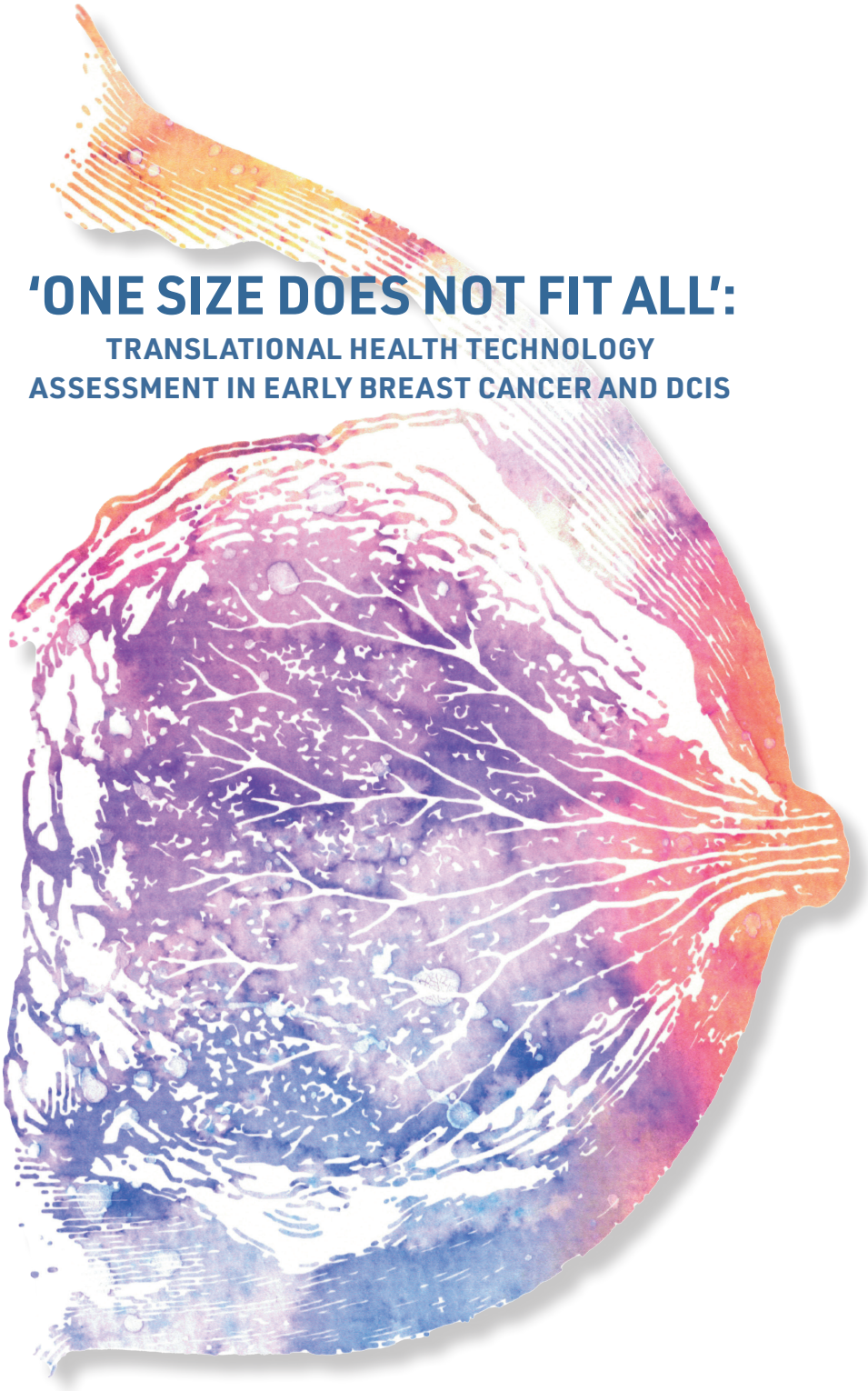
‘ONE SIZE DOES NOT FIT ALL’:

TRANSLATIONAL HEALTH TECHNOLOGY ASSESSMENT IN EARLY BREAST CANCER AND DCIS



Danalyn Byng

‘ONE SIZE DOES NOT FIT ALL’:

TRANSLATIONAL HEALTH TECHNOLOGY ASSESSMENT IN EARLY BREAST CANCER AND DCIS

Danalyn Tulagan Byng

This dissertation has been approved by:

Supervisor
prof. dr. W.H. van Harten

Co-supervisor
dr. V.P. Retèl

The research described in this dissertation was conducted at the Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital, Amsterdam, the Netherlands. In cooperation with the Grand Challenge PRECISION (PREvent ductal Carcinoma In Situ Invasive Overtreatment Now) Consortium, a recipient of the Cancer Research UK Grand Challenge Award 2017, and Duke University, Durham, North Carolina. This thesis is also part of the Health Sciences Series, HSS 22-45, department Health Technology and Services Research, University of Twente, Enschede, the Netherlands. ISSN: 1878-4968.

Research funding was provided by Cancer Research UK and the Dutch Cancer Society (KWF Kankerbestrijding) (grant number C38317/A24043), and through an unrestricted grant from Agendia BV.

