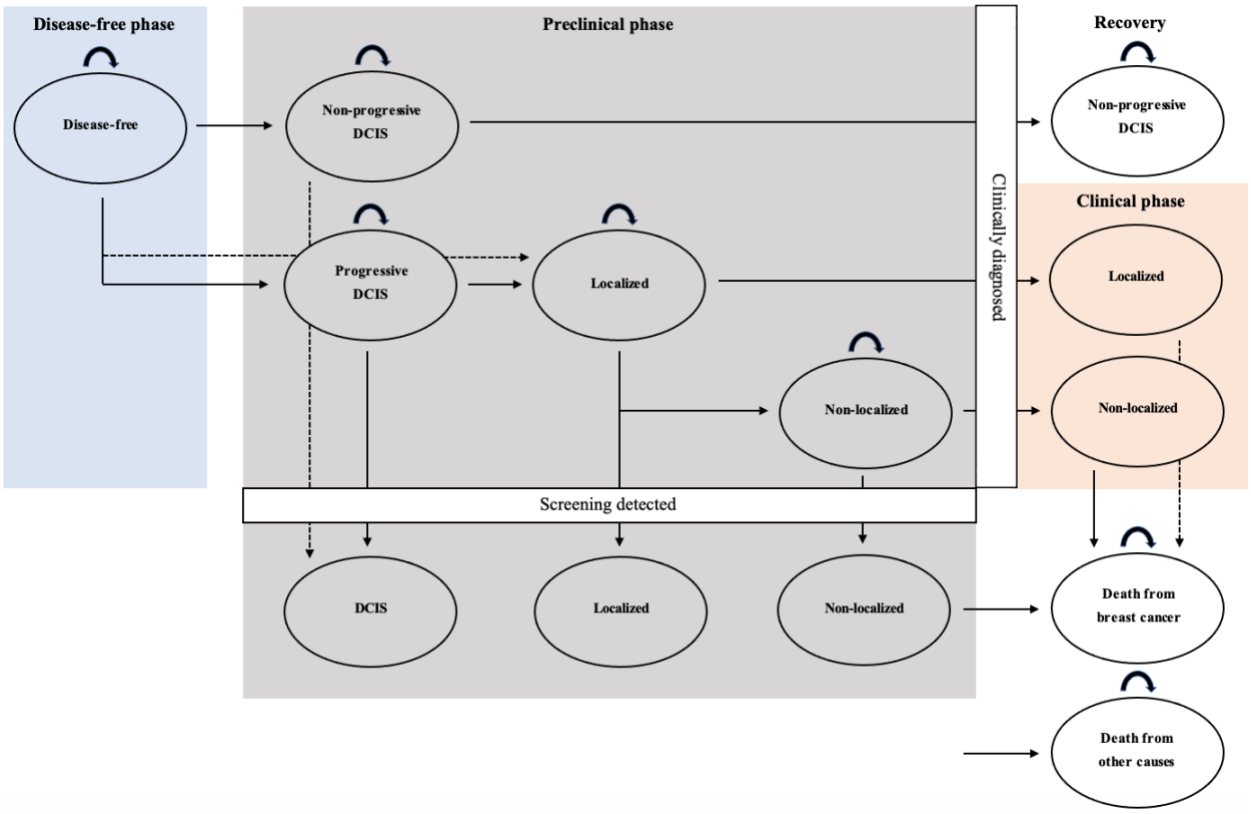
Model Structure

The model built for this analysis, illustrated in Figure 5, is structurally similar to the model used in van Luijt, Heijnsdijk and de Koning (2017). The model is adapted to reflect the disease stage classification system used by the FCR. Apart from invasive breast cancer, the model also includes DCIS as a precursor to IBC. DCIS is split into two separate health states,

progressive DCIS (pDCIS) and non-progressive DCIS (npDCIS), based on the assumption that not all DCIS lesions advance to IBC if left untreated. Accordingly, npDCIS allows for recovery, which is modelled as a separate health state rather than disease regression.

Figure 5

Visual presentation of the model structure.



Note. All straight arrows represent transition possibilities between health states, while curved arrows above health states represent probabilities to stay in the respective health state.

The evidence on disease progression and regression concerning carcinoma in situ is predominantly concerned with Ductal Carcinoma in Situ (DCIS), which is the most common form of in situ lesions, accounting for over 90% of all diagnosed in situ lesions in Finland in 2019 (FCR, n.d.-d). Despite the ambiguity regarding the natural history of DCIS and its role as a

precursor to IBC, in situ lesions are strongly associated with overdiagnosis and excluding them from the model would likely under-estimate the total overdiagnosis of mammography screening and, consequently, over-estimate its true effectiveness (S. Heinävaara, personal communication, 20.01.2022). DCIS rarely presents any symptoms and is predominantly discovered through mammography screening (Ernster & Barclay, 1997; Venkatesh, Oseni, & Bahl, 2021). To this background and due to the structure of the available evidence on DCIS progression to IBC, this model does not allow for preclinical DCIS progression to a clinical phase. The separation between non-progressive and progressive DCIS lesions into two different health states was motivated by the limited availability of estimated transition rates from DCIS to IBC, that were also compatible with the IBC natural history model structure.

The simulated population is assumed to start in the disease-free health state, defined as either absence of breast cancer, presence of progressive cancer lesions that are neither symptomatic nor detectable by screening, or presence of screening-detectable non-progressive LBC. The possibility of having non-progressive LBC is not modelled as a health state such as with DCIS and, thus, the probability of having it is not guided by transition probabilities. The prevalence of non-progressive LBC does not matter for the natural history of breast cancer, rather, is a way of modelling overdiagnosis associated with screening.

According to the model structure, the disease-free population can, in terms of disease progression, only transition to the preclinical phase of npDCIS, pDCIS or LBC within a cycle. In practice, however, because cancer lesions advance at different rates and the cycle length is set to one year, some lesions may pass through one or more health states within a single cycle, unless identified by screening. This is reflected in the natural history parameters used in the model, according to which a certain share of the population transition straight from the disease-free

health state to clinical NLBC, for example. However, this does not mean that some simulated patients break the natural breast cancer disease progression, as outlined in the model structure. Rather, it is the result of STMs only recognizing health states in which simulated patients reside in at the end of each cycle, which necessitates such transition possibilities to reflect the varying rates of progression. Consequently, in the absence of screening, disease-free patients can progress through all preclinical health states related to IBC within the span of a cycle and end up in the clinical phase in the next one, just as patients with preclinical LBC can pass through preclinical NLBC and end up with clinical NLBC within the same amount of time. These transition possibilities do not, however, apply to the proportion of the population going through the pre-invasive health state of pDCIS. In other words, progression from pDCIS through preclinical IBC to clinical IBC cannot occur within a single cycle, the reason of which is the lack of evidence on annual transition probabilities between pDCIS and clinical IBC.

When a screening scenario is introduced to the model, the population in the screening-detectable preclinical phase, within the age-range covered by the screening strategy, may exit the natural history disease progression every other year, if they participate in and their cancer is correctly identified by the screening process. The effect of screening is reflected in the model by denying patients who are correctly identified by the screening process to stay in the same preclinical health state or transition to other health states within that cycle, immediately assigning them to treatment followed by surveillance and, eventually, death.

Primary treatment of breast cancer, predominantly consisting of invasive surgery followed by adjuvant radio- and chemotherapy, is usually initiated within four weeks after diagnosis (Helsinki University Hospital, n.d.; Roine et al., 2019). Therefore, regardless of whether a patient is diagnosed in the clinical phase or through screening, treatment is assumed to

be initiated the same cycle as they are diagnosed in. Death from breast cancer or from other causes are “absorbing” health states, meaning that transition from them to other health states is impossible. The whole simulated population runs a certain age-specific risk of baseline mortality (i.e., mortality from other causes than breast cancer) every cycle, regardless of the health state they reside in, expect for patients diagnosed with metastatic breast cancer, who are assumed to die from breast cancer specific mortality only.

The time between diagnosis and death is called surveillance, consisting of different phases depending on cancer stage. NLBC is modelled as either non-metastatic or metastatic breast cancer, according to Lehtinen et al. (2019) reporting age-specific stage distributions of breast cancers recorded in the Turku University Hospital’s specialized care unit in VSSHP between 2004 and 2014. Following this study, the surveillance phase of the model separates between non-metastatic and metastatic NLBC, the former consisting of lymph node positive lesions and breast cancers of unknown spread.

The distinction is important for modelling survival, as non-metastatic breast cancer patients run certain risks of recurrence depending on age and disease stage, primarily within the first five years after treatment (Rintasyöpä.fi, 2021). According to Lehtinen et al. (2019), based on registry data from women diagnosed with breast cancer in Kuopio University Hospital between 1992 and 2011, the overall probability of local recurrence from diagnosis for non-metastatic breast cancer was estimated to 30%, with higher observed recurrence rates in younger ages and more advanced stages of breast cancer. The patients entering LBC and non-metastatic NLBC surveillance upon diagnosis are separated by the age-and stage specific recurrence rates into two survival functions. A proportion equal to the recurrence rate are assigned to all-cause breast cancer survival, reflecting the fact that this share of patients will at some point experience

local breast cancer recurrence within five years from diagnosis. The remaining patients who will not experience local recurrence are assigned to baseline survival. While the timing of recurrence is unknown, distinction between patients experiencing recurrence and not is necessary to avoid underestimating all-cause breast cancer survival. All-cause survival is a combination of baseline and breast cancer specific survival (BCSS). Full recovery from breast cancer is defined as absence of recurrence for a specific period after treatment, in this model assumed to be five years (10 years in case of recurrence). Because DCIS can only be diagnosed through screening, all DCIS patients are assumed to fully recover and follow baseline survival. Patients diagnosed with metastatic breast cancer follow BCSS only as they are assumed to never recover or, put differently, run a 100% probability of recurrence (Lehtinen et al., 2019). Recurrence rates and NLBC stage distributions applied in the model are presented in Appendix A2.

Model Parameters

Natural history parameters. Movement between the modelled health states associated with the natural history of breast cancer are guided by estimated transition rates and transition probabilities, retrieved from three different studies regarded as best available evidence based on a literature search. Searches were conducted by combinations of search terms relating to the natural history of breast cancer and DCIS, state-transition modelling, as well as the preferred setting of the Finnish female population. Due to observed differences in breast cancer incidence rates and survival on the one hand, and the unique FCR tumor classification system on the other, populating the model with transition probabilities estimated from other than the Finnish female population would be complicated. Furthermore, because aging has a negative impact on the rate of transition between different stages of breast cancer (i.e., the younger the patient, the faster the rate of tumor growth) (Cheung et al., 2017), age-specific transition rates were required. Relevant

studies were singled out according to the population from which the parameters were estimated, in terms of geographical setting and age. Details of the process of searching for parameters and synthetizing relevant evidence are presented in Appendix B.

Transition probabilities were informed by Wu et al. (2010), Duffy et al. (1997) and Yen et al. (2003), as these estimates were most compatible with one another and relevant to the Finnish setting. Wu et al. (2010) estimated natural history parameters for IBC based on non-randomized screening data from the Pirkanmaa Hospital District of Finland, using a five-state model adopting the same tumor stage classification of the FCR. However, as the data used was collected between 1988 and 2000, transition probabilities were estimated only for ages 50–59, the target age group of the district municipalities' screening program at the time. Duffy et al. (1997) estimated IBC transition probabilities for the age groups 40–49, 50–59 and 60–69 using a five-state Markov-chain model based on the Swedish Two County Trial screening data. The Swedish Two County Trial was the first randomized breast cancer screening trial suggesting a reductive effect of mammography on breast cancer mortality. Yen et al. (2003), also based on the Swedish two county trial data, estimated age-specific (i.e., including age groups 40–49, 50–59 and 60–69) natural history transition rates for entry into and exit from non-progressive and progressive DCIS, respectively. The study used a Markov process model fitted to the screening trial data to estimate the transition rates. Following Edlin et al. (2015, p. 83), transition rates (r) were converted to annual probabilities (p) according to:

$$p = 1 - e^{-r}$$

Unknown and uncertain parameter values were calibrated using the seven-step approach of Vanni et al. (2011), consisting of choices regarding; 1) which parameters to be calibrated; 2) what constitute appropriate calibration targets; 3) goodness-of-fit (GOF) measures to be applied;

4) parameter search strategy to be used; 5) convergence criteria; 6) stopping rule and 7) integration of calibration results with model parameters.

Rather than just the unknown parameters, all epidemiological parameters were included in the calibration process to account for variability in the precision of parameter estimates (Vanni et al., 2011). Age-specific pre-screening era (1977–1987) breast cancer incidence in the VSSHP (FCR, n.d.-a) were used as natural history calibration targets, reflecting both the relevant time period of the outcomes to which parameter values are being calibrated against, as well as the relevant population for which decisions are to be made (Hunink et al., 2014, p. 345; Vanni et al., 2011). GOF-measures are used to evaluate the proximity of model predictions to the calibration targets, usually in the form of quantitative distance and likelihood measures (Hunink et al., 2014, p. 345). The straightforward relative distance GOF-measure of minimizing the percentage deviation between the model predictions E_i and calibration targets O_i was used for the natural history parameter calibration, according to:

$$\sum_i \frac{E_i - O_i}{O_i}$$

Parameter search strategies refer to methods by which parameter values that best fit the calibration targets are found (Vanni et al., 2011). Some commonly used search strategies include the methods of grid search, generalized reduced gradient and downhill simplex (also known as the Nelder-Mead method). The most widely used approach, also used for calibration of this model's parameters, is the random search method. Random search operates through sampling random values from distributions assigned to the parameters which are allowed to vary in the calibration process, producing unique combinations of randomly generated parameter values.

Parameter distributions represent the interval from which random values are drawn in the calibration process. The transition rates associated with the natural history of DCIS, except for the probability of recovering from npDCIS, had reported measures of variation (95% CI) from which random values were directly drawn and converted to transition probabilities. The other transition probabilities were assigned appropriate distributions based on the nature of the data informing them. Probability parameters informed by binomial data were fitted with the beta distribution, which is constrained to the 0–1 interval and determined by two parameters: alpha and beta (Briggs et al., 2006, pp. 86-88). The Dirichlet distribution, defined in Briggs et al. (2006, p. 88) as a "multivariate generalization of the beta distribution with parameters equal to the number of categories in the multinomial distribution", was fitted to parameters included in a set of polytomous transitions from a health state.

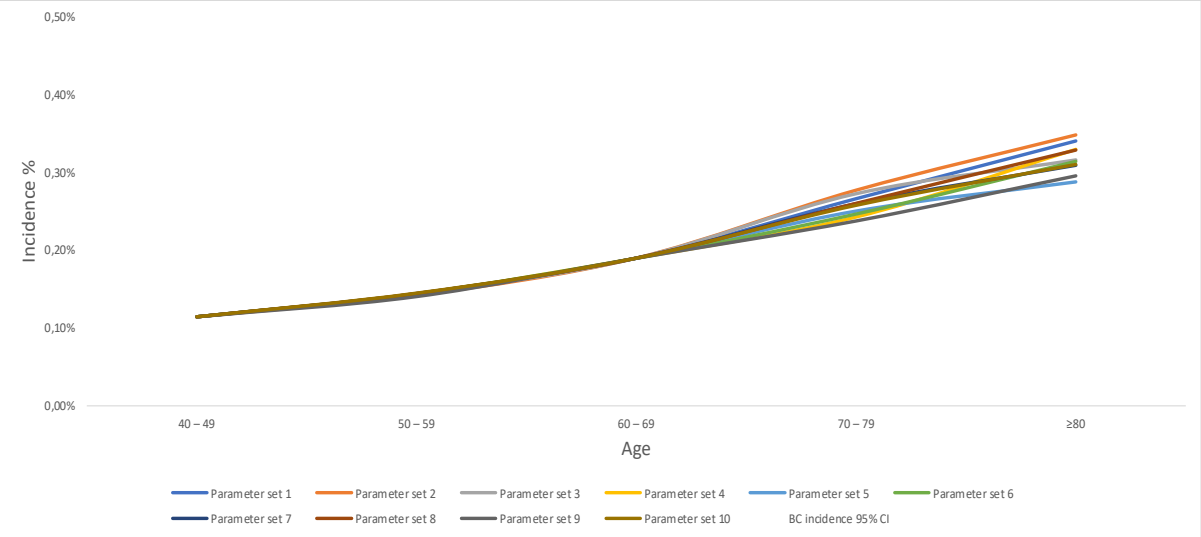
The unknown transition probabilities between pDCIS and preclinical LBC and NLBC, respectively, were assigned a wide interval from which to sample random values, assuming a larger proportion of transitions out of pDCIS end up in preclinical LBC within a given cycle. Because the model applies a lifetime horizon, natural history transition probabilities for ages over 69 are needed. Due to the lack of evidence on estimated natural history parameters for ages 70 and over, parameter distributions for this age group were approximated based on the corresponding calibrated parameters for ages 60–69. Methods for assigning distributions to unknown parameters and an overview of the calibration parameters are presented in Appendix B.

All natural history parameters were calibrated in a stepwise manner according to age in order to fit the age-specific structure of the parameters. Following the approach of Mittman et al. (2015), in which parameters were calibrated against observed incidence rates to model a scenario without screening, each of the parameter age groups presented in Table B1 (Appendix B1) were

calibrated separately from youngest to oldest, against observed age-specific historical incidence rates. The natural history parameters for ages ≥70 were calibrated against incidence rates for both the age group 70–79 and ≥80, respectively. By comparing model incidence rates against the corresponding age-specific calibration targets, the combination of parameter values producing the best fit were identified as the one minimizing the GOF-measure. This constitutes the convergence criteria. The stopping rule of this calibration process is defined as the completion of 1 000 iterations of simulating random parameter values. Finally, the calibration results were integrated with the rest of the model parameters, by replacing the inputs for the calibration process retrieved from the literature with point estimates of the best fit calibrated parameters. When completed for ages 40–49, the process is repeated for ages 50–59 and so on, holding the parameters already calibrated constant. Figure 6 shows the 10 best fit calibrated natural history parameter sets plotted against the incidence rate 95% CI.

Figure 6

Natural history parameter calibration goodness of fit.