Cover design: 'Breast Anatomy' by Erzebetth

Print & Lay-out: proefschrift-AIO

ISBN: 978-90-365-5433-6

DOI: 10.3990/1.9789036554336

‘ONE SIZE DOES NOT FIT ALL’:

TRANSLATIONAL HEALTH TECHNOLOGY ASSESSMENT IN EARLY BREAST CANCER AND DCIS

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof. dr. ir. A. Veldkamp,
on account of the decision of the Doctorate Board

to be publicly defended

on Friday 28 October 2022 at 14.45 hours

by

Danalyn Tulagan Byng

born on the 11th of September, 1991
in Toronto, Canada

Graduation Committee:

Chair:

prof. dr. T. Bondarouk

Supervisor:

prof. dr. W.H. van Harten - University of Twente

Co-supervisor:

dr. V.P. Retèl - University of Twente

Committee Members:

prof. dr. C.J.M. Doggen - University of Twente

prof. dr. M.A. Joore - Maastricht University

prof. dr. M.K. Schmidt - Leiden University

prof. dr. S. Siesling - University of Twente

prof. dr. C.A. Uyl-de Groot - Erasmus University Maastricht

Table of contents


1

Chapter 1

General introduction

Overtreatment and avoidable treatment-related adverse effects not only have important health-related implications, but come at an unsustainable price – even for high-income countries.[1] There is great potential to make oncology care more financially sustainable through de-escalation. However, this largely depends on identifying (cost)-effective prognostic and predictive tools to inform patient management, allowing certain therapies to be safely forgone in some patients and potentially intensified in others. Perhaps more crucial is ensuring the smooth adoption of these tools into routine clinical practice: through acceptance by physicians, patients, and policy makers.

This PhD dissertation aims to understand the factors that may affect the use of interventions that lend themselves to de-escalating low-value interventions for early breast cancer and ductal carcinoma in situ (DCIS). Using approaches grounded in health technology assessment (HTA), it considers clinical effectiveness of current and possible future diagnosis and treatment pathways, as well as all associated economic implications and wider implications for the patient.[2]

Addressing the issue of overtreatment is a complex undertaking which challenges existing therapeutic approaches that have long been central to oncological care. Current treatment strategies for invasive early-stage breast cancer and DCIS are based on active management. Surgical removal of "abnormal" tissue, possibly followed by local and systemic adjuvant therapies, are intended to effectively rid all signs of the tumour and minimize chances of return or progression to advanced or metastatic disease.[3,4]

Costly targeted therapies have been less pervasive in first-line treatment of the majority of women with early-stage disease, keeping in-line with the long-held mantra that early detection may lend itself to earlier, less aggressive treatment. Indeed, cure rates and survival following a breast cancer diagnosis have steadily risen over the past several decades following the introduction of early detection through population-based breast cancer screening and greater access to effective surgical and systemic treatment strategies.[5] More recently however, these improvements have stagnated or plateaued for many women with tumour characteristics associated with very good prognoses for women in industrialized countries.[5] Furthermore, there is clear evidence that a significant proportion of screen-detected breast cancers are overdiagnosed – meaning they were indolent precancers or progressive cancers detected through screening in women who would have otherwise died from non-breast cancer-related causes before a clinical (symptomatic) diagnosis of breast cancer could occur.[6]

A growing number of trials of breast cancer treatments have demonstrated that omission of many standard treatments could actually result in equivalent or non-inferior outcomes in selected good-prognosis patient groups. Many therapeutic approaches for breast cancer are now understood to be low-value, meaning they incur high costs, potentially cause harm and degrade quality of life, while providing no survival benefit.[7,8] Despite these therapies' known benefits, their limitations can have long-term and far-reaching consequences for their recipients. For systemic therapies, this can include the overtreatment of women already cured by effective locoregional therapy, treatment-related toxicities, including the emergence of chronic, long-term, and sometimes life-threatening adverse effects, and financial toxicities created when out-of-pocket treatment costs cause financial problems for the patient.[9]

As a result, the past few years have ushered in a sweeping reform in the treatment of early breast cancer towards a multidisciplinary paradigm of de-escalation (also referred to as 'de-implementation' or 'risk-adapted modulation') for certain women with early breast cancer. Various oncology groups have made recommendations for surgical de-escalation in breast cancer care.[7,10] In 2021, the new St. Gallen International Consensus Guidelines for treatment of early breast cancer were released with a special focus on customizing local and systemic therapies for women with early breast cancer.[11] Among their recommendations, the Panel exhibited renewed and strengthened enthusiasm for genomic testing to identify estrogen receptor (ER) positive, HER2 negative, node positive, early-stage breast cancers that do not warrant chemotherapy. The recently published long-term follow-up data derived from the Microarray in Node-Negative and 1 to 3 Positive Lymph Node Disease May Avoid Chemotherapy (EORTC 10041/BIG 3-04 MINDACT) trial, the Trial Assigning Individualized Options for Treatment (TAILORx), and A Clinical Trial RX for Positive Node, Endocrine Responsive Breast Cancer (RxPonder) have finally established the role of genomic signatures in treatment management for early-stage breast cancer.[12-14] Complimentarily to the efforts to address overtreatment, understanding the optimal level of screening among healthy women[15] and post-diagnosis surveillance imaging[16] among the growing number of women with a history of breast cancer and DCIS have also become important research undertakings.

'Less is More': Health technology assessment (HTA) to assess value

Without a doubt, improvements in screening, diagnosis, treatment, and follow-up for breast cancer have had major influences on breast cancer incidence, mortality and treatment-related morbidity. Yet, how do we adopt newer technologies and begin to change treatment practices which allow us to intervene and provide treatment to those who benefit most, and safely de-escalate treatment for those who will not

benefit? How do we ensure that women who can forgo treatment receive the correct follow-up care and surveillance that can catch any recurrences or progression of their disease?

In order to ensure smooth adoption of prognostic and predictive tools, mechanisms from different perspectives should be taken into account. The domain of health technology assessment (HTA) can support to capture these multi-faceted perspectives. We need evidence on patient quality of life, costs, and preference to complement clinical evidence stemming from prospective studies. With methods derived from the field of HTA, we aim to provide this evidence.[17]

HTA processes were initially introduced to inform decisions about the economic value of new and existing healthcare technologies, including drugs and other medical interventions.[18] Interdisciplinary processes were established to identify, measure, value and compare the costs and consequences of treatment alternatives which are being considered for a particular disease area. Drummond et al. has been instrumental in establishing the methods for economic evaluation of healthcare programmes over several decades,[17] and in 2008 set forth the key principles for improving the conduct of health technology assessments for healthcare resource allocation decisions about new technologies.[18] Today, many countries have commissioned organized HTA bodies to inform healthcare resource allocation decisions about new technologies. These decisions are usually based upon a consolidation of all existing evidence as to whether such technologies are safe and effective within the financial constraints of the healthcare system. Central to this is the analysis of cost-effectiveness, which requires measuring the trade-offs and balance between costs and health outcomes resulting from investing in a new technology, with this question at its core: **"How can the scarce health resources allocated to healthcare best be used in order to maximize the health gain obtained by them?"**[19]

In cost-effectiveness analysis, the costs and effects of a given treatment pathway are calculated and compared to one or more alternatives.[20] A measure of effect that is most widely used is the quality-adjusted life-year (QALY). It captures two very important features of a health intervention: its effect on survival, and on quality of life.

Box 1. Utilities. *Definition derived from York Health Economics Consortium.*[21]

In economic evaluation of healthcare interventions utilities (also called health state preference values) are used to represent the strength of individuals' preferences for different health states. When utility values are averaged over a population of responders they can be considered to be valuations of health states. Conventionally the valuations fall between 0 and 1, with 1 representing the valuation of a state of perfect health and 0 representing the valuation of death (non-existence). In some scoring systems a negative utility value is also possible, which indicates that a (very poor) health state is valued as less preferable than death. Sequences of utility values reported over periods of time for individual patients or cohorts of patients may be aggregated to derive quality-adjusted life years (QALYs), commonly used as outcomes in economic evaluation.

When comparing two treatment options, we consider incremental costs, incremental QALYs, and the incremental cost-effectiveness (or cost-utility) ratio (ICER).

$$ICER = \frac{Cost_a - Cost_b}{QALY_a - QALY_b} = \frac{\Delta Cost}{\Delta QALY}$$

HTA bodies must conduct cost-effectiveness analyses and ensure that subsequent decisions are informed by the values of the society they represent. It is thus expected that preferences and valuations assigned to particular health states differ by country, patient group, and even within patient groups. This variation within a population leads to uncertainty in modeling the cost-effectiveness of a new intervention, often without clear-cut results of whether an intervention is cost-effective, as illustrated by the cost-effectiveness plane in Figure 1.

Nevertheless, HTA bodies must make decisions with all available evidence on whether to list new drugs on a national formulary, whether to provide coverage under health insurance, or to establish any guidance on the technologies' use within the scope of the healthcare system.

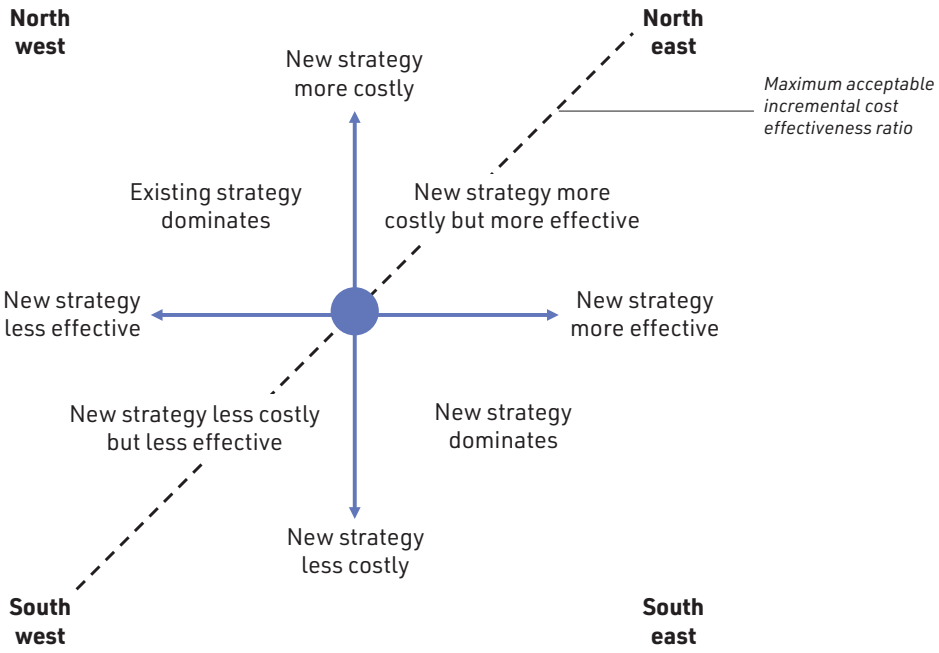


Figure 1. Cost-effectiveness plane, adapted from Gray et al. 2011.[19] The x axis shows the difference in effectiveness between the new treatment strategy and the comparator and the y axis shows the difference in cost. The slope of the line from any point on the figure to the origin is the incremental cost effectiveness ratio.