Screening-specific parameters. Age-specific invasive breast cancer sensitivity and specificity for the age groups 40–49 and 50–59 were informed by Duffy et al. (1997) and Wu et al. (2010), respectively. For ages 40–49, sensitivity was estimated to 83% (95% CI = 0.76–0.91), whereas for ages 50–59 it was estimated to be 85% (95% CI = 0.75–0.95). Because no measure of variation was estimated for the specificity for ages 40–49, it assumed to be the average of the point estimate for that age group and the lower and upper bound measures of variation for ages 50–59, respectively. Thus, the specificity for ages 40–49 was assumed to be 92%, while for ages 50–59 it was estimated to 99.97% (95% CI = 0.9989–1).

IBC sensitivity for older age groups were assumed to follow a linear increase based on the sensitivity estimated for ages 40–49 and 50–59, giving sensitivities of approximately 86% and 89% for ages 60–69 and 70–74, respectively. IBC specificity for these age groups were calculated as the average of their respective lower and upper bound measures of variation, the lower bound assumed to be equal to the point estimate of the specificity for the younger age group. The upper bound was assumed to be 1, following the estimated upper bound 95% CI for ages 50–59 reported in Wu et al. (2010). The sensitivity for DCIS in all age groups were assumed to be 15% higher than the sensitivity of IBC, following Ernstner et al. (2002), which also found no significant impact of age on DCIS sensitivity. Due to lack of better evidence, specificity for DCIS was assumed to be equal to that of IBC. An overview of the parameters determining the mammography test's accuracy, as well as the method for assigning measures of variation to the estimates for which none were reported, are provided in Appendix B2.

The rate of screening program participation directly impacts the total number of breast cancers detected by screening. In 2019, the overall national participation rate was approximately 82%, with no significant difference between age groups (Anttila et al., 2021). Accordingly,

participation rate was applied to all age groups in the model. Because screening sensitivity is never 100%, even if all simulated patients covered by the screening strategy did participate, some false-negative breast cancers overlooked in one cycle will advance to the clinical phase in the next one, following the natural history disease progression. These are so called interval cancers, in so far as they turn symptomatic in between two rounds of screening (Alanko, 2019).

Breast cancer specific survival. BCSS is estimated based on breast cancer specific Kaplan-Meier (KM) survival curves obtained from Karihtala et al. (2021), estimated based on patient data from Oulu University Hospital collected between 2003 and 2013. KM curves were estimated according to age, for the age groups <41, 41–69 and ≥70, as well as for early breast cancer (EBC) and metastatic breast cancer (MBC), respectively. For EBC, the time to event for which BCSS was calculated was defined as the period between surgical tumor removal and breast cancer specific death or end of follow-up. For MBC, it was defined as as the time from diagnoses of distant metastases to breast cancer specific death or end of follow-up. All patients diagnosed with non-metastatic NLBC or LBC who will experience recurrence in the surveillance phase are assigned to all-cause EBC survival, which is a function of EBCSS and baseline survival. All patients diagnosed with metastatic NLBC follow MBCSS.

The BCSS curves in Karihtala et al. (2021) were estimated based on prospective follow-up of a population cohort for a median of 102 months. In evaluating interventions with long term survival benefits realized after the end of a study's follow-up period, if patients are still alive at this point, it is necessary to extrapolate beyond the end of follow-up (Latimer & Adler, 2022). This is often the case with cancer treatment, when delaying disease progression improves survival. Extrapolation involves predicting how survival would proceed after the end of follow-up over the time horizon adopted for evaluating an intervention which, in the context of CEA,

usually is done by means of parametric survival modelling (Drummond et al., 2015, p. 317; Hoyle & Henley, 2011). Parametric models follow a distribution of time to event, specified by a set of parameters such as shape and scale, for example (Hoyle & Henley, 2011; Latimer & Adler, 2022). By fitting parametric models to survival data, values are estimated for its distribution parameters used to derive transition probabilities for disease-specific mortality, over a specified time horizon extending beyond that used for estimating the survival data.

Standard distributions, typically used for survival analysis in the field of health technology assessment (HTA), include the exponential, Weibull, Gompertz, log-logistic and log-normal specifications (Hoyle & Henley, 2011; Latimer & Adler, 2022). The choice of specification for parameterization of the survival model is based on its goodness-of-fit to estimated survival, usually measured in terms of best fit between modelled and estimated survival curves, as well as producing the lowest Akaike's Information Criteria (AIC) and Bayesian Information Criterion (BIC).

EBCSS for all age groups and MBCSS for ages 40–69 was extrapolated following the methods of Hoyle and Henley (2011), using a Microsoft Excel spreadsheet available from the journal article's publication platform to fit parametric models to survival data. The respective KM survival curves were digitized and transformed to data points using Engauge Digitizer (version 12.1). The data points were adapted to six-month intervals in Stata/SE 17.0 using linear interpolation to fit the Hoyle and Henley template. The template was filled out with the interpolated empirical survival probabilities and the size of the population at risk at the beginning of follow-up, reported in Karihtala et al. (2021), based on which the number of events were estimated. Once populated, an R language code, which can be found in the template, was generated and ran in RStudio (version 2021.09.2) for comparison of different distributions' fit to

survival data by maximum likelihood, and for estimating the distribution parameters. More specifically, the means and standard deviations of the best fit distribution parameters maximizing the likelihood, as well as their covariance and the variance-covariance matrix used for Cholesky decomposition, were recorded from the parametric model fit output.

Between the exponential, Weibull, log-logistic and log-normal distributions specifications, the Weibull specification achieved the best fit to all three relevant survival curves, producing the lowest AIC and BIC values. For each survival curve fit, the mean value of the distribution's shape and scale parameters, gamma γ and lambda λ, are used for calculating the deterministic probability of BCSS for time t according to:

$$\mathrm{S}(t) = exp(-\lambda t^{\gamma})$$

Distribution parameters for calculating probabilistic survival are simulated according to:

$$\left(\frac{\hat{\lambda}}{\hat{\gamma}}\right) + C_z$$

where C is the covariance matrix and z is a vector of standard independent normal variables (Hoyle & Henley, 2011).

The age- and stage specific survival probabilities were transformed to annual probabilities and calculated for a maximum of 70 and 40 years for the age-groups 40–69 and ≥70, respectively, assuming all simulated patients die at latest at the age of 110. In Figures B3.1-3.3 (Appendix B3), Weibull specification fits are plotted against the replicated age- and stage specific KM survival curves reported in Karihtala et al. (2021), visualizing their goodness-of-fit.

Baseline mortality. Age-specific baseline mortality rates were calculated as the number of deaths, excluding breast cancer-related deaths, divided by the female population size. Mean

death and population counts by 5-year age groups for the Finnish female population were obtained from Official Statistics Finland, combining national mortality data (OSF, 2020a) and population data (OSF, 2020b) between 2000 and 2018. The rates were converted to annual probabilities. To capture the impact of aging on survival, baseline survival applies to all health states, excluding metastatic NLBC. DCIS is assumed to follow baseline survival because of 1) the assumption that it would never be detected in the absence of screening and, therefore, cannot be assigned a disease-specific survival function in the scenario of no screening, and 2) the lack of relevant evidence on DCIS-specific survival.

Overdiagnosis. In the model, breast cancer patients are over diagnosed when, in the absence of screening, they would have never progressed to the clinical phase (or to invasive breast cancer in the case of DCIS) in their lifetime, or when they die from other causes before the disease progression would have occurred (Puliti et al., 2012). An intuitive approach to model overdiagnosis is separating between non-progressive and progressive lesions, the former of which, when identified by screening, are assumed to be over diagnosed (Tan et al., 2013).

The parameters used for modelling the natural history only apply this distinction to DCIS, whereas all invasive breast cancers entering the preclinical phase are assumed to be progressive. However, if one assumes that an unknown proportion of patients in the disease-free health state in fact have non-progressive breast cancer, overdiagnosis can be estimated by allowing these patients to be identified by screening and subtracted from the disease-free population. Because directly observing the degree of overdiagnosis for IBC is impossible, the number of over diagnosed cancers are simply assumed to be a proportion of the total number of histologically confirmed cancers, following the methods of Socialstyrelsen (2019). In this model, the over

diagnosed non-progressive cancers within a cycle are assumed to be a proportion of the total number of screening-detected LBC only as NLBC, by definition, is already progressed.

Estimates of overdiagnosis due to breast cancer screening vary considerably in the published literature, ranging from 0% to 50% of all histologically confirmed breast cancers (Puliti et al., 2012). The methodology for estimating overdiagnosis varies between studies, mainly depending on whether estimated for both IBC and DCIS, as well as whether estimated over the whole follow-up period of the study or only over the screening period, for women participating in screening (Puliti et al., 2012; Socialstyrelsen, 2019). Proportions of overdiagnosis applied to this model were obtained from a meta-analysis by Marmot et al. (2013), reporting an overdiagnosis rate of 19% (95% CI = 0.15–0.23) of screening-detected invasive breast cancer over the screening period, for women invited to screening. This estimate is used by Socialstyrelsen (2019) and supported by a Norwegian study which estimated an overdiagnosis rate for IBC between 15% and 25% (Kalager, Adami, Bretthauer, & Tamimi, 2012).

Cost parameters. This analysis applies a limited societal perspective as defined in Kim et al. (2020), following the HILA (2019) instructions for economic evaluations attached to application for reimbursement status. Thus, all direct healthcare and non-healthcare costs, informal care costs and productivity losses are accounted for. Estimates informing cost parameters are presented separately for screening- and treatment related costs.

Direct healthcare costs related to screening consist of invitation, primary screening test and diagnostic follow-up examination costs. According to cost-effectiveness analyses on breast cancer screening in Finland and Sweden by Leivo et al. (1999) and the Socialstyrelsen (2019), respectively, follow-up examination costs depend on whether or not biopsy was required to rule out breast cancer for false-positive results from the primary mammography test. Age-specific

proportions of false-positives ruled out through biopsy were obtained from the FCR annual review of the FBCSP, based on the total number of participants in 2019 (Anttila et al., 2021). The mean proportion for ages 50–59 and 60–69 was 28% and 40%, respectively. Due to lack of relevant evidence on the corresponding proportions for younger and older ages, the proportions for the age groups 40–49 and ≥70 were set equal to those of 50–59 and 60–69, respectively. True-positive breast cancers are always histologically confirmed through biopsy prior to surgical referral (Anttila et al., 2021). Costs of primary mammography screening tests were obtained from Mäklin and Kokko et al. (2021), estimating the average unit cost in Finland in 2017. Costs of follow-up examinations with and without biopsy, as well as pathological examinations of biopsies, were obtained from unit price lists of the VSSHP Division of Medical Imaging (2017). Due to lack of better evidence on costs of invitation to screening, the estimated cost by Socialstyrelsen (2019) was used.

Treatment costs for breast cancer are obtained from Lehtinen et al. (2019), which estimated annual age- and stage-specific (In situ, LBC, lymph node positive NLBC, NLBC of unknown spread and metastatic NLBC) treatment costs for different treatment phases, over a 10 year period since time of diagnosis. The costs were estimated by combining patient data from women diagnosed with breast cancer in VSSHP between 2004 and 2013 with estimated age- and stage-specific survival and risk of recurrence, as well as the hospital district's healthcare service unit price lists of 2017. The study reports average costs for the first year since diagnosis, the last year before death and the years in between. Following the study, if a patient dies within a year from diagnosis, the time counts towards last year before death only.

Direct non-healthcare costs of screening, consisting of transportation to and from primary mammography screening as well as follow-up examinations, were obtained from Leivo et al.

(1999) in lack of more updated evidence. Due to the same reason, estimates of productivity loss, in terms of opportunity costs of the time spent on primary mammography tests and follow-up examinations, as well as travel and waiting time, were informed by the same study.

Costs of transportation, informal care and productivity loss related to breast cancer treatment were obtained from Roine et al. (2019), estimating average costs for different breast cancer states over periods of six months. The cross-sectional survey and registry study is based on Social Insurance Institution records and data collected from questionnaires sent to breast cancer patients in the Helsinki and Uusimaa Hospital District between 2009 and 2010. The disease states, according to time lapsed since diagnosis, for which costs were estimated included primary treatment (<6 months from diagnosis), rehabilitation (6–18 months from diagnosis), remission (>18 months from diagnosis) and metastatic disease. Costs of informal care and productivity loss related to palliative care (i.e., end-of-life care after the termination of life-prolonging treatment for metastatic breast cancer) were obtained from Haltia et al. (2018), estimated for the average duration of palliative care of 59 days.

Utility parameters. Each health state is associated with a certain value of HRQoL expressed in terms of QALYs. The HRQoL of a cycle spent in the disease-free health state is assumed equal to 1 QALY, representing full HRQoL. The same quality of life applies to patients in all preclinical health states, as they do not yet know of or experience any symptoms from the disease. Total experienced disutility from screening is equal to the sum of utility decrements related to initial mammography tests and follow-up assessment. A literature search for screening related disutility was conducted, the details of which are presented in Appendix B4. As the literature suggests no relevant impact of mammography on HRQoL (Pataky et al., 2014; Rijnsburger et al., 2004), this model assumes zero disutility from primary screening tests.

No relevant sources for disutility associated with follow-up examinations excluding biopsy were found. Therefore, a utility weight incorporating both diagnostic mammography and the risk of biopsy used in Pataky et al. (2014) was applied to all patients called back to follow-up examinations collectively.

Utility weights for breast cancer treatment were informed by Rautalin et al. (2018), in which HRQoL of breast cancer patients treated in the Helsinki and Uusimaa Hospital District of Finland was estimated for five mutually exclusive disease stages, separately, based on time since diagnosis and tumor stage. Similar to Roine et al. (2019), the disease stage separation included primary treatment (<6 months from diagnosis), recovery (6–18 months from diagnosis), remission (>18 months from diagnosis), metastatic disease and palliative care.

Parametrization of cost and utility estimates. All cost and disutility estimates were adjusted to the modelled cycle length of one year. Where needed, costs were converted to euros and adjusted to the 2021 price level, according to the same methods used for adjusting the estimates presented in Table 1.

Productivity loss was applied to patients under 62 years of age only, reflecting the average effective retirement age for women in Finland 2021 according to the Finnish Centre for Pensions (ETK, 2022). For the purpose of sensitivity analysis, cost and disutility estimates for which no measures of variation was reported were assigned gamma and beta distributions, respectively, reflecting the nature of the data informing them (Briggs et al., 2006, pp. 91-92). An overview of cost and utility parameters used in the model is presented in Appendix B4.

Model Uncertainty

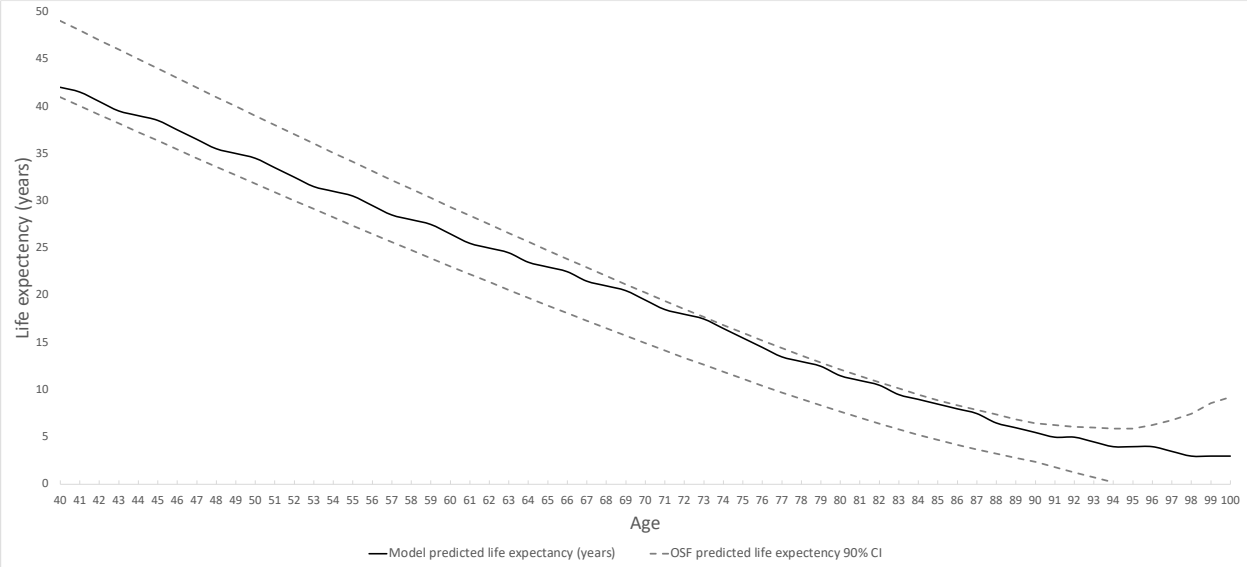
External model validation. In addition to estimating unknown parameters, the natural history parameter calibration process verified consistency between model outcomes and

observed rates of pre-screening era incidence. External validation supports credibility of model predictions and accuracy of input parameters, as well as their implementation in the model (Hunink et al., 2014, p. 351). In addition to incidence, observed mortality rates are commonly used for external model validation

Observed breast cancer specific mortality could not be calibrated against in estimating natural history parameters, as pre-screening era mortality rates would disregard the improvement in breast cancer treatment since the FBCSP was launched. Therefore, consistency between model predicted and post-screening era observed all-cause mortality, when allowing patients aged 50–69 to be detected by screening, was assessed through life expectancy predictions. In Figure 7, model predicted life expectancy is plotted against the 90% CI OSF predicted age-specific life expectancies for women aged 40 to 100 in 2018 (OSF, 2022b), supporting credibility of the model implementation of baseline and disease-specific mortality parameters.

Figure 7

Model predicted life expectancy compared to OSF predicted life expectancy.



Discounting and half-cycle correction. Where preferences with regard to the timing of costs and utilities are present, future outcomes should be duly expressed in terms of present value (Drummond et al., 2015, p. 241; Edlin et al., 2015, p. 85). People generally have positive time preference rates (i.e., costs and benefits are preferred today rather than in the future) due to uncertainty about the future or being present-oriented, for example. Because outcomes are predicted for every cycle over the entire time horizon of the model, they should be discounted to present values. As recommended by HILA (2019), both costs and utilities were discounted with an annual factor of 3%.

According to the model structure, the distribution of the simulated population between model health states depends on the state where patients reside at the end of each cycle, after all potential events have occurred. By construction, this means that all events are assumed to occur at the start of each cycle. However, the timing of events within a specific cycle is unknown. Rather than assuming all events to occur at the start of a cycle, it is more intuitive to assume the timing of events to be symmetrically distributed over the cycle (Briggs et al., 2006, pp. 33-36; Edlin et al., 2015, pp. 84-85). Half-cycle correction (HCC) reflects the assumption that, on average, events are assumed to occur half-way through a cycle. HCC is applied to all discounted outcomes for every cycle, calculated as the average between a cycle t and cycle $t+1$.