Early HTA

When limited concrete evidence exists of the safety, effectiveness, and utility of new potential treatment strategies (i.e. in the form of prospective clinical studies), steps must be taken in these early stages of these studies to produce supportive information for decision makers. Methods within early HTA allow us to anticipate developments and preferences which may be encountered during the early introduction of new treatment practices in clinical practice, while also predicting costs and utilities. The new definition of HTA shown in Box 1 best illustrates the flexible and dynamic processes of HTA across the lifecycle of a new treatment strategy or intervention.

Box 2. The new definition of HTA. *Adapted from O'Rourke et al.*[2]

HTA is a multidisciplinary process that uses explicit methods to determine the value of a health technology at different points in its lifecycle. The purpose is to inform decision-making in order to promote an equitable, efficient, and high-quality health system.

Note 1: A health technology is an intervention developed to prevent, diagnose or treat medical conditions; promote health; provide rehabilitation; or organize healthcare delivery. The intervention can be a test, device, medicine, vaccine, procedure, program, or system.

Note 2: The process is formal, systematic, and transparent, and uses state-of-the-art methods to consider the best available evidence.

Note 3: The dimensions of value for a health technology may be assessed by examining the intended and unintended consequences of using a health technology compared to existing alternatives. These dimensions often include clinical effectiveness, safety, costs and economic implications, ethical, social, cultural and legal issues, organizational and environmental aspects, as well as wider implications for the patient, relatives, caregivers, and the population. The overall value may vary depending on the perspective taken, the stakeholders involved, and the decision context.

Note 4: HTA can be applied at different points in the lifecycle of a health technology, that is, pre-market, during market approval, post-market, through to the disinvestment of a health technology.

HTA has a particularly unique role to play in the case of evaluating new technologies to guide treatment de-escalation for early-stage breast cancer and DCIS. As highly heterogeneous diseases, conventional approaches to assess prognosis (e.g. based on the TNM Classification of Malignant Tumors[22]) are not sufficient to clinicians nor patients to decide upon the best treatment approach. Numerous predictive biomarkers have shown promise in guiding treatment decisions in breast cancer, yet very few make it to clinical practice.[23] Furthermore, evidence derived from prospective clinical trials may not be sufficient to inform the necessary adaptations to treatment that individuals require. Clinicians and patients must make inferences from these large studies and continuously updated new data, while customizing this information to individual situations informed by the patient's own preference.[11] For treatment de-escalation, preference may be the most important guiding principle to assess the concept of value. HTA can consolidate the measurement of clinical effects and health outcomes, with the impact of different outcomes on patients' perceived

value.[24] Put simply, forgoing treatment may mean more to one patient than the other given her unique preference-informed values and risk profile.

Early Breast Cancer and DCIS: One-size does not fit all

Among all cancers, breast cancer has become the most commonly occurring cancer affecting women worldwide. In 2020, nearly 2.3 million new cases were diagnosed.[5] It is now the world's most prevalent cancer: at the end of 2020 there were 7.8 women alive with a personal history of breast cancer occurring within the previous 5 years.[25]

Uncovering the highly heterogeneous nature of breast cancer has helped to shape understanding of individual patient prognosis and determine different possibilities for treatment options.[26] It is breast cancer which brought the concept of personalized medicine to the forefront of oncologic care: allowing for the identification of women who benefit most from treatment, and safely de-escalate treatment for those who will not benefit.[27] There is consensus among care providers that the needs of a specific patient may be better defined through consideration of individualized approaches to care. This has also given rise to the prioritization of a woman's quality of life: acknowledging that surviving breast cancer is not merely a question of tackling mortality, but upholding the value of a life minimally affected by treatment-related morbidity.

Before the molecular era[28] of breast cancer treatment and the discovery of new surgical approaches, standard treatment was extensive. Radical mastectomy required the removal of the whole tumor, pectoral muscles, lymphatic vessels and the axillary lymph nodes; a mutilating procedure with profound side effects. Eventually, following important findings about the molecular profile of breast cancer subtypes (Table), physicians began to combine treatment modalities and utilize targeted therapies. Partial breast surgery soon became the mainstay, and was supplemented by adjuvant radiotherapy, chemotherapy and targeted therapy when appropriate.

Despite making up 20-25% of screen-detected breast lesions,[30,31] DCIS is far less understood than invasive breast cancer. DCIS is a nonobligate precursor to invasive breast cancer. This means some, but not all women with DCIS, will progress to invasive breast cancer within their lifetime. The considerable uncertainty about optimal management of the disease results in women continuing to be treated with a "one size fits all" scenario despite mounting evidence pointing to its heterogeneous nature.[32] A consequence of this approach is that women with a diagnosis of DCIS have tended to vastly overestimate their risk of progression to invasive breast cancer and likelihood of dying from breast cancer.[33,34] Many efforts are underway to address communication gaps between patients and providers.[35,36]

Table. Surrogate definitions of intrinsic subtypes of breast cancer (adapted from the 2013 St Gallen Consensus Conference[29] and Ades et al.[28])

	Luminal A	Luminal B	ER2- enriched	Basal like
IHC Surrogate	ER(+) and/or PR(+), HER2(−), Ki67 < 14% (St Gallen)	ER(+) and/or PR(+), HER2(−), Ki67 ≥ 14% (St Gallen)	ER(±), PR(±), HER2(+) (St Gallen)	ER(−), PR(−), HER2(−) (St Gallen)
Prognosis	Good	Intermediate	Poor	Poor
Treatment vulnerability	Endocrine treatment	Endocrine treatment + cytotoxic chemotherapy	HER2 blockade	Cytotoxic chemotherapy

ER: estrogen receptor. HER2: human epidermal growth factor receptor 2. IHC: immunohistochemistry. PR: progesterone receptor.