Sensitivity analysis. The impact of overdiagnosis relating to invasive breast cancer on model predictions was assessed through deterministic sensitivity analysis, recording the ICERs for various values of the overdiagnosis rate while holding all other parameters constant. The range within which the overdiagnosis rate is varied is set to 0%–40%, reflecting the variation in literature regarding estimated rates of overdiagnosis associated with invasive breast cancer.

For the purpose of PSA, all calibrated parameters were assigned beta and Dirichlet distributions based on the nature of the data informing them. PSA was conducted simultaneously for both screening intervention strategies, using Monte-Carlo simulation to draw random values from all input parameters' respective distributions. For each of a total of 1 000 iterations, parameter values and model outcomes were recorded in order to calculate the intervention strategies' and their comparators' ICERs and NMBs associated with every specific iteration. ICERs were plotted on a CE-plane to display the uncertainty surrounding its true value, whereas the probability of cost-effectiveness at given WTP-thresholds are presented with CEAFs.

Due to the lack of an explicit WTP-threshold in Finland, a threshold in terms of costs per QALY gained of €50 000 or twice the GDP per capita is often used (Hallinen & Soini, 2011), the latter of which would be approximately €80 000 as of 2021 (OSF, 2022a). These values are used as lower and upper bound WTP-thresholds in the sensitivity analyses, reflecting an acceptable range within which society is willing to pay per incremental QALY gained. All sensitivity analyses use discounted and half-cycle corrected costs and outcomes.

Value of information analysis. EVPI can be estimated directly from the PSA output as the highest mean NMB of all simulated iterations at a given threshold, subtracted by the mean of the highest NMB at each simulated iteration for the interventions being compared (Drummond et al., 2015, p. 411), according to the following formula:

$$\text{EVPI} = E_{\theta} max_j \text{NMB}(\text{j}, \theta) - \text{max}_j E_{\theta} NMB(j, \theta) \quad (1)$$

where θ represents an uncertain parameter and j a particular value of θ.

EVPPI is computationally more demanding to calculate, requiring both an inner and outer simulation loop, the former estimating expected NMB for each iteration of θ_1 and the latter randomly sampling possible values for θ_1 (Drummond et al., 2015, p. 411), according to:

$$EVPPI_{\theta_1} = E_{\theta}\max_j E_{\theta_0|\theta_1} NMB(j, \theta_1, \theta_2) - max_j E_{\theta_0,\theta_1} NMB(j, \theta_0, \theta_1) \quad (2)$$

VoI analysis was conducted for both intervention screening strategies using the University of Sheffield Accelerated Value of Information (SAVI) online application (version 2.2.0), available at https://savi.shef.ac.uk/SAVI/. SAVI estimates EVPI and EVPPI through non-parametric methods explained in Strong, Oakley and Brennan (2014), based on PSA outputs that are uploaded to the application. It is preferred to the nested Monte Carlo approach in equation (2) as it produces very similar outcomes using less computational power. Individual and population EVPI, as well as EVPPI for parameter groups consisting of natural history parameters, screening-specific parameters, cost and utility components and the overdiagnosis rate were estimated using SAVI. Population EVPI (pEVPI) is calculated based on annual BC prevalence in Finland (Cancer Society of Finland, n.d.).

In estimating EVPPI, the lower bound WTP-threshold was applied for the current practice screening strategy, while the upper bound was applied for the more extensive screening strategy. For both intervention strategies, the EVPPI was estimated for the following parameter groups: 1) Natural history parameters 2) Screening-specific parameters (i.e., sensitivities and specificities), 3) Direct treatment costs, 4) Productivity loss (i.e., costs due to sick days and early retirement associated with treatment), 5) Informal care costs, 6) Transportation costs associated with treatment, 7) Screening costs (i.e., costs of primary mammography test and follow-up procedures, as well of as transportation and productivity loss associated with screening), 8) Utility weights associated with treatment, 9) Disutility associated with screening and 10) The rate of overdiagnosis.

Results

Base-Case Analysis

The base-case analysis results for current practice of screening women aged 50–69 compared to no screening, as well as screening women aged 40–74 compared to current practice, are presented in ICERs representing incremental costs per QALY gained. The ICERs presented in Table 2 are calculated for both undiscounted and discounted half-cycle corrected costs and outcomes, the latter by applying a discount rate of 3% to both costs and outcomes.

Table 2

Undiscounted and discounted ICERs of the intervention screening strategies.

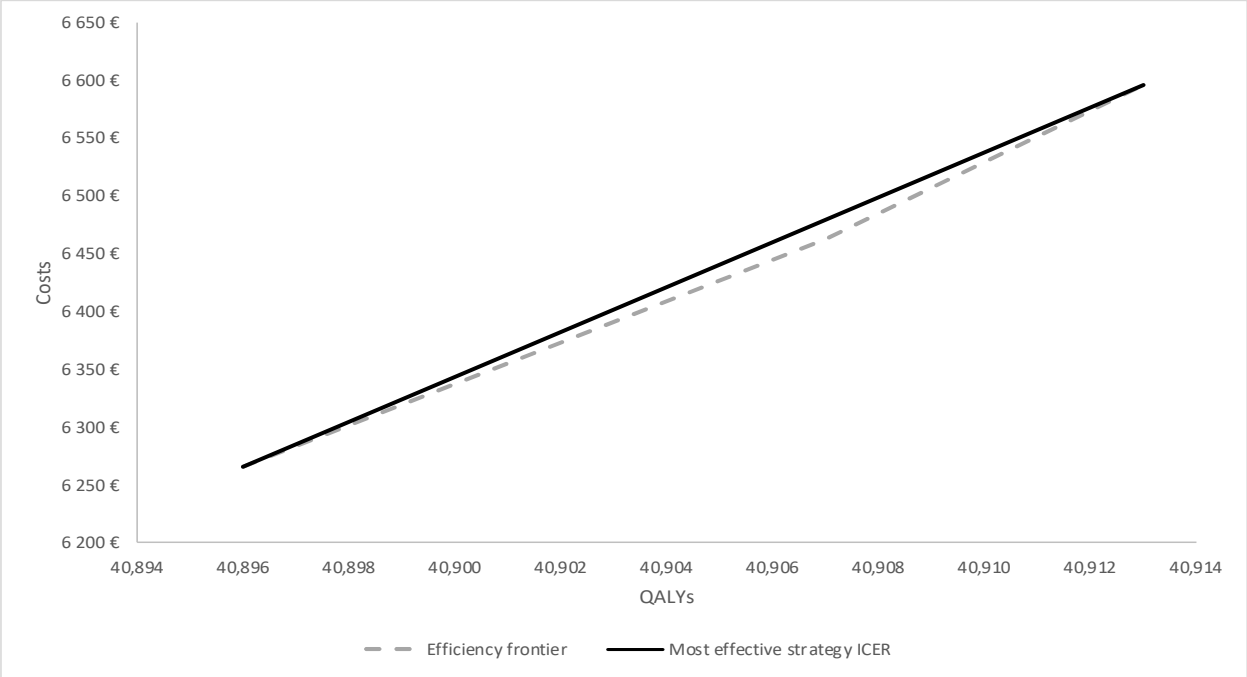
	NO SCREENING	**SCREENING 50–69**	**SCREENING 40–74**
Undiscounted costs	€6 885	€7 086	€7 223
Undiscounted QALYs	43.96	43.97	43.98
ICER (EUR/QALY)		€15 989	€18 667
Discounted costs €	€6 266	€6 462	€6 596
Discounted QALYs	40.896	40.907	40.913
ICER (EUR/QALY)		**€18 548**	**€21 580**

The discounted ICER for the current practice screening strategy compared to no screening was €18 548, whereas for screening all women aged 40–74 compared to current practice it was €21 580 per QALY gained. Based on these findings, it appears as if both screening strategies are cost-effective, when compared against the WTP-threshold interval of €50 000–€80 000 per QALY. This interpretation, suggesting that none of the strategies were strictly dominated (i.e., producing less QALYs to higher or the same cost), is supported by Figure 8

showing both strategies' ICERs on the efficiency frontier. The efficiency frontier consists of all non-dominated alternatives' ICERs when comparing more than two (Briggs et al., 2006, p. 149).

Figure 8

Cost-efficiency frontier and the ICER of the most cost-effective strategy.



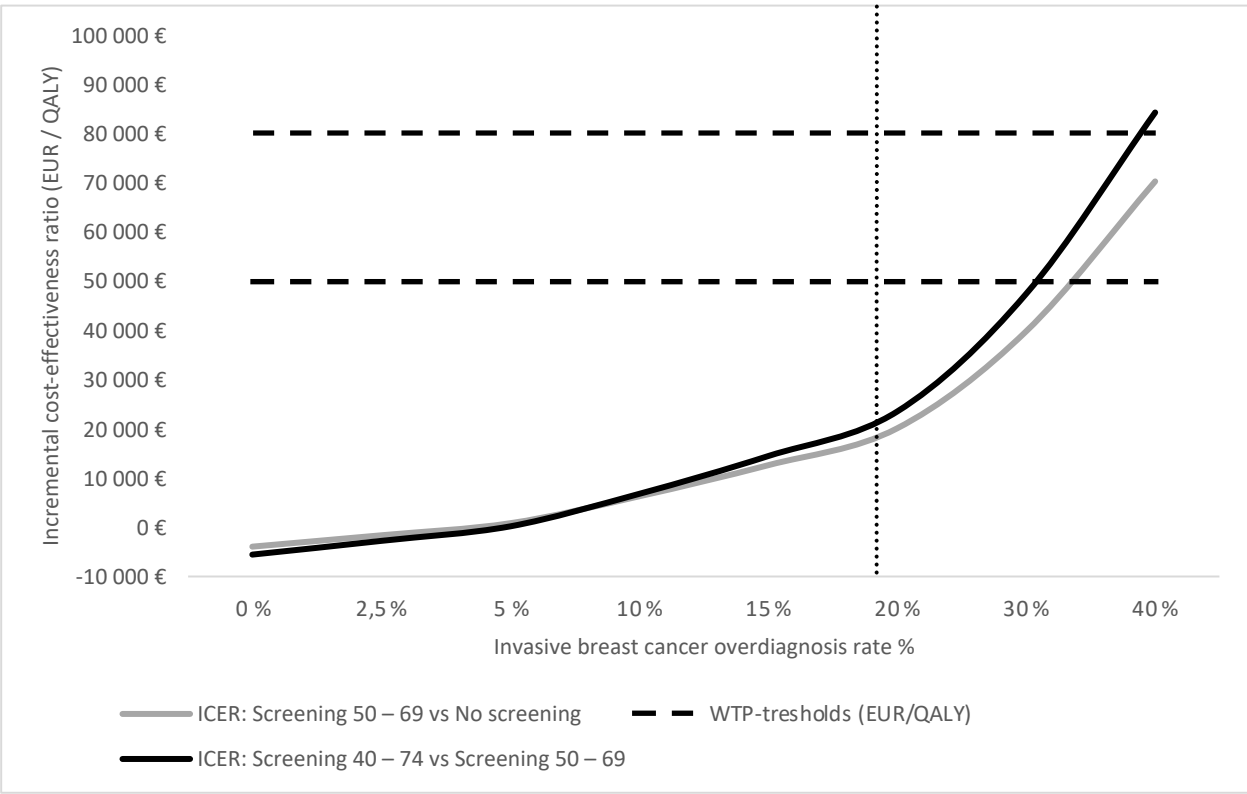
Impact of Overdiagnosis on Model Predictions

The impact of overdiagnosis associated with IBC on model outcomes is presented in Figure 9, plotting the predicted ICERs for different overdiagnosis rates against the lower and upper bound WTP-thresholds. Within the range of overdiagnosis rates included in the DSA, the ICER of current practice screening never exceeds the upper bound threshold, whereas that of screening women aged 40–74 exceeds it at a rate just below 40%. The lower bound WTP-threshold is exceeded by both strategies at rates of around 30%–35%. At low rates (0%–5%) both intervention strategies appear cost-saving, suggesting that they produce more QALYs at

lower costs than their comparators. The overdiagnosis rate applied in the base-case scenario is marked by a vertical dotted line, at which the discounted ICERs in Table 2 were estimated.

Figure 9

Impact of the overdiagnosis rate on the cost-effectiveness of screening strategies.



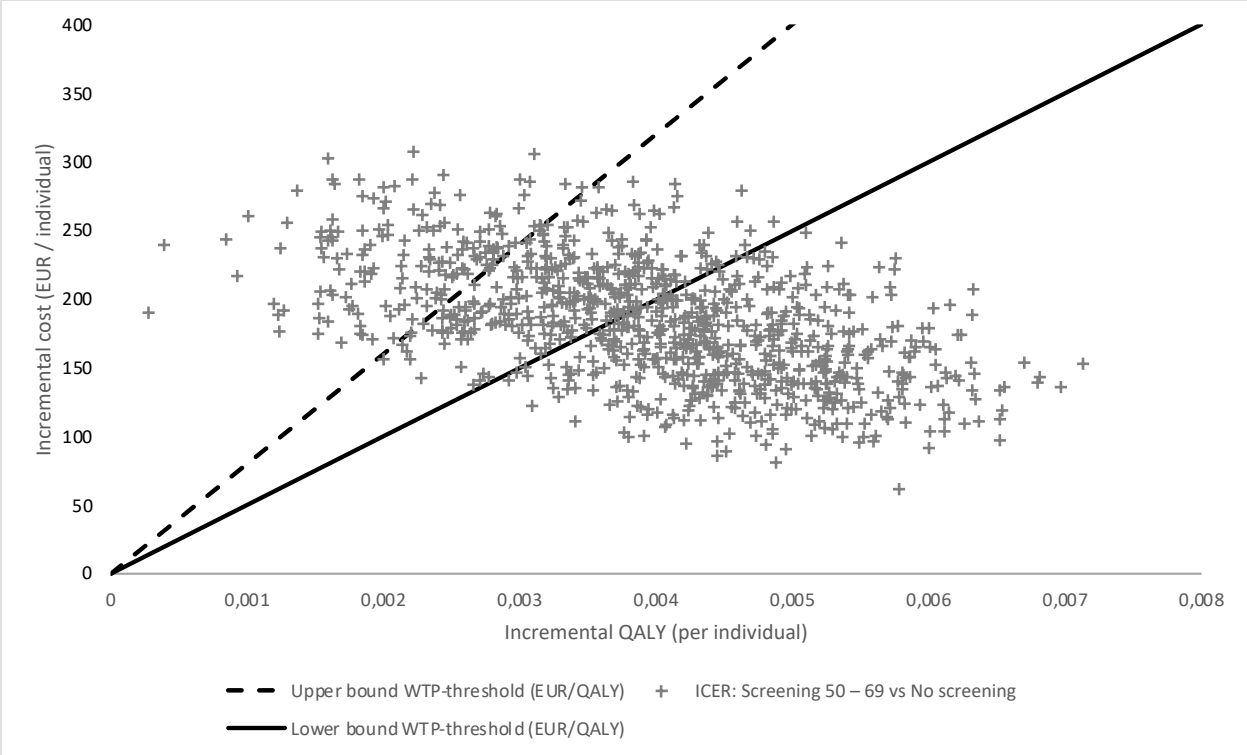
Probabilistic Sensitivity Analysis

The results of the probabilistic sensitivity analysis are presented in CE-planes and CEAFs for each intervention strategy. The data points in the CE-plane represent 1 000 simulated ICERs plotted against the lower and upper bound WTP-thresholds. The dispersion of data points represents the impact of parameter uncertainty on the ICER, a more dispersed cluster of data points reflecting higher parameter uncertainty. Because the decision between screening strategies affects not only breast cancer patients, but the whole population participating in screening, the ICERs plotted in Figures 10 and 12 represent the individual incremental cost per QALY, for each

PSA simulation. The CEAFs are drawn against a range of threshold values between €0 and €300 000, allowing for the possibility of the interventions and their comparators to converge to a 100% probability of being cost-effective relative to each other. The shaded rectangles in Figures 11 and 13 represent the lower and upper bound WTP-thresholds.

Figure 10

CE-plane for the current practice screening strategy compared to no screening.



The ICERs sampled for the intervention strategy of current screening practice compared to no screening are all in the north-east quadrant of the CE-plane, suggesting that over all iterations of the PSA simulation, the intervention strategy produces positive incremental QALYs to higher costs. The associated CEAF indicates that the current practice screening strategy has a probability of over 50% to be cost-effective compared to no screening over the relevant WTP-threshold interval, converging to 100% at a threshold value of ca. €180 000 per QALY.

Figure 11

CEAF for current the practice screening strategy compared to no screening.

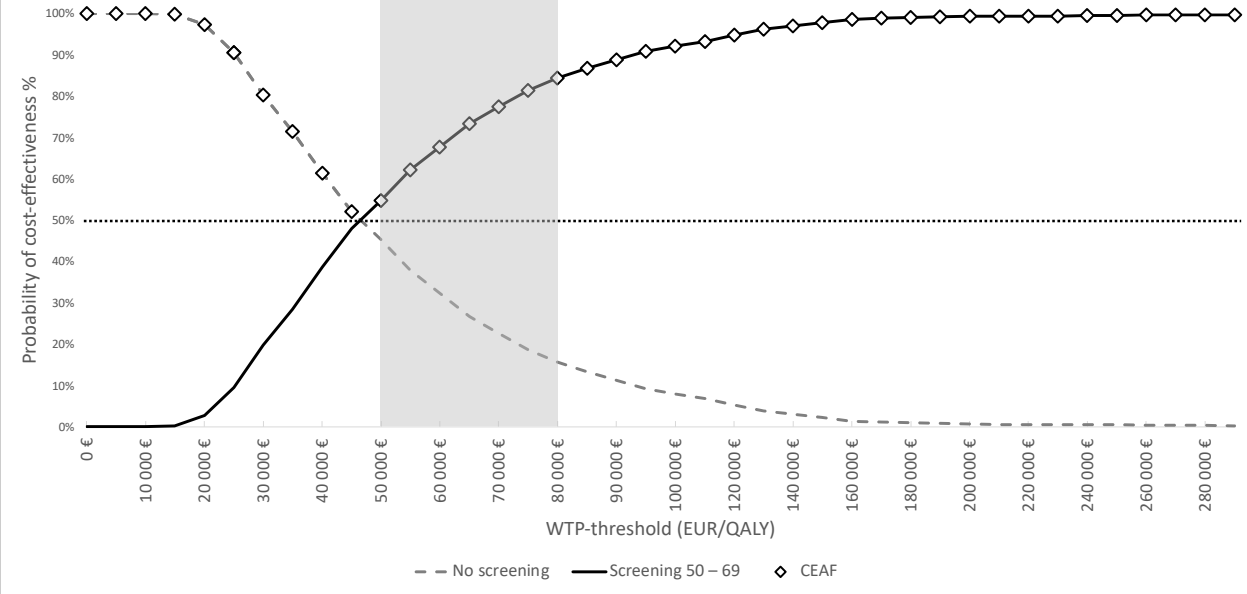
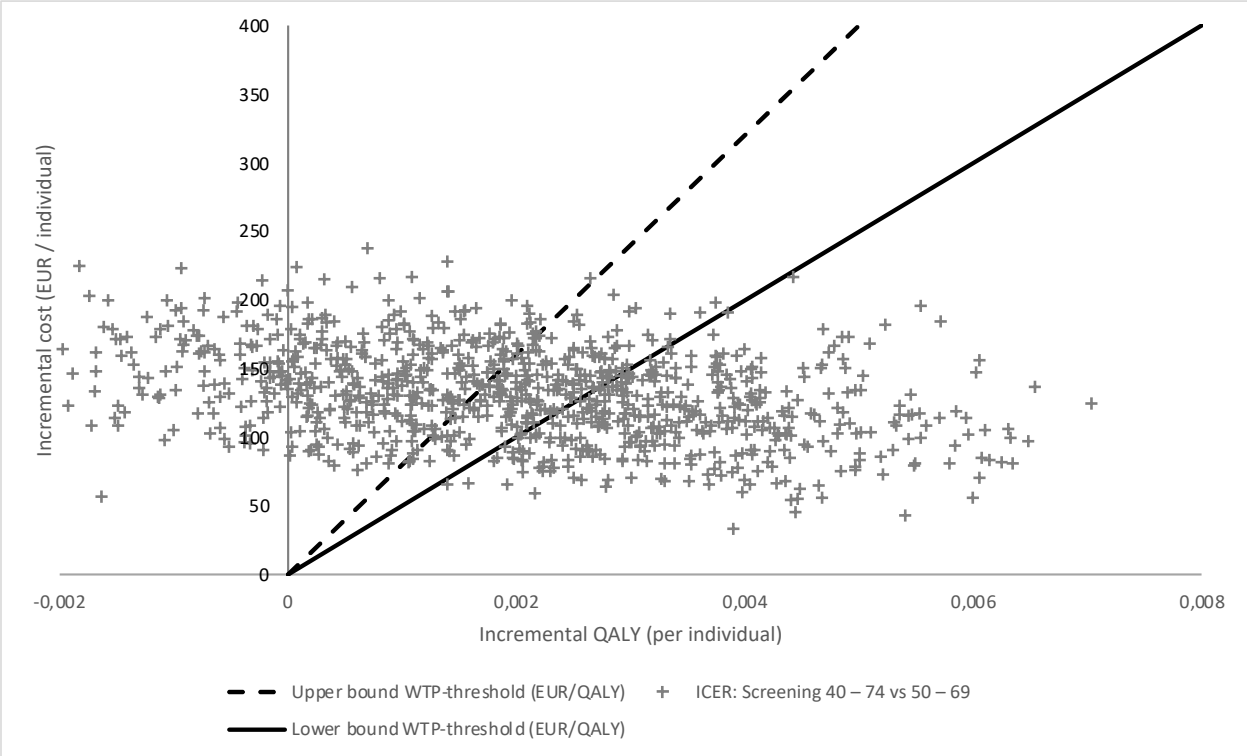


Figure 12

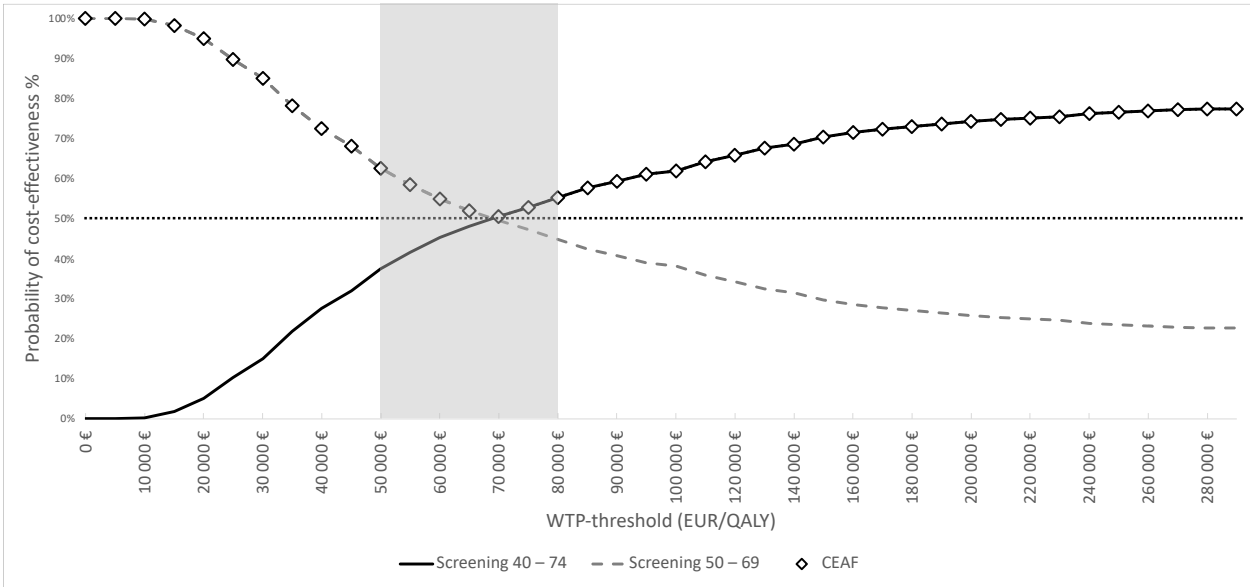
CE-plane for screening women aged 70–74 compared to current practice.



For the intervention strategy of screening women aged 40–74, some ICERs have been predicted in the north-west quadrant of the CE-plane, suggesting that adverse effects of the more strategy outweigh the gains of early detection. Most of the iterations, however, are in the nort-east quadrant. The CEAF for the more extensive screening strategy demonstrates a higher probability of being cost-effective than current practice, however, only at a WTP-threshold of ca. €70 000 per QALY, never converging to 100% over the thresholds included.

Figure 13

CEAF for screening women aged 40–74 compared to current practice.

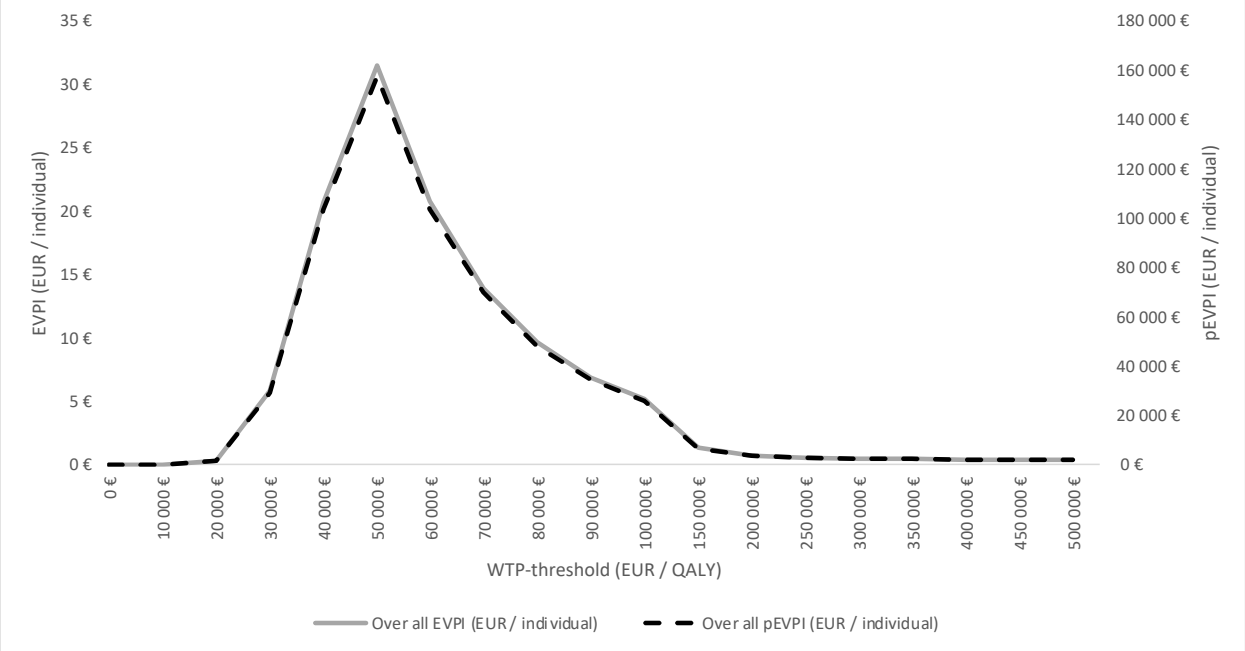


Expected Value of Perfect Information

According to Figure 14, as the threshold value increases from €0 to €50 000 per QALY, the EVPI and pEVPPI for the current practice screening strategy compared to no screening also increase. They peak at the lower bound WTP-threshold, at which individual and population EVPI were €31.5 and €157 400, respectively. When further increasing the threshold, the EVPI and EVPPI decrease, the former going below €1 at the threshold value of €200 000 per QALY.

Figure 14

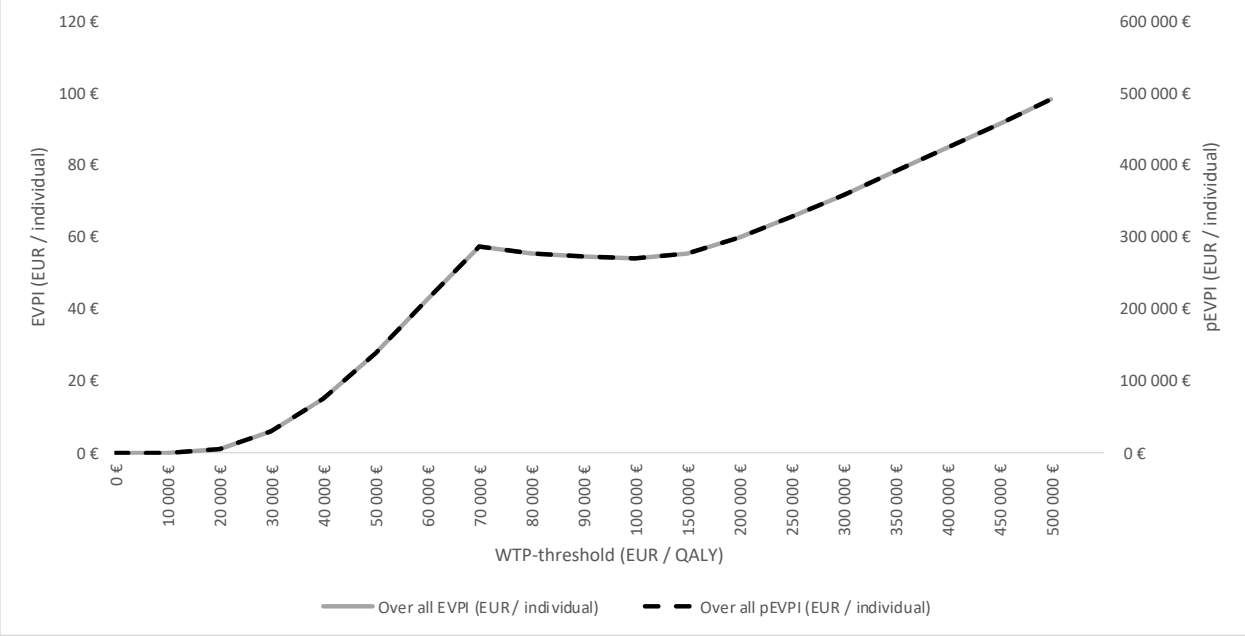
Individual and population EVPI of current practice compared to no screening.



For the more extensive screening strategy compared to current practice, EVPI and EVPPI peak at the WTP-threshold of ca. €60 000 per QALY, as seen in Figure 15. When further increasing the WTP-threshold they decrease slightly up until a threshold of around €90 000 per QALY, after which they increase again, reaching a value of approximately €100 for EVPI and €500 000 for pEVPI at the threshold of €500 000 per QALY.

Figure 15

Individual EVPI for screening women aged 40–74 compared to current practice.



Expected Value of Partial Perfect Information

The EVPPI estimated for current practice screening compared to no screening, presented in Figure 16, suggests that collecting more evidence informing the parameter groups of the overdiagnosis rate, natural history parameters, treatment costs and screening-specific parameters would generate the highest rate of return to research.

For the more extensive screening strategy compared to current practice screening, the parameter groups associated with the highest EVPPI are largely the same, as seen in Figure 17, although in different order according to the rate of return to further research. The screening-specific parameters showed the highest expected value of collecting more evidence, while the informal care cost parameter group surpassed the overdiagnosis rate in the same regard.

Figure 16

EVPPI of current practice screening compared to no screening.

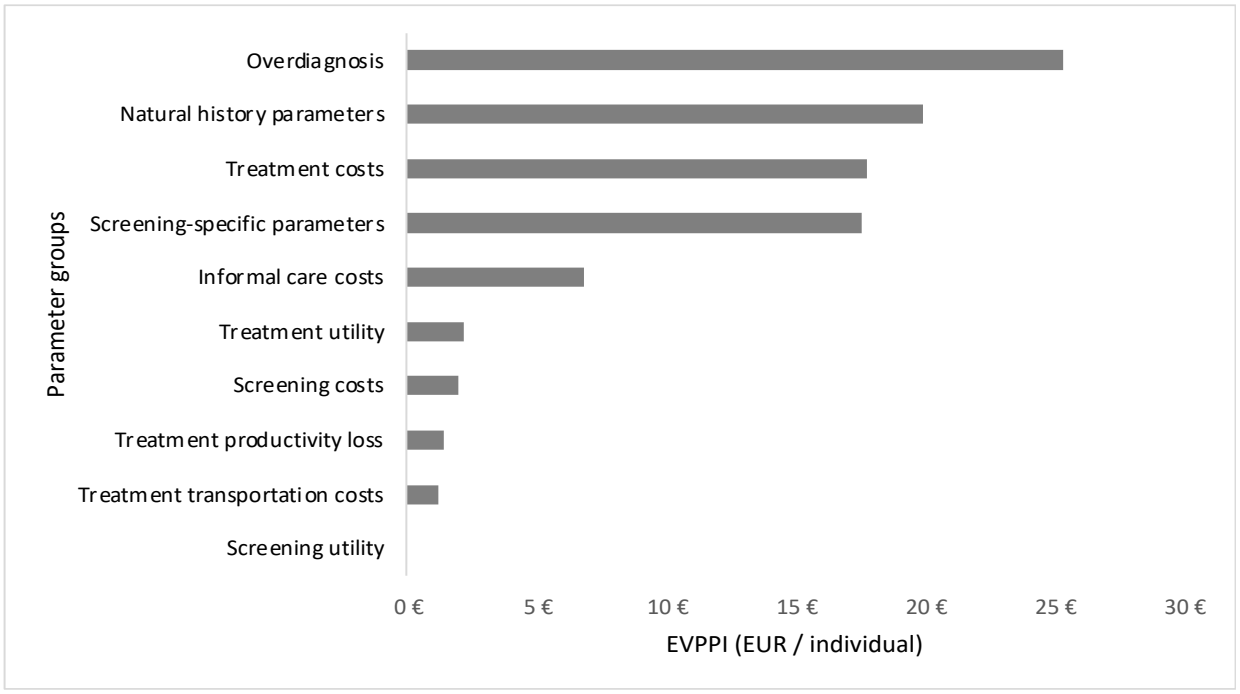
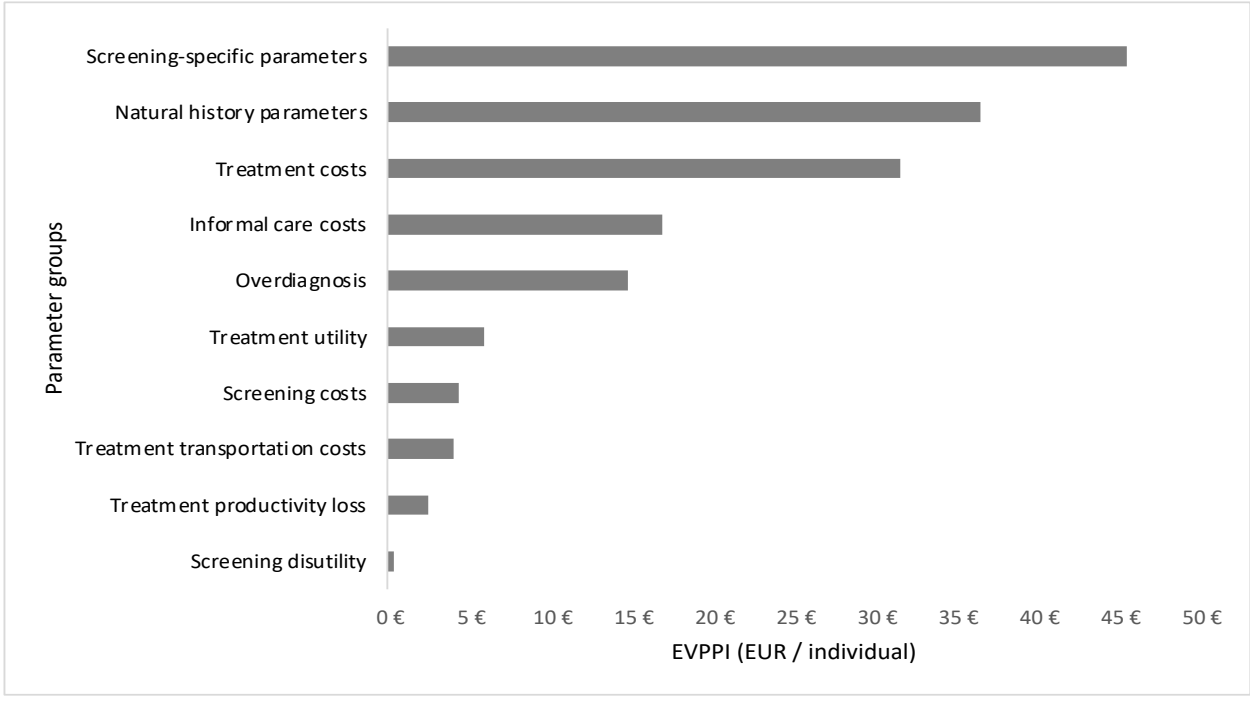


Figure 17

EVPPI of screening women aged 40–74 compared to current practice.