Now, discussions surrounding the safe de-escalation of locoregional and adjuvant treatment in early breast cancer and DCIS have taken center-stage.[37,38] When addressing issues of over-treatment, attention is being closely paid to women with screen-detected ductal carcinoma in situ (DCIS) and early-stage Luminal A breast cancer. It is among these women, that several can be identified as being unlikely to progress. It is for these women that the risks of treatment may outweigh the potential benefits.

Women with asymptomatic DCIS and early-stage breast cancer represent a significant proportion of women diagnosed through screen-detection methods.[30] Since 1989, a nation-wide biennial screening mammography programme for women aged 50–70 (extended to 75 in 1999) has existed in the Netherlands. Similar organized breast cancer screening programs exist across Europe, while opportunistic breast cancer screening exists in most countries worldwide.[39-41]

Following the introduction of full-field digital mammography (FFDM) screening in 2009, a sharp increase of detection of low-grade DCIS and low-grade invasive carcinomas was observed (Figure 2).[30] Mammography screening sensitivity has also been demonstrated to be highly associated with breast cancer subtypes, with aggressive subtypes showing the lowest sensitivity, though this effect is mediated by grade.[42] The increased incidence of good-prognosis, low-grade breast cancer over time has followed trends of decreasing mortality over the same period. To what extent improvements in survival are attributable to screening or to advancing treatments continues to be debated.[43,44]

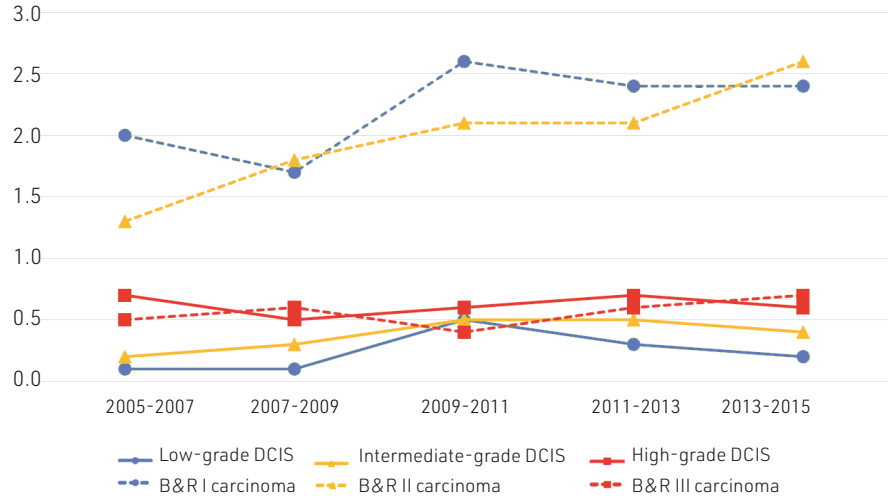


Figure 2: Rate per 1000 screened women. Luiten et al. 2017.[30] B&R = Bloom & Richardson grade.

Breast cancer-related morbidity and mortality represents a major burden to society in the Netherlands. It is estimated that breast cancer is responsible for approximately 3,100 deaths, 26,000 life years lost, 65,000 Disability Adjusted Life Years (DALYs) and an economic burden of €1.27 billion per year.[45] It can be expected that these numbers may grow with continued improvements in screening technologies which may identify more women with breast cancer and DCIS, and with the subsequent high cost of treating greater numbers of patients.[46]

In light of these changing trends and shifting population demographics, efforts are being made to counteract the unintended consequences of improved highly sensitive screening. Many ongoing studies are evaluating new technologies which may better select women at "high" and "low" risk of progression. For DCIS, this means identifying women at risk of progressing to invasive carcinoma within their lifetime. For women with early-stage breast cancer, this has meant identifying those at risk of progressing to distant metastasis following locoregional treatment. With this information, care teams can more accurately select individuals most likely to benefit from locoregional treatment or adjuvant systemic therapy. This can leave women at "low" risk of progression with the choice to forgo high-cost invasive treatments that pose no survival benefit. Though this should be supplemented by regular, high-quality and accessible surveillance in the form of continued breast surveillance.

While the introduction of population-based breast cancer screening has changed the incidence of early-stage disease and DCIS, it remains an imperfect modality. For each 1000 women participating in screening, six will be diagnosed with a screen-detected

invasive breast cancer or DCIS. It is estimated that approximately 20% of these are over-diagnosed, meaning their disease would have remained indolent and undetected had it not been for screening.[6] A further two per 1000 women screened will have breast cancer diagnosed outside of screening in the interval before the next screening round.[47] These are known as interval breast cancers. About 1 in 2 of these breast cancers are considered "missed", minimal signs or occult at the last screening round, a deeply unfortunate and partly preventable situation. The remaining 50% will have a *true* interval cancer that was not visible on mammography at the last screening round. Due to aggressive tumour growth developed during the interval, this subsequently leads to diagnosis through clinical symptoms. Compared to screen-detected breast cancers, symptomatic and interval cancers have poorer prognostic characteristics and lower breast cancer specific survival.[47-49] Even among high-risk cancers with the same genetic make-up, the mode of detection (screen-detected vs. interval-detected) makes a significant difference to the subsequent risk of distant metastasis and survival.[50]

Overview of the dissertation

Chapters in this dissertation are based on three complimentary themes in the management of early-stage breast cancer. The first theme focuses on screen-detected primary DCIS, with select chapters characterizing disease etiology, treatment and surveillance outcomes, real-world health care utilization, and potential of biomarkers to select low-risk women for an active surveillance strategy. Research was performed within the PRECISION (PREvent ductal Carcinoma In Situ Invasive Overtreatment Now) Consortium. The second theme focuses on treatment de-escalation for early-stage breast cancer, based on the first results of the EORTC 10041/BIG 3-04 MINDACT (Microarray in Node-Negative and 1 to 3 Positive Lymph Node Disease May Avoid Chemotherapy) phase 3 randomized control trial of the 70-gene signature. Finally, a complementary final theme and chapter highlights a promising new technology: artificial intelligence for to improve cancer detection at breast cancer screening to decrease the interval cancer rate.