Discussion

Main Findings

Current practice screening for breast cancer does seem to increase overall QALYs at an acceptable cost compared to no screening. The more extensive screening strategy may also provide gains in QALYs compared to current practice, however, at somewhat higher costs. The expected incremental cost per QALY from expanding the FBCSP to include all women aged 40–74 is within the acceptable WTP-threshold range, however, when accounting for parameter uncertainty, it is the preferred strategy only at a threshold value above €70 000 per QALY.

The ICER of €18 548 per QALY for biennial screening of ages 50–69 is largely in line with the relevant literature presented in Table 1, and very similar to Socialstyrelsen (2019) as well as van Luijt, Heijnsdijk and de Koning (2017), which estimated ICERs of €22 513/QALY for the Swedish and €21 500/QALY for the Norwegian population, respectively. The corresponding ICER estimated by Mittmann et al. (2015) of €80 389/QALY is about four times as high as the estimates for the Swedish and Norwegian settings. It was, however, estimated for the female population of Canada, for a different healthcare system using a different modelling approach. Studies have shown variation in tumor subtype according to ethnicity (Ooi, Martinez, & Li, 2011; Parise, Bauer, & Caggiano, 2010) and while healthcare provision in Canada is similar to the Nordic model, healthcare system funding is different, suggesting inferior transferability of the findings of the study to the Finnish setting, compared to those for the Swedish and Norwegian settings.

The oval shape of the "cloud" of ICERs for the current practice screening strategy, presented in Figure 10, suggests that outcomes are more sensitive to parameter uncertainty than costs. Most of the iterations are concentrated around the lower bound WTP-threshold, indicating

a probability of at least 50% of the intervention strategy to be cost-effective within acceptable WTP-threshold range. This is more evident when looking at the CEAF, which reveals the intersection between the CEACs of the intervention and comparator strategies, representing the threshold value at which the decision flips (i.e., when the intervention has a higher probability of being cost-effective). Because this critical threshold is lower than the lower bound WTP-threshold of €50 000 per QALY, the probability of the intervention strategy to be cost-effective is ≥50% over the acceptable threshold range, illustrated by the shaded grey area in Figure 11.

For the intervention strategy of screening women aged 40–74 compared to current practice, more horizontal variation in the simulated ICERs seen in Figure 12, compared to the CE-plane for the current practice screening strategy, suggest even higher sensitivity of outcomes to parameter uncertainty. The iterations in the north-west corner of the CE-plane are, however, rather limited and can be the result of extreme values of one or more parameters for these iterations. In Figure 13, the CEAF for the intervention strategy of screening women aged 40–74 shows a higher WTP-threshold associated with the decision flip, the intersection between the CEACs of the intervention and comparator strategies appearing at a threshold of ca. €70 000 per QALY. Furthermore, while the CEAC for current practice screening converges to a 100% probability of being cost-effective compared to no screening, the CEAC of the more extensive strategy compared to current practice never does over the same threshold range.

These remarks suggest higher decision uncertainty surrounding the intervention strategy of screening women aged 40–74 compared to current practice, than for current practice screening compared to no screening. This is supported by the respective strategies' EVPI and pEVPI presented in Figures 14 and 15, respectively, showing higher expected values of perfect information for the more extensive screening strategy over all WTP-thresholds. While the EVPI

and pEVPI for current practice screening converge to €0 with higher threshold values after peaking, the expected value of perfect information for the more extensive strategy falls slightly after peaking at a WTP-threshold of around €70 000 per QALY gained, only to increase again when the threshold exceeds €100 000 per QALY. The decision uncertainty increases with higher thresholds at almost the same rate as before peaking, suggesting considerable costs associated with making the wrong decision.

The ICERs of the intervention strategies appear sensitive to changes in the assumed proportion of over diagnosed patients with IBC. According to Figure 9, the critical values of the overdiagnosis rate, at which the ICERs exceed the lower bound WTP-threshold, is around 30%–35%. For an overdiagnosis rate up to around 38%, both intervention strategies produce an ICER within the upper bound WTP-threshold, which is exceeded by the more extensive screening strategy at higher rates. With both intervention strategies predicting negative ICERs (due to negative incremental costs) and positive ICERs above the average WTP-threshold when varying the overdiagnosis rate between the extreme values of 0% and 40%, one can conclude that the cost-effectiveness of screening is sensitive to the rate of overdiagnosis. This is confirmed by the EVPPI, especially for the current practice screening strategy, for which overdiagnosis was associated with the highest rate of return to research.

To gain more insight into why the more extensive strategy produces the incremental costs and QALYs compared to current practice presented in Table 2, one needs to assess the impact of screening ages 40–49 and 70–74 on the ICER, separately. Interestingly, screening all women aged 40–69 compared to current practice produces a deterministic ICER of €4 349 per QALY, whereas screening women aged 50–74 produces an ICER of -€14 829 due to negative QALYs. Therefore, to some extent, screening in ages 70–74 thwarts the effectiveness of screening women

aged 40–49. These findings are contradictory to those of Socialstyrelsen (2019), which found screening women aged 50–74 favorable to screening ages 40–69, in terms of incremental costs per QALY gained. The intuitive answer to why the model predicts a more favorable ICER for screening in ages 40–49, is that it offsets more costs and utility decrements of later stage breast cancer than it inflicts through overdiagnosis and false-positives, compared to screening in ages 70–74. This, in turn, is possible due to a number of reasons, including 1) age-specific natural history and screening parameters, 2) differences in offsets of productivity loss and 3) differences in disease-specific survival prospects for ages 41–69 and ≥70.

In the study by Socialstyrelsen (2019), it was noted that the results were particularly sensitive to the method for calculating costs associated with productivity loss. This suggest that the offset of productivity loss, which is contingent upon the proportion of cancers prevented to progress to later stages with higher associated productivity loss, can explain at least in part the difference between the ICERs estimated for the respective screening scenarios included in the study. In this model, because the rate of tumor progression impacts the benefit of early detection in preventing later stage cancer and the associated higher cost and disutility, the natural history and screening-specific parameters for ages 40–49 may overestimate the effectiveness of screening in this age group. As the EVPPI for the strategy of screening women aged 40–74 suggests, these parameter groups, consisting of parameters estimated separately for the age groups 40–49 and ≥70, are associated with the highest returns to further research. Likewise, productivity loss, which is only relevant to simulated patients under the age of 62, can explain some of the difference between the ICERs of the screening strategies evaluated in this analysis. Compared to current practice screening, total discounted productivity loss costs were over €3 million higher for screening women aged 50–74 than for screening women aged 40–69. It may

also be considered unethical to include productivity losses in the analysis, as it disfavors the elderly non-working population (Socialstyrelsen, 2019).

The third potential background to the favorable ICER of screening in ages 40–49 over ages 70–74 may offer some explanation, in so far as the rate of overdiagnosis is overestimated for older ages, falsely assigning too many patients with nonprogressive cancers to breast cancer specific survival. If this is the case, screening in ages 70–74 may underestimate total utility in this age group, if the disutility from overdiagnosis offsets the utility of early detection. This is further explained in the next section on model limitations.

Model Limitations

The model developed for estimating the cost-effectiveness of the strategies suffers a number of limitations, which can mainly be attributed to 1) the nature of the evidence informing the natural history parameters, assumptions made in synthetization of the evidence and modelling the natural disease progression, 2) evidence used and assumptions made in modelling the surveillance phase of simulated patients diagnosed with breast cancer and 3) the approach used for incorporating overdiagnosis associated with screening in the model.

While the studies informing the parameters used for modelling the natural history of disease were estimated from Finnish and Swedish female populations, which are not expected to differ significantly between themselves in terms of patient characteristics, the natural history models used for estimating invasive breast cancer natural history parameters differed with regard to the stage classification adopted. Duffy et al. (1997) separated between LN positive and negative tumors and, despite the similarity between estimated transition probabilities for the 50–59 age group compared to those in Wu et al. (2010), the potential inconsistency between the FCR tumor stage classification and one according to lymph node involvement may contest their

compatibility with the model structure, even after calibration. These parameter estimates were, however, considered best available evidence given the limited published literature on natural history parameters approximated through a five-state model. Furthermore, integrating the DCIS natural history parameters estimated by Yen et al. (2003) with the invasive breast cancer transition parameters required adjustment to the underlying incidence associated with both DCIS and IBC, as the study assumed all invasive breast cancers to be preceded by DCIS. Also, no separation between localized and non-localized tumors regarding progression from DCIS to IBC was made. However, because limiting progression to either LBC or NLBC would have distorted the IBC stage distribution, adjustments to the underlying incidence parameters were necessary.

The assumptions made in modelling the natural history of breast cancer directly affect decisions regarding for what ages screening should be initiated and terminated. The input parameters and their distributions used for calibration follow the general pattern of decreasing rates of progression and increasing underlying incidence rates with age. However, simply fitting model predicted incidence to observed data does not guarantee correct parameter estimation, as there are multiple intercorrelated elements behind the modelled natural disease progression.

The surveillance phase of the model required more specific separation regarding disease progression of NLBC, due to the way the data on recurrence rates and treatment costs were structured. Importantly, modelling survival after diagnosis required assumptions regarding the probability of local recurrence which, in the model, determines whether patients are assigned to baseline or all-cause breast cancer survival. Ideally, after diagnosis, patients would be subject to annual probabilities of experiencing recurrence within a specified period of time, based on which they are assigned to survival functions allowing for disease-specific mortality. However, because the rates of local recurrence applied in the model were estimated by Lehtinen et al. (2019) as

overall probabilities over an undefined time period, simulated patients were simply assigned to cohorts that either will or will not experience local recurrence at some point within five years after diagnosis, based on age- and stage-specific recurrence rates. Because the timing of recurrence could not be identified, patients assigned to the cohort that will experience recurrence were assumed to be in remission for 10 years, which potentially overestimates time spent in remission and the disutility and costs associated with it. Distant recurrence, defined as tumors detected in a more advanced disease stage than initially treated for, were not considered in the model due to lack of relevant evidence on the rates associated with distant recurrence and infeasibility to incorporate it in the model. This, however, may overestimate the effectiveness of EBC treatment, especially for NLBC which can recur as distant metastasis (Wu et al., 2016).

Finally, the method of modelling overdiagnosis associated with invasive breast cancer relies on the assumption that screening detected progressive cancers are never over diagnosed. This may not hold, as overdiagnosis includes not only tumors that would never have progressed to the clinical phase in the absence of screening, but also tumors that don't progress before the patient dies from natural causes. Just as the rate of overdiagnosis is highly uncertain, the share of all over diagnosed cancers being nonprogressive is unknown and, therefore, adjustment for tumors not progressing before the patient dies cannot be made. Although the latter represents only a small share of all over diagnosed cancers in younger ages due to low associated baseline mortality, the total number of over diagnosed cancers in older ages with higher associated baseline mortality may be overestimated, as the same rate of overdiagnosis regarding nonprogressive invasive breast cancer is applied.

Model Strengths

The model developed for this analysis allows for assessment of costs and consequences associated with various screening strategies, as they are simply superimposed on a model of the natural history of breast cancer, allowing them to interrupt the natural disease progression. Albeit the source of many potential limitations of the model, the simulation of patients through a natural disease history, reflecting the stage classification of the FCR, is also its main strength. The calibration process further improved the credibility of the uncertain and unknown natural history parameters, generating consistency between model predicted and observed pre-screening era incidence rates.

The state values associated with each modelled health state were highly relevant for the population setting for which the intervention strategies were evaluated, including all major components associated with the limited societal perspective recommended in national guidelines. This enabled expected costs and outcomes of breast cancer screening in the Finnish setting, over an appropriate time horizon, to be evaluated in a CUA, which is vital for making decisions on publicly funded health interventions for diseases of great burden such as breast cancer.

Transferability and Suggested Further Research

The findings of this study are transferrable in so far as they are applied in the context of a similar setting with a similar healthcare system, when comparing expected costs and consequences of the intervention strategies and comparators used in this analysis, from a similar perspective. The cost-effectiveness of screening strategies was estimated based on a model designed to reflect costs and outcomes associated with breast cancer and screening relevant for the Finnish setting. The model's cost parameters, in particular, are estimated specifically for the healthcare system and the female population of Finland and, therefore, are unlikely to be relevant

for other settings. Caution is to be made regarding the population to which the findings of this study are applied. While the applicability of the natural disease progression may not be exclusive to the Finnish female population, it might not be appropriate to reflect that of all ethnicities.

To support decision making on the optimal screening strategy of the FBCSP, further research is recommended particularly on the natural history of breast cancer, as well as the sensitivity and specificity of mammography screening for different age groups. The rate of overdiagnosis, greatly affecting the true effectiveness of screening, should be further explored.

References

Abrahamsson, L., Isheden, G., Czene, K., & Humphreys, K. (2020). Continuous tumour growth models, lead time estimation and length bias in breast cancer screening studies. *Stat Methods Med Res, 29*(2), 374-395. doi:10.1177/0962280219832901

Alanko, J. (2019). The current state of breast cancer screening and new imaging opportunities. *LÄÄKETIETEELLINEN AIKAKAUSKIRJA DUODECIM, 135*(19), 1904-1911. Retrieved from https://www.duodecimlehti.fi/duo15149

Anttila, A., Lehtinen, M., Mäki, S., Leivonen, A., Heinävaara, S., & Sarkeala, T. (2021). Rintasyöpäseulonnan vuosikatsaus 2021. Retrieved from https://syoparekisteri.fi/assets/files/2021/09/Rintasyovan_seulontaohjelman_vuosikatsaus_2021.pdf

Aslam, S., & Emmanuel, P. (2010). Formulating a researchable question: A critical step for facilitating good clinical research. *Indian J Sex Transm Dis AIDS, 31*(1), 47-50. doi:10.4103/0253-7184.69003

Briggs, A., Sculpher, M., & Claxton, K. (2006). *Decision modelling for health economic evaluation*: Oup Oxford.

Broeders, M. J. M., Allgood, P., Duffy, S. W., Hofvind, S., Nagtegaal, I. D., Paci, E., . . . Bucchi, L. (2018). The impact of mammography screening programmes on incidence of advanced breast cancer in Europe: a literature review. *BMC Cancer, 18*(1), 860. doi:10.1186/s12885-018-4666-1

Cancer Society of Finland. (n.d.). Bröstcancer. Retrieved 27.04.2022 from https://www.alltomcancer.fi/information-om-cancer/cancersjukdomar/brostcancer/

Cheung, S., Hutton, J. L., & Brettschneider, J. A. (2017). *Review of sojourn time calculation models used in breast cancer screening.* University of Warwick, Retrieved from https://warwick.ac.uk/fac/sci/statistics/crism/research/17-04/17-04w.pdf

Croswell, J. M., Ransohoff, D. F., & Kramer, B. S. (2010). Principles of cancer screening: lessons from history and study design issues. *Semin Oncol, 37*(3), 202-215. doi:10.1053/j.seminoncol.2010.05.006

de Koning, H. J. (2003). Mammographic screening: evidence from randomised controlled trials. *Ann Oncol, 14*(8), 1185-1189. doi:10.1093/annonc/mdg319

Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). *Methods for the Economic Evaluation of Health Care Programmes* (4th ed.): Oxford University Press.

Duffy, S. W., Day, N. E., Tabar, L., Chen, H. H., & Smith, T. C. (1997). Markov models of breast tumor progression: some age-specific results. *J Natl Cancer Inst Monogr*(22), 93-97. doi:10.1093/jncimono/1997.22.93

Duffy, S. W., Nagtegaal, I. D., Wallis, M., Cafferty, F. H., Houssami, N., Warwick, J., . . . Lawrence, G. (2008). Correcting for lead time and length bias in estimating the effect of screen detection on cancer survival. *Am J Epidemiol, 168*(1), 98-104. doi:10.1093/aje/kwn120

Duffy, S. W., Vulkan, D., Cuckle, H., Parmar, D., Sheikh, S., Smith, R. A., . . . Moss, S. M. (2020). Effect of mammographic screening from age 40 years on breast cancer mortality (UK Age trial): final results of a randomised, controlled trial. *Lancet Oncol, 21*(9), 1165-1172. doi:10.1016/S1470-2045(20)30398-3

Edlin, R., McCabe, C., Hulme, C., Hall, P., & Wright, J. (2015). *Cost Effectiveness Modelling for Health Technology AssessmentA Practical Course* (1st ed.). doi:https://doi.org/10.1007/978-3-319-15744-3

Ekpo, E. U., Alakhras, M., & Brennan, P. (2018). Errors in Mammography Cannot be Solved Through Technology Alone. *Asian Pac J Cancer Prev, 19*(2), 291-301. doi:10.22034/APJCP.2018.19.2.291

Ernster, V. L., Ballard-Barbash, R., Barlow, W. E., Zheng, Y., Weaver, D. L., Cutter, G., . . . Geller, B. M. (2002). Detection of ductal carcinoma in situ in women undergoing screening mammography. *J Natl Cancer Inst, 94*(20), 1546-1554. doi:10.1093/jnci/94.20.1546

Ernster, V. L., & Barclay, J. (1997). Increases in ductal carcinoma in situ (DCIS) of the breast in relation to mammography: a dilemma. *J Natl Cancer Inst Monogr*(22), 151-156. doi:10.1093/jncimono/1997.22.151

Esserman, L. J., Study, W., & Athena, I. (2017). The WISDOM Study: breaking the deadlock in the breast cancer screening debate. *NPJ Breast Cancer, 3*, 34. doi:10.1038/s41523-017-0035-5

EUnetHTA. (2015). *Methods for health economic evaluations - A guideline based on current practices in Europe*. Retrieved from https://www.eunethta.eu/wp-content/uploads/2018/03/Methods_for_health_economic_evaluations.pdf

European Central Bank. (n.d.). Euro foreign exchange reference rates. Retrieved 23.05.2022 from https://www.ecb.europa.eu/stats/policy_and_exchange_rates/euro_reference_exchange_rates/html/index.en.html

European Commission Initiatives on Breast and Colorectal Cancer. (2022). European breast cancer guidelines. Screening ages and frequencies. Retrieved 14.06.2022 from https://healthcare-quality.jrc.ec.europa.eu/ecibc/european-breast-cancer-guidelines/screening-ages-and-frequencies

Fenwick, E., Steuten, L., Knies, S., Ghabri, S., Basu, A., Murray, J. F., . . . Rothery, C. (2020). Value of Information Analysis for Research Decisions-An Introduction: Report 1 of the ISPOR Value of Information Analysis Emerging Good Practices Task Force. *Value Health, 23*(2), 139-150. doi:10.1016/j.jval.2020.01.001

Finnish Cancer Registry. (n.d.-a). Cancer statistics. Retrieved 23.01.2022 from https://cancerregistry.fi/statistics/cancer-statistics/

Finnish Cancer Registry. (n.d.-b). General information on registration. Retrieved from https://syoparekisteri.fi/assets/files/2017/07/variable_list_eng_net.pdf

Finnish Cancer Registry. (n.d.-c). Rintasyövän seulonta. Retrieved 23.01.2022 from https://syoparekisteri.fi/seulonta/rintasyovanseulonta/

Finnish Cancer Registry. (n.d.-d). Statistics. Retrieved 12.02.2022 from https://cancerregister.fi/statistik/

Finnish Centre for Pensions. (2022). Effective retirement age in the earnings-related pension scheme. Retrieved 23.04.2022 from https://tilastot.etk.fi/pxweb/en/ETK/ETK__130elakkeellesiirtymisika/esiirtymisika01.px/table/tableViewLayout1/

Finnish Office for Health Care Technology Assessment. (2001). The effects of extending the use of mammography screening. A report on the cost-effectiveness of breast cancer screening in 60 to 69 year-old women. *International Network of Agencies for Health Technology*

Assessment, 2001(3). Retrieved from https://www.inahta.org/upload/Briefs_1/01-03%20FinOHTA.pdf

Glasziou, P., & Houssami, N. (2011). The evidence base for breast cancer screening. *Prev Med, 53*(3), 100-102. doi:10.1016/j.ypmed.2011.05.011

Goldie, S. J. (2003). Chapter 15: Public health policy and cost-effectiveness analysis. *J Natl Cancer Inst Monogr*(31), 102-110. doi:10.1093/oxfordjournals.jncimonographs.a003471

Hakama, M., Pukkala, E., Heikkila, M., & Kallio, M. (1997). Effectiveness of the public health policy for breast cancer screening in Finland: population based cohort study. *BMJ, 314*(7084), 864-867. doi:10.1136/bmj.314.7084.864

Hallinen, T., & Soini, E. J. (2011). The Impact of the Pharmaceutical Pricing System on Cost-Effectiveness Results: Finnish Analysis. *The Open Pharmacoeconomics & Health Economics Journal, 3*, 6-10. Retrieved from https://openpharmacoeconomicsandhealtheconomicsjournal.com/contents/volumes/V3/TOPHARMEJ-3-6/TOPHARMEJ-3-6.pdf

Haltia, O., Farkkila, N., Roine, R. P., Sintonen, H., Taari, K., Hanninen, J., . . . Saarto, T. (2018). The indirect costs of palliative care in end-stage cancer: A real-life longitudinal register- and questionnaire-based study. *Palliat Med, 32*(2), 493-499. doi:10.1177/0269216317729789

Heinävaara, S., Sarkeala, T., & Anttila, A. (2014). Overdiagnosis due to breast cancer screening: updated estimates of the Helsinki service study in Finland. *British journal of cancer, 111*(7), 1463-1468. doi:10.1038/bjc.2014.413

Heinävaara, S., Sarkeala, T., & Anttila, A. (2016). Impact of organised mammography screening on breast cancer mortality in a case-control and cohort study. *British journal of cancer, 114*(9), 1038-1044. doi:10.1038/bjc.2016.68

Heller, S. L., Plaunova, A., & Gao, Y. (2021). Ductal Carcinoma In Situ and Progression to Invasive Cancer: A Review of the Evidence. *Journal of Breast Imaging, 3*(2), 135-143. doi:10.1093/jbi/wbaa119

Helsinki University Hospital. (n.d.). Care pathaway for a breast cancer patient - Access to treatment for a breast cancer patient. Retrieved 17.05.2022 from https://www.hus.fi/en/treatments-and-examinations/care-pathway-breast-cancer-patient#access-to-treatment-for-a-brea

Holland, W. W., & Stewart, S. (2005). *Screening in Disease Prevention. What works?*(1st ed.).

Holmberg, F. (2017). Mammografi innan du fyllt 50 – onödigt eller nödvändigt? *Yleisradio (YLE)*. Retrieved 23.01.2022 from https://svenska.yle.fi/artikel/2017/03/16/mammografi-innan-du-fyllt-50onodigt-eller-nodvandigt

Hoyle, M. W., & Henley, W. (2011). Improved curve fits to summary survival data: application to economic evaluation of health technologies. *BMC Med Res Methodol, 11*, 139. doi:10.1186/1471-2288-11-139

Hunink, M. G. M., Weinstein, M. C., Wittenberg, E., Drummond, M. F., Pliskin, J. S., Wong, J. B., & Glasziou, P. P. (2014). *Decision Making in Health and Medicine: Integrating Evidence and Values* (2 ed.). Cambridge: Cambridge University Press.

International Agency for Research on Cancer. (2016). Effectiveness of Breast Cancer Screening. In *IARC Handbooks of Cancer Prevention. Breast cancer screening* (Vol. 15).

Jatoi, I. (2011). The impact of advances in treatment on the efficacy of mammography screening. *Prev Med, 53*(3), 103-104. doi:10.1016/j.ypmed.2011.06.012

Jayasekera, J., & Mandelblatt, J. S. (2020). Systematic Review of the Cost Effectiveness of Breast Cancer Prevention, Screening, and Treatment Interventions. *J Clin Oncol, 38*(4), 332-350. doi:10.1200/JCO.19.01525

Kalager, M., Adami, H. O., Bretthauer, M., & Tamimi, R. M. (2012). Overdiagnosis of invasive breast cancer due to mammography screening: results from the Norwegian screening program. *Ann Intern Med, 156*(7), 491-499. doi:10.7326/0003-4819-156-7-201204030-00005

Karihtala, P., Jääskeläinen, A., Roininen, N., & Jukkola, A. (2021). Real-world, single-centre prospective data of age at breast cancer onset: focus on survival and reproductive history. *BMJ Open, 11*(1), e041706-e041706. doi:10.1136/bmjopen-2020-041706

Kielhorn, A., & Schulenburg, J. M. (2000). *The health economics handbook*: Adis International.

Kim, D. D., Silver, M. C., Kunst, N., Cohen, J. T., Ollendorf, D. A., & Neumann, P. J. (2020). Perspective and Costing in Cost-Effectiveness Analysis, 1974–2018. *Pharmacoeconomics, 38*(10), 1135-1145. doi:10.1007/s40273-020-00942-2

Latimer, N. R., & Adler, A. I. (2022). Extrapolation beyond the end of trials to estimate long term survival and cost effectiveness. *BMJ Medicine, 1*(1), e000094. doi:10.1136/bmjmed-2021-000094

Lauby-Secretan, B., Scoccianti, C., Loomis, D., Benbrahim-Tallaa, L., Bouvard, V., Bianchini, F., & Straif, K. (2015). Breast-Cancer Screening — Viewpoint of the IARC Working Group. *N Engl J Med, 372*(24), 2353-2358. doi:10.1056/NEJMsr1504363

Lehtinen, M. (2021). Rintasyöpäseulonnan hyödyt ja haitat ovat ikäsidonnaisia. Retrieved 05.06.2022 from https://www.syopajarjestot.fi/ajankohtaista/blogit/rintasyopaseulonnan-hyodyt-ja-haitat-ovat-ikasidonnaisia/

Lehtinen, M., Heinävaara, S., Sarkeala, T., Karlsson, A., Tukiainen, M., Huovinen, R., & Anttila, A. (2019). Potilaan ikä ja syövän levinneisyys vaikuttavat rintasyövän hoitokustannuksiin. *LÄÄKETIETEELLINEN AIKAKAUSKIRJA DUODECIM, 135*, 953-960.

Leivo, T., Sintonen, H., Tuominen, R., Hakama, M., Pukkala, E., & Heinonen, O. P. (1999). The cost-effectiveness of nationwide breast carcinoma screening in Finland, 1987-1992. *Cancer, 86*(4), 638-646. doi:10.1002/(sici)1097-0142(19990815)86:4<638::aid-cncr12>3.0.co;2-h

Loberg, M., Lousdal, M. L., Bretthauer, M., & Kalager, M. (2015). Benefits and harms of mammography screening. *Breast Cancer Res, 17*, 63. doi:10.1186/s13058-015-0525-z

Mäklin, S., & Kokko, P. (2021). Terveyden- ja sosiaalihuollon yksikkökustannukset Suomessa vuonna 2017. Retrieved from https://www.julkari.fi/handle/10024/142882

Marmot, M. G., Altman, D. G., Cameron, D. A., Dewar, J. A., Thompson, S. G., & Wilcox, M. (2013). The benefits and harms of breast cancer screening: an independent review. *British journal of cancer, 108*(11), 2205-2240. doi:10.1038/bjc.2013.177

339, (2011).

Mittmann, N., Stout, N. K., Lee, P., Tosteson, A. N., Trentham-Dietz, A., Alagoz, O., & Yaffe, M. J. (2015). Total cost-effectiveness of mammography screening strategies. *Health Rep, 26*(12), 16-25.

Official Statistics Finland. (2020a). Deaths by underlying cause of death (time series classification), age and gender, 1969-2020. Retrieved 03.03.2022 from https://pxnet2.stat.fi/PXWeb/pxweb/en/StatFin/StatFin__ter__ksyyt/statfin_ksyyt_pxt_11bs.px/

Official Statistics Finland. (2020b). Population according to age (5-year) and sex, 1865-2021. Retrieved 03.03.2022 from https://pxnet2.stat.fi/PXWeb/pxweb/en/StatFin/StatFin__vrm__vaerak/statfin_vaerak_pxt_11rc.px/

Official Statistics Finland. (2021). Consumer Price Indices, overall index, yearly data, 1972-2021. Retrieved 09.03.2022 from https://pxnet2.stat.fi/PXWeb/pxweb/en/StatFin/StatFin__hin__khi__vv/statfin_khi_pxt_11xt.px/table/tableViewLayout1/

Official Statistics Finland. (2022a). Gross domestic product per capita. Retrieved 17.05.2022 from https://www.stat.fi/tup/suoluk/suoluk_kansantalous_en.html

Official Statistics Finland. (2022b). Life table by age and sex, 1986-2020. Retrieved 20.05.2022 from pxnet2.stat.fi/PXWeb/pxweb/en/StatFin/StatFin__vrm__kuol/statfin_kuol_pxt_12ap.px/

Ooi, S. L., Martinez, M. E., & Li, C. I. (2011). Disparities in breast cancer characteristics and outcomes by race/ethnicity. *Breast Cancer Res Treat*, *127*(3), 729-738. doi:10.1007/s10549-010-1191-6

Parikh, R., Mathai, A., Parikh, S., Sekhar, G., C., & Thomas, R. (2008). Understanding and using sensitivity, specificity and predictive values. *Indian Journal of Opthalmology, 56*(2008 Jan-Feb), 45-50.

Parise, C. A., Bauer, K. R., & Caggiano, V. (2010). Variation in breast cancer subtypes with age and race/ethnicity. *Crit Rev Oncol Hematol, 76*(1), 44-52. doi:10.1016/j.critrevonc.2009.09.002

Parvinen, I. (2014). *The Effects of the Breast Cancer Mammography Screening Programme in Women aged 40 to 84 years in Turku, Finland (1987-2009).* University of Turku, Turku. Retrieved from https://www.utupub.fi/bitstream/handle/10024/100946/AnnalesD1142Parvinen.pdf?sequence=2

Parvinen, I., Helenius, H., Pylkkänen, L., Anttila, A., Immonen-Räihä, P., L., K., . . . Klemi, P. J. (2006). Service screening mammography reduces breast cancer mortality among elderly women in Turku. *J Med Screen, 13*(1), 34-40. doi:10.1258/096914106776179845

Pataky, R., Ismail, Z., Coldman, A. J., Elwood, M., Gelmon, K., Hedden, L., . . . Peacock, S. (2014). Cost-effectiveness of annual versus biennial screening mammography for women with high mammographic breast density. *J Med Screen, 21*(4), 180-188. doi:10.1177/0969141314549758

Pharmaceuticals Pricing Board. (2019). *PREPARING A HEALTH ECONOMIC EVALUATION TO BE ATTACHED TO THE APPLICATION FOR REIMBURSEMENT STATUS AND WHOLESALE PRICE FOR A MEDICINAL PRODUCT.* Retrieved from https://www.hila.fi/content/uploads/2020/01/Instructions_TTS_2019.pdf

Pitkäniemi, J., Malila, N., Tanskanen, T., Degerlund, H., Heikkinen, S., & Seppä, K. (2022). *Syöpä 2020. Tilastoraportti Suomen syöpätilanteesta*. Retrieved from https://syoparekisteri.fi/assets/files/2022/06/Syopa-2020-raportti_fin.pdf

Puliti, D., Duffy, S. W., Miccinesi, G., de Koning, H., Lynge, E., Zappa, M., . . . Group, E. W. (2012). Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. *J Med Screen, 19 Suppl 1*, 42-56. doi:10.1258/jms.2012.012082

Rautalin, M., Farkkila, N., Sintonen, H., Saarto, T., Taari, K., Jahkola, T., & Roine, R. P. (2018). Health-related quality of life in different states of breast cancer - comparing different instruments. *Acta Oncol, 57*(5), 622-628. doi:10.1080/0284186X.2017.1400683

Rijnsburger, A. J., Essink-Bot, M. L., van Dooren, S., Borsboom, G. J., Seynaeve, C., Bartels, C. C., . . . de Koning, H. J. (2004). Impact of screening for breast cancer in high-risk women on health-related quality of life. *British journal of cancer*, *91*(1), 69-76. doi:10.1038/sj.bjc.6601912

Rim, S. H., Allaire, B. T., Ekwueme, D. U., Miller, J. W., Subramanian, S., Hall, I. J., & Hoerger, T. J. (2019). Cost-effectiveness of breast cancer screening in the National Breast and Cervical Cancer Early Detection Program. *Cancer Causes Control, 30*(8), 819-826. doi:10.1007/s10552-019-01178-y

Rintasyöpä.fi. (2021). Rintasyövän ennuste ja seuranta. Retrieved 19.05.2022 from https://rintasyopa.fi/tietoa-rintasyovasta/rintasyovan-ennuste-ja-seuranta/

Roine, E., Farkkila, N., Sintonen, H., Taari, K., Roine, R. P., & Saarto, T. (2019). Costs in Different States of Breast Cancer. *Anticancer Res*, *39*(1), 353-359. doi:10.21873/anticanres.13119

Saarenmaa, I., Salminen, T., Geiger, U., Heikkinen, P., Hyvarinen, S., Isola, J., . . . Haka, M. (2001). The effect of age and density of the breast on the sensitivity of breast cancer diagnostic by mammography and ultasonography. *Breast Cancer Res Treat, 67*(2), 117-123. doi:10.1023/a:1010627527026

Schiller-Fruhwirth, I. C., Jahn, B., Arvandi, M., & Siebert, U. (2017). Cost-Effectiveness Models in Breast Cancer Screening in the General Population: A Systematic Review. *Appl Health Econ Health Policy, 15*(3), 333-351. doi:10.1007/s40258-017-0312-3

Schunemann, H. J., Lerda, D., Quinn, C., Follmann, M., Alonso-Coello, P., Rossi, P. G., . . . European Commission Initiative on Breast Cancer Contributor, G. (2020). Breast Cancer Screening and Diagnosis: A Synopsis of the European Breast Guidelines. *Ann Intern Med, 172*(1), 46-56. doi:10.7326/M19-2125

Siebert, U., Alagoz, O., Bayoumi, A. M., Jahn, B., Owens, D. K., Cohen, D. J., & Kuntz, K. M. (2012). State-transition modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-3. *Med Decis Making, 32*(5), 690-700. doi:10.1177/0272989X12455463

Smith, R. A., Duffy, S. W., & Tabar, L. (2012). Breast cancer screening: the evolving evidence. *Oncology (Williston Park), 26*(5), 471-475, 479-481, 485-476.

Smith, R. A., Mettlin, C. J., & Eyre, H. (2003). Methodologic Issues in the Evaluation of Early Detection Programs. In D. W. Kufe, R. E. Pollock, R. R. Weichselbaum, R. C. Bast, T. S. Gansler, J. F. Holland, & E. Frei (Eds.), *Holland-Frei Cancer Medicine* (6th ed.).