Introduction to the PRECISION Consortium

In a multi-national effort to fill in the gaps of knowledge on DCIS, the PRECISION Consortium was created. The aim of PRECISION is to reduce the burden of overtreatment of DCIS through the development of novel tests that promote informed shared decision-making between patients and clinicians, without compromising the excellent outcomes for DCIS presently achieved. Project objectives are addressed in seven work packages, including a working group on Early Health Technology Assessment. The aim of this working group is to identify, measure, value and compare the consequences of alternative management strategies for DCIS with currently (and "nearly") available clinical evidence.

Leveraging historical cancer registry data

Cancer registry data remains an important and rich source of real-world evidence to model the impact of local treatment options for DCIS. To do so, a dataset which has a substantial number of patients with known treatment status (including no local treatment) and detailed patient and clinical-pathological characteristics (including grade) is required.

Using data on N=85,982 women with primary DCIS from the Surveillance, Epidemiology, and End Results (SEER) Program, a detailed illustration the entire disease process of DCIS is provided in **Chapter 2**. We modeled the competing and intermediate event risks from the point of diagnosis of the primary lesion to death, capturing the intermediary risks of ipsilateral and contralateral invasive breast cancer. We provide a comparison of the disease process for women undergoing different treatment strategies: mastectomy, breast conserving surgery followed by radiotherapy, breast conserving surgery alone, and a cohort of women identified as having not undergone any local treatment. We then further parsed out women with low-risk features, the same women who would have been included in the ongoing prospective clinical trials, to understand their probability of experiencing the competing risks of progression and death.

Real-world uptake of surveillance imaging after DCIS diagnosis

For women diagnosed with primary DCIS treated with breast conserving surgery, clinical guidelines from the American College of Radiology, National Comprehensive Cancer Network, and American Society of Clinical Oncology recommend annual surveillance screening. Guideline adherence remains poorly understood and the highest quality study to date was published in the JCO in 2009 when Nekhlyudov and colleagues[51] reported on mammography uptake in a historic cohort of n=3,037 patients with private insurance. In **Chapter 3** we provide an updated characterization of contemporary imaging surveillance after primary DCIS and relate surveillance uptake to the rate of detection with invasive cancer. In particular, we analyzed a novel US-based cohort of 12,559 DCIS patients with detailed longitudinal follow-up.

Understanding patient and provider treatment and follow-up preferences

In **Chapter 4** we report the results of a preference-based study conducted among 172 women participating in a prospective active surveillance trial for DCIS, and 30 radiation and surgical oncologists involved in the care of women with DCIS. This article provides an unparalleled insight into the first period of recruitment into prospective trials for treatment de-escalation for low-risk DCIS. We describe the challenges associated with enrolment in these de-escalation trials, and the large

numbers of women declining randomization due to their preference for the non-intervention arm. In the Netherlands, the LORD trial was recently changed from an RCT to a preference-based design to address difficulties with enrolment. We took this opportunity to quantitatively measure preferences of the women enrolling into the LORD trial and understand what aspects of a treatment strategy factor into their decision-making process. We did this with a discrete choice experiment, a stated-preference method commonly used in the field of health technology assessment. The same experiment was conducted with oncologists in order to compare their responses with patients.

Could active surveillance for low-risk DCIS be a cost-effective strategy?

Drawing together all the findings presented in Chapters 2, 3, and 4, **Chapter 5** aims to characterize the costs and quality-adjusted health outcomes associated with (non)-interventional strategies for women with low-risk DCIS. A semi-Markov model was constructed based on the modelling approach employed in Chapter 2. The cost-effectiveness analysis explores two opportunities for selecting low-risk women with primary DCIS who could opt for an active surveillance strategy.

Introduction to the 70-Gene Signature and the MINDACT Trial

In 2016, Cardoso and colleagues published the first results of the Microarray in Node-Negative and 1 to 3 Positive Lymph Node Disease May Avoid Chemotherapy (EORTC 10041/BIG 3-04 MINDACT) phase 3 randomized control trial.[52] N=6,693 women with early-stage breast cancer were enrolled and had their risk of distant recurrence evaluated by means of the 70-gene signature and a modified version of Adjuvant! Online. Women were assigned to a genomic and clinical "high" or "low" risk category. In patients with discordant risk results, either the genomic risk or the clinical risk was used to determine the use of adjuvant chemotherapy. The primary goal was to assess whether, among patients with high-risk clinical features and a low-risk gene-expression profile who did not receive chemotherapy, the lower boundary of the 95% confidence interval for the rate of 5-year survival without distant metastasis would be 92% (i.e., the noninferiority boundary) or higher.

Demonstrating the cost-effectiveness of the 70-Gene Signature

In **Chapter 6** we report the results of a cost-effectiveness and budget impact analysis of treatment strategies guided by the 70-gene signature versus treatment decisions based on clinical risk assessment alone for a target group of patients with ER+/HER2- early breast cancer. The analysis is based on patient-level outcome data from the MINDACT trial, information on breast cancer-specific quality of life, as well

as costs for six countries: Belgium, France, Germany, the Netherlands, the United Kingdom, and the United States. This is the first cost-effectiveness analysis on a genomic signature for breast cancer that directly utilizes patient-level data from a large prospective RCT.