Swedish National Board of Health and Welfare. (2019). *Värdet av populationsbaserad screening för bröstcancer*. Retrieved from https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/dokument-webb/nationella-screeningprogram/screening-brostcancer-halsoekonomi-analys.pdf

Strong, M., Oakley, J. E., & Brennan, A. (2014). Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample: a

nonparametric regression approach. *Med Decis Making, 34*(3), 311-326. doi:10.1177/0272989X13505910

Tan, K. H., Simonella, L., Wee, H. L., Roellin, A., Lim, Y. W., Lim, W. Y., . . . Cook, A. R. (2013). Quantifying the natural history of breast cancer. *Br J Cancer, 109*(8), 2035-2043. doi:10.1038/bjc.2013.471

Torkki, P., Leskela, R. L., Linna, M., Maklin, S., Mecklin, J. P., Bono, P., . . . Karjalainen, S. (2018). Cancer costs and outcomes for common cancer sites in the Finnish population between 2009-2014. *Acta Oncol, 57*(7), 983-988. doi:10.1080/0284186X.2018.1438656

van den Ende, C., Oordt-Speets, A. M., Vroling, H., & van Agt, H. M. E. (2017). Benefits and harms of breast cancer screening with mammography in women aged 40-49 years: A systematic review. *Int J Cancer, 141*(7), 1295-1306. doi:10.1002/ijc.30794

van Luijt, P. A., Heijnsdijk, E. A., & de Koning, H. J. (2017). Cost-effectiveness of the Norwegian breast cancer screening program. *Int J Cancer, 140*(4), 833-840. doi:10.1002/ijc.30513

van Oortmarssen, G. J., Habbema, J. D., van der Maas, P. J., de Koning, H. J., Collette, H. J., Verbeek, A. L., . . . Lubbe, K. T. (1990). A model for breast cancer screening. *Cancer, 66*(7), 1601-1612. doi:10.1002/1097-0142(19901001)66:7<1601::aid-cncr2820660727>3.0.co;2-o

Vanni, T., Karnon, J., Madan, J., White, R. G., Edmunds, W. J., Foss, A. M., & Legood, R. (2011). Calibrating models in economic evaluation: a seven-step approach. *Pharmacoeconomics, 29*(1), 35-49. doi:10.2165/11584600-000000000-00000

Vehmanen, L. (2020). Rintasyövän toteaminen, alatyypit ja ennuste. Retrieved from https://www.terveyskirjasto.fi/dlk00818

Venkatesh, S. L., Oseni, T. O., & Bahl, M. (2021). Symptomatic ductal carcinoma in situ (DCIS): Upstaging risk and predictors. *Clin Imaging, 73*, 101-107. doi:10.1016/j.clinimag.2020.11.050

VSSHP. (2017). T11 Kuvanatamisen toimialue - Hinnasto 2018. Retrieved 12.05.2022 from https://docplayer.fi/68998741-T11-kuvantamisen-toimialue-hinnasto-2018.html

Weber, R. J., Klompenhouwer, E. G., Voogd, A. C., Strobbe, L. J., Broeders, M. J., & Duijm, L. E. (2015). Comparison of the diagnostic workup of women referred at non-blinded or blinded double reading in a population-based screening mammography programme in the south of the Netherlands. *British journal of cancer*, *113*(7), 1094-1098. doi:10.1038/bjc.2015.295

Wilson, J. M., & Jungner, Y. G. (1968). Principles and practice of mass screening for disease. *Bol Oficina Sanit Panam, 65*(4), 281-393.

World Bank. (2020). Global Economic Monitor (GEM). Retrieved 12.05.2022 from https://databank.worldbank.org/source/global-economic-monitor-(gem)/preview/on

World Health Organization. (2021). Breast cancer. Retrieved 23.01.2022 from https://www.who.int/news-room/fact-sheets/detail/breast-cancer

Wu, J. C., Hakama, M., Anttila, A., Yen, A. M., Malila, N., Sarkeala, T., . . . Chen, H. H. (2010). Estimation of natural history parameters of breast cancer based on non-randomized organized screening data: subsidiary analysis of effects of inter-screening interval, sensitivity, and attendance rate on reduction of advanced cancer. *Breast Cancer Res Treat, 122*(2), 553-566. doi:10.1007/s10549-009-0701-x

Wu, X., Baig, A., Kasymjanova, G., Kafi, K., Holcroft, C., Mekouar, H., . . . Muanza, T. (2016). Pattern of Local Recurrence and Distant Metastasis in Breast Cancer By Molecular Subtype. *Cureus, 8*(12), e924. doi:10.7759/cureus.924

Yen, M. F., Tabar, L., Vitak, B., Smith, R. A., Chen, H. H., & Duffy, S. W. (2003). Quantifying the potential problem of overdiagnosis of ductal carcinoma in situ in breast cancer screening. *Eur J Cancer, 39*(12), 1746-1754. doi:10.1016/s0959-8049(03)00260-0

Appendix A

A1. Literature search on the cost-effectiveness of organized breast cancer screening

Searches were conducted on the Google Scholar and PubMed databases using the following search strategy: (cost-effectiveness analysis) OR (cost-utility analysis) OR (CEA) OR (CUA) AND (breast cancer screening) OR (mammography) AND (Finland). Additional searches were conducted in Finnish and Swedish. Roughly 40 000 hits were generated without "Finland".

A2. Recurrence rates and NLBC stage distribution

Table A.2.1

Age-specific NLBC stage distribution.

Age group	Value (share)	Source
40 – 49		
Lymph node positive	0,84	Lehtinen et al. 2019
Unknown spread	0,14	Lehtinen et al. 2019
Metastasised	0,03	Lehtinen et al. 2019
50 – 54		
Lymph node positive	0,83	Lehtinen et al. 2019
Unknown spread	0,14	Lehtinen et al. 2019
Metastasised	0,03	Lehtinen et al. 2019
55 – 59		
Lymph node positive	0,85	Lehtinen et al. 2019
Unknown spread	0,11	Lehtinen et al. 2019
Metastasised	0,04	Lehtinen et al. 2019
60 – 64		
Lymph node positive	0,81	Lehtinen et al. 2019
Unknown spread	0,16	Lehtinen et al. 2019
Metastasised	0,02	Lehtinen et al. 2019
65 – 69		
Lymph node positive	0,72	Lehtinen et al. 2019
Unknown spread	0,28	Lehtinen et al. 2019
Metastasised	0,00	Lehtinen et al. 2019
70 – 74		
Lymph node positive	0,71	Lehtinen et al. 2019
Unknown spread	0,27	Lehtinen et al. 2019
Metastasised	0,02	Lehtinen et al. 2019
≥75		
Lymph node positive	0,52	Lehtinen et al. 2019
Unknown spread	0,45	Lehtinen et al. 2019
Metastasised	0,04	Lehtinen et al. 2019

Table A.2.2

Age- and stage specific recurrence rates.

Age group	Probability	Source
40 – 49		
Local	0,3	Lehtinen et al. 2019
Unknown spread	0,3	Lehtinen et al. 2019
Lymph node positive	0,5	Lehtinen et al. 2019
50 – 69		
Local	0,2	Lehtinen et al. 2019
Unknown spread	0,2	Lehtinen et al. 2019
Lymph node positive	0,4	Lehtinen et al. 2019
≥70		
Local	0,25	Lehtinen et al. 2019
Unknown spread	0,25	Lehtinen et al. 2019
Lymph node positive	0,4	Lehtinen et al. 2019

Appendix B

B1. Natural history parameter literature search and synthetization

In searching for relevant natural history parameter estimates, searches conducted on the PubMed and Google Scholar databases using combinations of the words “natural history model”, “breast cancer” and “Finland” yielded roughly 49 000 hits. Among the top results was a study by Wu et al. (2010) which estimated IBC natural history parameters for women in the Pirkanmaa hospital district of Finland. Removing the word “Finland” from the search algorithm increased the number of hits to over 1.5 million, calling for more targeted search criteria. After adding the term “age-specific” to the algorithm, the number of hits were limited to roughly 30 000. In determining the search results’ relevance for the analysis, the population was given less importance in favor of the model structure adopted for estimating the natural history parameters. As expected, an identical model structure to Wu et al. (2010) was not found due to the FCR’s unique tumor stage classification. A study by Duffy et al. (1997) estimated transition probabilities for the age groups 40–49, 50–59 and 60–69 using a five-state Markov-chain model according to lymph node status. The unadjusted underlying incidence rate for LBC for the age group 50–59 was retrieved from Duffy et al. (1997) instead of Wu et al. (2010), as it fit the increasing function between incidence and age better.

Searching for relevant transition parameters for the natural history of DCIS was carried out using the same search algorithm as for that of invasive breast cancer, only changing the search term “breast cancer” to “DCIS”. Again, the relevance of the search results depended more on the compatibility of the model structure used for estimating the natural history parameters with the model structure used by Wu et al. (2010). No relevant studies conducted in the Finnish population were found. Excluding the word “Finland” yielded approximately 11 000 hits, among

which a study by Yen et al. (2003) was deemed relevant. Similar to Duffy et al. (1997), the study was based on the Swedish Two-County Trial screening data, to which a Markov process model was fitted in order to estimate age-specific (i.e., including age groups 40–49, 50–59 and 60–69) natural history transition rates for entry into and exit from non-progressive and progressive DCIS, respectively. The underlying incidence parameters for IBC and DCIS were adjusted downwards according to the modelled proportion of IBCs preceded by progressive DCIS, in this model assumed to be 18%, the same proportion applied in a version of the MISCAN model (van Oortmarssen et al., 1990).

The assumption that a larger proportion of pDCIS should end up in preclinical LBC was based on the pattern of the estimated underlying incidence rates, presented in Table B1, with respect to the proportion of transitions into preclinical LBC within a cycle, compared to that of NLBC. Accordingly, the proportions of LBC and NLBC preceded by DCIS assumed a higher proportion (50%–100%) were LBC and a lower proportion (0%–50%) were NLBC

For ages 70 an over, parameters guiding the underlying incidence rate were assigned lower bound measures of variation equal to the calibrated value of the corresponding parameters for ages 60–69. Due to its inconsistent variation with age, the lower bound for the transition rate from pDCIS to preclinical IBC was derived the same way. To sufficiently allow for the underlying incidence to increase with age, the parameters were assigned upper bound measures of variation equal to the corresponding parameters' calibrated value multiplied by a factor of 2. Other transition probabilities for ages ≥70 were assigned upper bounds equal to the calibrated value of the corresponding parameter for ages 60–69 and lower bounds equal to the same value multiplied by a factor of 0.5, reflecting the decreasing rate of tumor progression with age.

Table B1

Natural history parameter calibration inputs.

Transition rates 40 – 49	**Deterministic**		**Lower bound CI 95%**	**Upper bound CI 95%**	**Source**
Healthy to Preclinical non-progresssive DCIS	0,0000072		0,000001	0,0000039	Yen et al. (2003)
Adjusted for progressive DCIS precedent	0,0000013		0,00000018	0,0000007	
Healthy to Preclinical progresssive DCIS	0,0015		0,0013	0,0021	Yen et al. (2003)
Adjusted for progressive DCIS precedent	0,00027		0,00023	0,00038	
Preclinical progressive DCIS to Preclinical invasive BC	4,93		1,24	8,61	Yen et al. (2003)
Transition probabilities 40 – 49	**Deterministic**	**Distribution**	**alpha**	**beta**	**Source**
Healthy to Preclinical LBC	0,0009				Duffy et al. (1997)
Adjusted for progressive DCIS precedent	0,00074	Beta	400	541205	
Healthy to Preclinical NLBC	0,00009				Duffy et al. (1997)
Adjusted for progressive DCIS precedent	0,00007	Beta	400	5419253	
Healthy to Clinical LBC	0,00014	Beta	400	2856342	Duffy et al. (1997)
Healthy to Clinical NLBC	0,00008	Beta	400	4999199	Duffy et al. (1997)
Preclinical non-progresssive DCIS to recovery	0,082	Beta	367	4102	Yen et al. (2003)
Preclinical progressive DCIS to Preclinical progressive DCIS	0,0072	Dirichlet	99,3	13637,7	Yen et al. (2003)
	Deterministic	**Distribution**	**Lower bound**	**Upper bound**	
Preclinical progressive DCIS to Preclinical LBC	-	Dirichlet	0,5	1,00	
Preclinical progressive DCIS to Preclinical NLBC	-	Dirichlet	0,0	0,5	
	Deterministic	**Distribution**	**alpha**	**beta**	
Preclinical LBC to Preclinical LBC	0,54	Dirichlet	183	156	Duffy et al. (1997)
Preclinical LBC to Preclinical NLBC	0,1	Dirichlet	360	3239	Duffy et al. (1997)
Preclinical LBC to Clinical LBC	0,2	Dirichlet	320	1279	Duffy et al. (1997)
Preclinical LBC to Clinical NLBC	0,16	Dirichlet	336	1763	Duffy et al. (1997)
Preclinical NLBC to Preclinical NLBC	0,12	Beta	352	2580	Duffy et al. (1997)
Preclinical NLBC to Clinical NLBC	0,88	Beta	47	6	Duffy et al. (1997)
Transition rates 50 – 59	**Deterministic**		**Lower bound CI 95%**	**Upper bound CI 95%**	**Source**
Healthy to Preclinical non-progresssive DCIS	0,00001		0,000000	0,000035	Yen et al. (2003)
Adjusted for progressive DCIS precedent	0,0000018		0,000001	0,000006	
Healthy to Preclinical progresssive DCIS	0,0015		0,0013	0,0018	Yen et al. (2003)
Adjusted for progressive DCIS precedent	0,00027		0,00023	0,00032	
Preclinical progressive DCIS to Preclinical invasive BC	2,99		1,22	4,76	Yen et al. (2003)
Transition probabilities 50 – 59	**Deterministic**	**Distribution**	**alpha**	**beta**	**Source**
Healthy to Preclinical LBC	0,0014				Duffy et al. (1997)
Adjusted for progressive DCIS precedent	0,0011	Beta	400	347632	
Healthy to Preclinical NLBC	0,00014				Wu et al. (2010)
Adjusted for progressive DCIS precedent	0,00011	Beta	400	3483520	
Healthy to Clinical LBC	0,00014	Beta	400	2856342	Duffy et al. (1997)
Healthy to Clinical NLBC	0,00004	Beta	400	9999199	Duffy et al. (1997)
Preclinical non-progresssive DCIS to recovery	0,0598	Beta	376	5908	Yen et al. (2003)
Preclinical progressive DCIS to Preclinical progressive DCIS	0,0503	Dirichlet	380	7173	Yen et al. (2003)
	Deterministic	**Distribution**	**Lower bound**	**Upper bound**	
Preclinical progressive DCIS to Preclinical LBC	-	Dirichlet	0,50	1,00	
Preclinical progressive DCIS to Preclinical NLBC	-	Dirichlet	0,00	0,50	
	Deterministic	**Distribution**	**alpha**	**beta**	
Preclinical LBC to Preclinical LBC	0,5343	Dirichlet	186	162	Wu et al. (2010)
Preclinical LBC to Preclinical NLBC	0,1357	Dirichlet	346	2201	Wu et al. (2010)
Preclinical LBC to Clinical LBC	0,2152	Dirichlet	314	1144	Wu et al. (2010)
Preclinical LBC to Clinical NLBC	0,1148	Dirichlet	354	2729	Wu et al. (2010)
Preclinical NLBC to Preclinical NLBC	0,2943	Beta	282	676	Wu et al. (2010)
Preclinical NLBC to Clinical NLBC	0,7057	Beta	117	49	Wu et al. (2010)
Transition rates 60 – 69	**Deterministic**		**Lower bound CI 95%**	**Upper bound CI 95%**	**Source**
Healthy to Preclinical non-progresssive DCIS	0,000012		0,0000015	0,000026	Yen et al. (2003)
Adjusted for progressive DCIS precedent	0,000002		0,0000003	0,000005	
Healthy to Preclinical progresssive DCIS	0,0027		0,002	0,003	Yen et al. (2003)
Adjusted for progressive DCIS precedent	0,0005		0,000	0,001	
Preclinical progressive DCIS to Preclinical invasive BC	6,13		1,92	10,33	Yen et al. (2003)

Transition probabilities 60 – 69	Deterministic	Distribution	alpha	beta	Source
Healthy to Preclinical LBC	0,0020				Duffy et al. (1997)
Adjusted for progressive DCIS precedent	0,0016	Beta	399	243102	
Healthy to Preclinical NLBC	0,00017				Duffy et al. (1997)
Adjusted for progressive DCIS precedent	0,00014	Beta	400	2868640	
Healthy to Clinical LBC	0,00018	Beta	400	2221421	Duffy et al. (1997)
Healthy to Clinical NLBC	0,00005	Beta	400	7999199	Duffy et al. (1997)
Preclinical non-progresssive DCIS to recovery	0,0027	Beta	399	148654	Yen et al. (2003)
Preclinical progressive DCIS to Preclinical progressive DCIS	0,0022	Dirichlet	399	182974	Yen et al. (2003)
	Deterministic	**Distribution**	**Lower bound**	**Upper bound**	
Preclinical progressive DCIS to Preclinical LBC	-	Dirichlet	0,50	1,00	
Preclinical progressive DCIS to Preclinical NLBC	-	Dirichlet	0,00	0,50	
	Deterministic	**Distribution**	**alpha**	**beta**	
Preclinical LBC to Preclinical LBC	0,71	Dirichlet	115	47	Duffy et al. (1997)
Preclinical LBC to Preclinical NLBC	0,11	Dirichlet	356	2879	Duffy et al. (1997)
Preclinical LBC to Clinical LBC	0,13	Dirichlet	348	2328	Duffy et al. (1997)
Preclinical LBC to Clinical NLBC	0,05	Dirichlet	380	7219	Duffy et al. (1997)
Preclinical NLBC to Preclinical NLBC	0,46	Beta	216	253	Duffy et al. (1997)
Preclinical NLBC to Clinical NLBC	0,54	Beta	183	156	Duffy et al. (1997)

Note. Alpha and beta parameters were estimated with a standard error of fiver percent, for parameter inputs assigned Beta or Dirichlet distributions.

B2. Screening-specific parameter synthetization

Measures of variation were applied directly from parameter estimate sources where available. Sensitivity parameters' upper bound variation measures were set equal to the deterministic value of older age groups' corresponding parameters, whereas lower bounds were set equal to the deterministic value of younger age groups, reflecting the pattern of increasing sensitivity with age. Specificity parameter's lower bound measures of variation were set equal to the deterministic value of younger age groups' corresponding parameters, whereas upper bounds were assumed to be 1 for age groups older than 50–59, following Wu et al. (2010).

Table B2

Screening-specific parameters.

Mammography accuracy 40 – 49	**Deterministic**	**Lower CI 95%**	**Upper CI 95%**	**Source**
IBC sensitivity	0,83	0,76	0,91	Duffy et al. (1997)
	Deterministic	**Lower bound**	**Upper bound**	
IBC specificity	0,92	0,85	0,9997	Duffy et al. (1997)
DCIS sensitivity	0,95	0,87	0,98	Assumption
DCIS specificity	0,92	0,85	0,9997	Assumption
Mammography accuracy 50 – 59	**Deterministic**	**Lower CI 95%**	**Upper CI 95%**	
IBC sensitivity	0,8483	0,7488	0,9479	Wu et al. (2010)
IBC specificity	0,9997	0,9989	1	Wu et al. (2010)
	Deterministic	**Lower bound**	**Upper bound**	
DCIS sensitivity	0,9755	0,9545	0,9971	Assumption
DCIS specificity	0,9997	0,9989	1	Assumption
Mammography accuracy 60 – 69	**Deterministic**	**Lower bound**	**Upper bound**	
IBC sensitivity	0,8670	0,85	0,89	Assumption
IBC specificity	0,9999	0,9997	1	Assumption
DCIS sensitivity	0,9971	0,9755	1	Assumption
DCIS specificity	0,99985	0,9997	1	Assumption
Mammography accuracy 70 – 74	**Deterministic**	**Lower bound**	**Upper bound**	
IBC sensitivity	0,8861	0,8670	1	Assumption
IBC specificity	0,9999	0,9999	1	Assumption
DCIS sensitivity	1	0,9971	1	Assumption
DCIS specificity	0,99993	0,99985	1	Assumption

B3. Extrapolated breast cancer specific survival goodness-of-fit

Figure B3.1

Weibull specification fitted to EBCSS curve for ages 41–69.