What happens if women forgo adjuvant treatment entirely?

The MINDACT trial also afforded opportunities to analyze different trends in adjuvant treatment across the study population. In the study presented in **Chapter 7**, we evaluated the survival of breast cancer patients participating in the MINDACT trial who did not receive any adjuvant systemic treatment after locoregional treatment. These women had favourable prognostic characteristics: ER+/HER2, node negative tumours ≤2 cm. Their breast cancer outcomes were compared to patients with similar characteristics who received endocrine therapy.

Using artificial intelligence to improve breast cancer screening

Throughout the preceding chapters, optimizing treatment through de-escalation was a central theme. In the final chapter, optimizing treatment is explored through emphasizing alternative approaches to reducing the number of cancers that progress to advanced stages.

Chapter 8 covers a study on future perspectives incorporating artificial (AI)-based technology for optimization of breast cancer screening. This was conducted on data from the German national breast cancer screening program, in collaboration with Vara (MX Healthcare GmbH) and the North Mammography Reference Center in Oldenburg, Germany, during the latter part of the PhD study time frame. The study was deemed fitting to include in the thesis by the University of Twente supervising team in view of the topics concerned. Using a large retrospective cohort of women with biopsy-confirmed interval cancer diagnoses (N=2,396), we evaluated the potential of artificial intelligence (AI) to detect and thereby reduce retrospectively visible cancers that would subsequently be clinically diagnosed in the 24-month interval after screening.[53]

Summary

This PhD dissertation delves into current challenges and opportunities for optimization within the management of early-stage breast cancer and DCIS. It addresses attempts made by the medical research community to provide optimal care with less treatment and better surveillance. It also highlights inequities in access to a continuum of care for some women.

References

1. Van Ommen-Nijhof A, Retèl VP, van den Heuvel M, Jager A, van Harten WH, Sonke GS. A revolving research fund to study efficient use of expensive drugs: big wheels keep on turning. *Ann Oncol* 2021; **32**(10): 1212-5.
2. O'Rourke B, Oortwijn W, Schuller T. The new definition of health technology assessment: A milestone in international collaboration. *Int J Technol Assess Health Care* 2020; **36**(3): 187-90.
3. Cardoso F, Kyriakides S, Ohno S, et al. Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2019; **30**(8): 1194-220.
4. Gradishar WJ, Moran MS, Abraham J, et al. NCCN guidelines® insights: Breast cancer, version 4.2021: Featured updates to the NCCN guidelines. *JNCCN* 2021; **19**(5): 484-93.
5. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021; **71**(3): 209-49.
6. Ryser MD, Lange J, Inoue LY, et al. Estimation of breast cancer overdiagnosis in a US breast screening cohort. *Ann Intern Med* 2022; **175**(4): 471-8.
7. Shubeck SP, Morrow M, Dossett LA. De-escalation in breast cancer surgery. *NPJ Breast Cancer* 2022; **8**(1): 25.
8. Trapani D, Franzoi M, Burstein H, et al. Risk-adapted modulation through de-intensification of cancer treatments: an ESMO classification. *Ann Oncol* 2022; **33**(7): 702-12.
9. Piccart MJ, Hilbers FS, Bliss JM, et al. Road Map to Safe and Well-Designed De-escalation Trials of Systemic Adjuvant Therapy for Solid Tumors. *J Clin Oncol* 2020; **38**(34): 4120-9.
10. Moran MS, Schnitt SJ, Giuliano AE, et al. Society of Surgical Oncology-American Society for Radiation Oncology consensus guideline on margins for breast-conserving surgery with whole-breast irradiation in stages I and II invasive breast cancer. *Int J Radiat Oncol Biol Phys* 2014; **88**(3): 553-64.
11. Burstein HJ, Curigliano G, Thürlimann B, et al. Customizing local and systemic therapies for women with early breast cancer: the St. Gallen International Consensus Guidelines for treatment of early breast cancer 2021. *Ann Oncol* 2021; **32**(10): 1216-35.
12. Piccart M, van't Veer LJ, Poncet C, et al. 70-gene signature as an aid for treatment decisions in early breast cancer: updated results of the phase 3 randomised MINDACT trial with an exploratory analysis by age. *Lancet Oncol* 2021; **22**(4): 476-88.
13. Sparano JA, Gray RJ, Ravdin PM, et al. Clinical and genomic risk to guide the use of adjuvant therapy for breast cancer. *N Engl J Med* 2019; **380**(25): 2395-405.
14. Kalinsky K, Barlow WE, Meric-Bernstam F, et al. Abstract GS3-00: First results from a phase III randomized clinical trial of standard adjuvant endocrine therapy (ET)+/-chemotherapy (CT) in patients (pts) with 1-3 positive nodes, hormone receptor-positive (HR+) and HER2-negative (HER2-) breast cancer (BC) with recurrence score (RS)< 25: SWOG S1007 (RxPonder). *Cancer Res* 2021; **81**(4_Supplement): GS3-00-GS3.
15. Shieh Y, Eklund M, Madlensky L, et al. Breast Cancer Screening in the Precision Medicine Era: Risk-Based Screening in a Population-Based Trial. *J Natl Cancer Inst* 2017; **109**(5): djw290.
16. Lam DL, Houssami N, Lee JM. Imaging Surveillance After Primary Breast Cancer Treatment. *AJR Am J Roentgenol* 2017; **208**(3): 676-86.
17. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press, 2015.