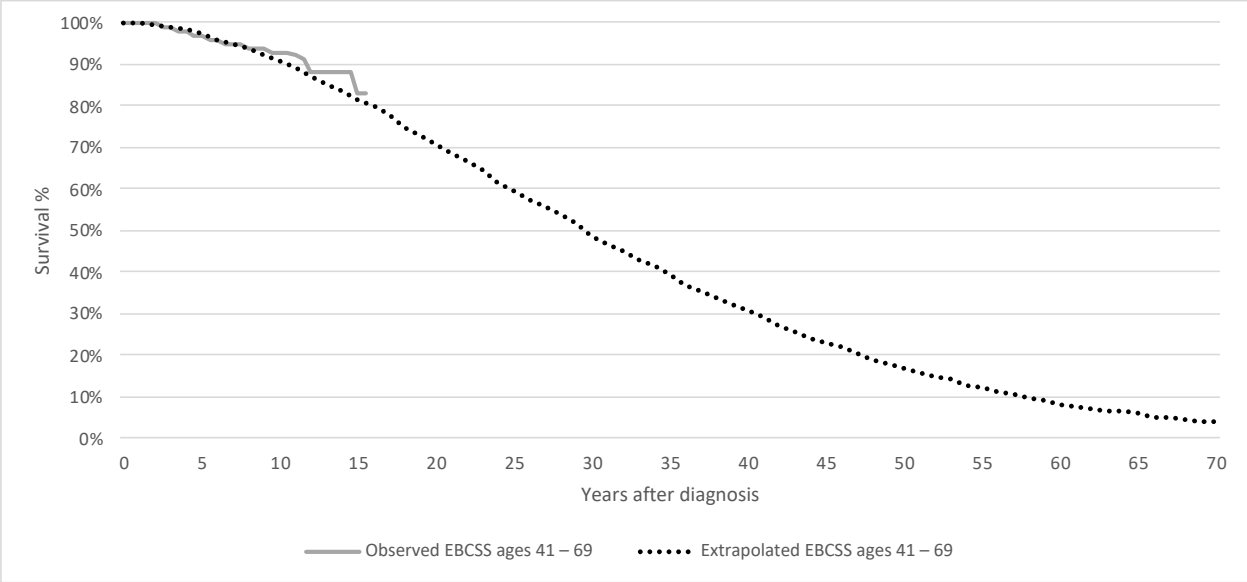


Figure B3.2

Weibull specification fitted to EBCSS curve for ages 70 and over.

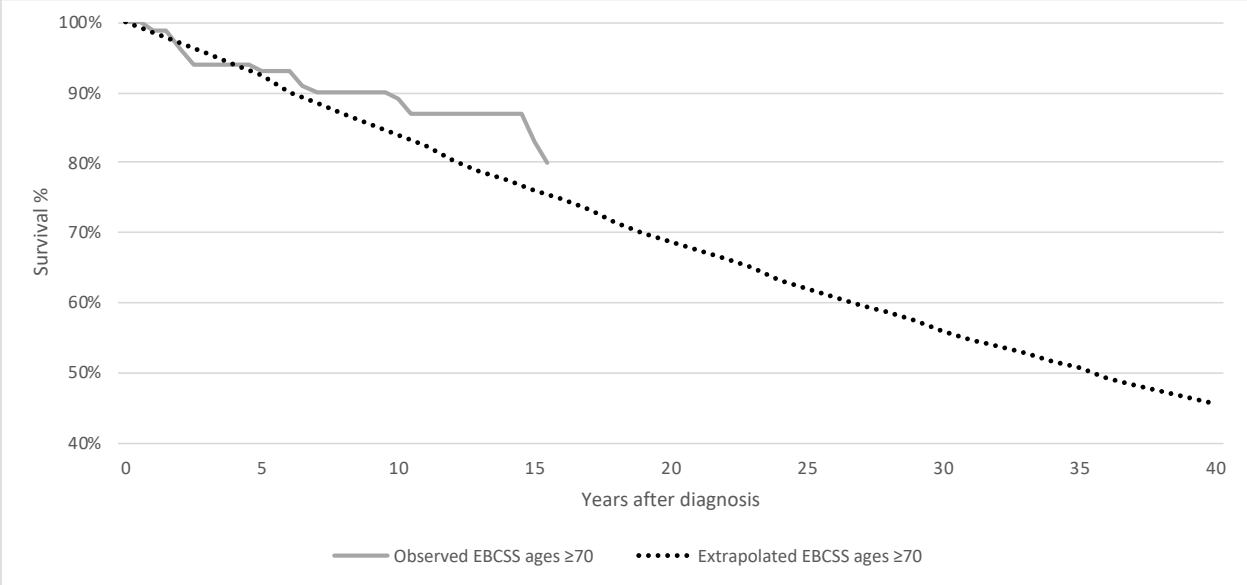
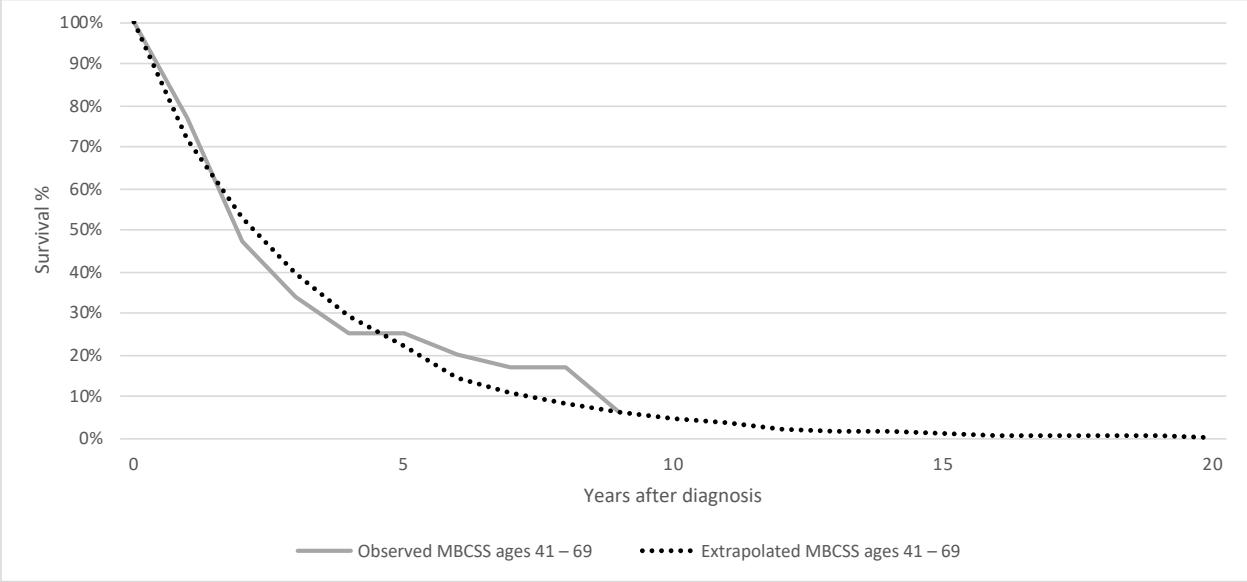


Figure B3.3

Weibull specification fitted to MBCSS curve for ages 41–69.



B4. Cost and utility parameter literature search and synthetization

Table B4.1

Direct healthcare costs.

BC treatment	Deterministic	Distribution	alpha	beta	Source
1st year after diagnosis					
DCIS					
<50	15220	Gamma	400	38	Lehtinen et al. (2019)
50-54	13066	Gamma	400	33	Lehtinen et al. (2019)
55-59	13928	Gamma	400	35	Lehtinen et al. (2019)
60-64	11547	Gamma	400	29	Lehtinen et al. (2019)
65-69	11770	Gamma	400	29	Lehtinen et al. (2019)
70-74	9430	Gamma	400	24	Lehtinen et al. (2019)
≥75	5000	Gamma	400	13	Lehtinen et al. (2019)
Local BC					
<50	20328	Gamma	400	51	Lehtinen et al. (2019)
50-54	18166	Gamma	400	45	Lehtinen et al. (2019)
55-59	18925	Gamma	400	47	Lehtinen et al. (2019)
60-64	16706	Gamma	400	42	Lehtinen et al. (2019)
65-69	16826	Gamma	400	42	Lehtinen et al. (2019)
70-74	14706	Gamma	400	37	Lehtinen et al. (2019)
≥75	10015	Gamma	400	25	Lehtinen et al. (2019)
Node positive BC					
<50	28384	Gamma	400	71	Lehtinen et al. (2019)
50-54	26326	Gamma	400	66	Lehtinen et al. (2019)
55-59	27045	Gamma	400	68	Lehtinen et al. (2019)
60-64	24801	Gamma	400	62	Lehtinen et al. (2019)
65-69	24877	Gamma	400	62	Lehtinen et al. (2019)
70-74	22586	Gamma	400	56	Lehtinen et al. (2019)
≥75	18119	Gamma	400	45	Lehtinen et al. (2019)
Metastatic BC					
<50	33332	Gamma	400	83	Lehtinen et al. (2019)
50-54	31110	Gamma	400	78	Lehtinen et al. (2019)
55-59	31981	Gamma	400	80	Lehtinen et al. (2019)
60-64	29601	Gamma	400	74	Lehtinen et al. (2019)
65-69	29835	Gamma	400	75	Lehtinen et al. (2019)
70-74	27571	Gamma	400	69	Lehtinen et al. (2019)
≥75	23162	Gamma	400	58	Lehtinen et al. (2019)
Stable					
DCIS					
<50	1998	Gamma	400	5	Lehtinen et al. (2019)
50-54	1636	Gamma	400	4	Lehtinen et al. (2019)
55-59	1336	Gamma	400	3	Lehtinen et al. (2019)
60-64	1130	Gamma	400	3	Lehtinen et al. (2019)
65-69	1334	Gamma	400	3	Lehtinen et al. (2019)
70-74	1136	Gamma	400	3	Lehtinen et al. (2019)
≥75	758	Gamma	400	2	Lehtinen et al. (2019)
Local BC					
<50	2405	Gamma	400	6	Lehtinen et al. (2019)
50-54	2030	Gamma	400	5	Lehtinen et al. (2019)
55-59	1750	Gamma	400	4	Lehtinen et al. (2019)
60-64	1539	Gamma	400	4	Lehtinen et al. (2019)
65-69	1813	Gamma	400	5	Lehtinen et al. (2019)
70-74	1561	Gamma	400	4	Lehtinen et al. (2019)
≥75	2405	Gamma	400	6	Lehtinen et al. (2019)

Node positive BC					
<50	3294	Gamma	400	8	Lehtinen et al. (2019)
50-54	2929	Gamma	400	7	Lehtinen et al. (2019)
55-59	2639	Gamma	400	7	Lehtinen et al. (2019)
60-64	2453	Gamma	400	6	Lehtinen et al. (2019)
65-69	2624	Gamma	400	7	Lehtinen et al. (2019)
70-74	2471	Gamma	400	6	Lehtinen et al. (2019)
≥75	2069	Gamma	400	5	Lehtinen et al. (2019)
Metastatic BC					
<50	6407	Gamma	400	16	Lehtinen et al. (2019)
50-54	6041	Gamma	400	15	Lehtinen et al. (2019)
55-59	5772	Gamma	400	14	Lehtinen et al. (2019)
60-64	5527	Gamma	400	14	Lehtinen et al. (2019)
65-69	5745	Gamma	400	14	Lehtinen et al. (2019)
70-74	5579	Gamma	400	14	Lehtinen et al. (2019)
≥75	5181	Gamma	400	13	Lehtinen et al. (2019)
Last year before death					
DCIS					
<50	29483	Gamma	400	74	Lehtinen et al. (2019)
50-54	23144	Gamma	400	58	Lehtinen et al. (2019)
55-59	19059	Gamma	400	48	Lehtinen et al. (2019)
60-64	16230	Gamma	400	41	Lehtinen et al. (2019)
65-69	17564	Gamma	400	44	Lehtinen et al. (2019)
70-74	8427	Gamma	400	21	Lehtinen et al. (2019)
≥75	4083	Gamma	400	10	Lehtinen et al. (2019)
Local BC					
<50	32958	Gamma	400	82	Lehtinen et al. (2019)
50-54	27089	Gamma	400	68	Lehtinen et al. (2019)
55-59	22728	Gamma	400	57	Lehtinen et al. (2019)
60-64	19818	Gamma	400	50	Lehtinen et al. (2019)
65-69	21316	Gamma	400	53	Lehtinen et al. (2019)
70-74	12058	Gamma	400	30	Lehtinen et al. (2019)
≥75	7829	Gamma	400	20	Lehtinen et al. (2019)
Node positive BC					
<50	38493	Gamma	400	96	Lehtinen et al. (2019)
50-54	32578	Gamma	400	81	Lehtinen et al. (2019)
55-59	28554	Gamma	400	71	Lehtinen et al. (2019)
60-64	25336	Gamma	400	63	Lehtinen et al. (2019)
65-69	26817	Gamma	400	67	Lehtinen et al. (2019)
70-74	17787	Gamma	400	44	Lehtinen et al. (2019)
≥75	13456	Gamma	400	34	Lehtinen et al. (2019)
Metastatic BC					
<50	55296	Gamma	400	138	Lehtinen et al. (2019)
50-54	49207	Gamma	400	123	Lehtinen et al. (2019)
55-59	44881	Gamma	400	112	Lehtinen et al. (2019)
60-64	41999	Gamma	400	105	Lehtinen et al. (2019)
65-69	43575	Gamma	400	109	Lehtinen et al. (2019)
70-74	34297	Gamma	400	86	Lehtinen et al. (2019)
≥75	30021	Gamma	400	75	Lehtinen et al. (2019)
Screening	**Deterministic**		**Lower CI 95%**	**Upper CI 95%**	**Source**
Invitation	1		0,65	0,95	Socialstyrelsen (2019)
Primary screening					
Mammography test	25		16	35	Mäklin & Kokko
	Deterministic	**Distribution**	**alpha**	**beta**	
Recall assessment					
Mammography & ultrasound	165	Gamma	400	0,41	VSSHP unit prices (2017)
Mammography, ultrasound & biopsy	332	Gamma	400	0,83	VSSHP unit prices (2017)
Pathological examination of biopsy test	291	Gamma	400	0,83	VSSHP unit prices (2017)

Note. Because costs were presented in the form of graphs, Engauge Digitizer (version 12.1) was used to produce data points representing the stage-specific average costs according to age.

Table B4.2

Direct non-healthcare costs.

Transportation	Deterministic		Lower CI 95%	Upper CI 95%	Source
1st year since diagnosis	250		131	369	Ferkkilä et al. (2018)
2nd year since diagnosis	69		23	117	Ferkkilä et al. (2018)
Remission (>2 years since diagnosis)	33		0	66	Ferkkilä et al. (2018)
Metastatic	970		664	1278	Ferkkilä et al. (2018)
	Deterministic	**Distribution**	**alpha**	**beta**	
Primary screening	5	Gamma	400	0,01	Leivo et al. (1999)
Recall assessment	8,48232	Gamma	400	0,02	Leivo et al. (1999)

Table B4.3

Indirect healthcare costs.

Informal care	Deterministic	Lower CI 95%	Upper CI 95%	Source
1st year since diagnosis	2908	957	4860	Ferkkilä et al. (2018)
2nd year since diagnosis	681	179	1184	Ferkkilä et al. (2018)
Remission (>2 years since diagnosis)	365	150	583	Ferkkilä et al. (2018)
Metastatic	6567	3782	9350	Ferkkilä et al. (2018)
Last year metastatic (incl. Palliative care)	8926	3158	15813	Haltia et al. (2018)

Table B4.4

Indirect non-healthcare costs.

Productivity loss	Deterministic		Lower CI 95%	Upper CI 95%	Source
1st year since diagnosis	9833		7260	12405	Ferkkilä et al. (2018)
2nd year since diagnosis	616		168	1062	Ferkkilä et al. (2018)
Remission (>2 years since diagnosis)	244		66	420	Ferkkilä et al. (2018)
Metastatic	3885		1958	5810	Ferkkilä et al. (2018)
Last year metastatic (incl. Palliative care)	16357		9320	23391	Haltia et al. (2018)
	Deterministic	**Distribution**	**alpha**	**beta**	
Productivity loss	4	Gamma	400	0,01	Leivo et al. (1999)
Productivity loss	5	Gamma	400	0,01	Leivo et al. (1999)

Estimated utility weights and decrements were searched for on the Google Scholar and PubMed databases using the search terms “HRQoL”, “Health-related quality of life”, “Disutility” and “QALY” combined with “Breast cancer screening” and “Mammography”, yielding approximately 32 000 hits. The relevance of studies was based on how well reported disutility values fit the structure of total screening utility decrements used in this model.

Table B4.5

Utility weights and decrements.

BC treatment utility weights	Deterministic	Distribution	alpha	beta	Source
1st year since diagnosis	0,86	Beta	13,14	2,14	Rautalin et al. (2018)
2nd year since diagnosis	0,86	Beta	13,65	2,31	Rautalin et al. (2018)
Remission (> 2 years since diagnosis)	0,84	Beta	15,16	2,89	Rautalin et al. (2018)
Metastatic	0,74	Beta	25,26	8,88	Rautalin et al. (2018)
Last year before death from metastatic BC	0,70	Beta	29,01	12,27	Rautalin et al. (2018)
Screening utility decrements	**Deterministic**	**Distribution**	**alpha**	**beta**	**Source**
Follow-up assessment without biopsy	0,01	Beta	99,34	14989,54	Pataky et al. (2014)
Follow-up assessment with biopsy	0,01	Beta	99,34	14989,54	Pataky et al. (2014)