18. Drummond MF, Schwartz JS, Jönsson B, et al. Key principles for the improved conduct of health technology assessments for resource allocation decisions. *Int J Technol Assess Health Care* 2008; **24**(3): 244-58; discussion 362-8.
19. Gray AM, Clarke PM, Wolstenholme JL, Wordsworth S. Applied methods of cost-effectiveness analysis in healthcare. Oxford: Oxford University Press, 2010.
20. Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med* 1977; **296**(13): 716-21.
21. York Health Economics Consortium. Utility. 2016. https://yhec.co.uk/glossary/utility/ (accessed February 27 2022).
22. Amin MB, Greene FL, Edge SB, et al. The eighth edition AJCC cancer staging manual: continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA Cancer J Clin* 2017; **67**(2): 93-9.
23. Miquel-Cases A, Schouten PC, Steuten LM, Retèl VP, Linn SC, van Harten WH. (Very) Early technology assessment and translation of predictive biomarkers in breast cancer. *Cancer Treat Rev* 2017; **52**: 117-27.
24. Mühlbacher AC. Patient-centric HTA: different strokes for different folks. *Expert Rev Pharm Out* 2015; **15**(4): 591-7.
25. World Health Organization. Breast Cancer. 2021. https://www.who.int/news-room/fact-sheets/detail/breast-cancer (accessed February 27 2022).
26. Ellis MJ, Perou CM. The genomic landscape of breast cancer as a therapeutic roadmap. *Cancer Discov* 2013; **3**(1): 27-34.
27. Olopade OI, Grushko TA, Nanda R, Huo D. Advances in breast cancer: pathways to personalized medicine. *Clinical Cancer Res* 2008; **14**(24): 7988-99.
28. Ades F, Tryfonidis K, Zardavas D. The past and future of breast cancer treatment-from the papyrus to individualised treatment approaches. *Ecancermedicalscience* 2017; **11**: 746-.
29. Goldhirsch A, Winer EP, Coates AS, et al. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Ann Oncol* 2013; **24**(9): 2206-23.
30. Luiten JD, Voogd AC, Luiten EJT, Duijm LEM. Trends in incidence and tumour grade in screen-detected ductal carcinoma in situ and invasive breast cancer. *Breast Cancer Res Treat* 2017; **166**(1): 307-14.
31. Chootipongchaivat S, van Ravesteyn NT, Li X, et al. Modeling the natural history of ductal carcinoma in situ based on population data. *Breast Cancer Res* 2020; **22**(1): 1-12.
32. Punglia RS, Bifolck K, Golshan M, et al. Epidemiology, biology, treatment, and prevention of ductal carcinoma in situ (DCIS). *JNCI Cancer Spectr* 2018; **2**(4): pky063.
33. Rakovitch E, Franssen E, Kim J, et al. A comparison of risk perception and psychological morbidity in women with ductal carcinoma in situ and early invasive breast cancer. *Breast Cancer Res Treat* 2003; **77**(3): 285-93.
34. Partridge A, Adloff K, Blood E, et al. Risk perceptions and psychosocial outcomes of women with ductal carcinoma in situ: longitudinal results from a cohort study. *J Natl Cancer Inst* 2008; **100**(4): 243-51.
35. Ozanne EM, Soeteman DI, Frank ES, et al. Commentary: Creating a patient-centered decision aid for ductal carcinoma in situ. *Breast J* 2020; **26**(7): 1498-9.
36. Rosenberg SM, Gierisch JM, Revette AC, et al. "Is it cancer or not?" A qualitative exploration of survivor concerns surrounding the diagnosis and treatment of ductal carcinoma in situ. *Cancer* 2022; **128**(8): 1676-83.

37. Hwang ES, Solin L. De-Escalation of Locoregional Therapy in Low-Risk Disease for DCIS and Early-Stage Invasive Cancer. *J Clin Oncol* 2020; **38**(20): 2230-9.
38. Curigliano G, Burstein HJ, Winer EP, et al. De-escalating and escalating treatments for early-stage breast cancer: the St. Gallen International Expert Consensus Conference on the Primary Therapy of Early Breast Cancer 2017. *Ann Oncol* 2017; **28**(8): 1700-12.
39. Ginsburg O, Yip CH, Brooks A, et al. Breast cancer early detection: A phased approach to implementation. *Cancer* 2020; **126**: 2379-93.
40. Zielonke N, Kregting LM, Heijnsdijk EA, et al. The potential of breast cancer screening in Europe. *Int J Cancer* 2021; **148**(2): 406-18.
41. Schünemann HJ, Lerda D, Quinn C, et al. Breast cancer screening and diagnosis: a synopsis of the European Breast Guidelines. *Ann Intern Med* 2020; **172**(1): 46-56.
42. Perron L, Chang S-L, Daigle J-M, et al. Breast cancer subtype and screening sensitivity in the Quebec Mammography Screening Program. *J Med Screen* 2019; **26**(3): 154-61.
43. Otten JD, Broeders MJ, Fracheboud J, Otto SJ, de Koning HJ, Verbeek AL. Impressive time-related influence of the Dutch screening programme on breast cancer incidence and mortality, 1975-2006. *Int J Cancer* 2008; **123**(8): 1929-34.
44. Youlden DR, Cramb SM, Dunn NA, Muller JM, Pyke CM, Baade PD. The descriptive epidemiology of female breast cancer: an international comparison of screening, incidence, survival and mortality. *Cancer Epidemiol* 2012; **36**(3): 237-48.
45. Vondeling GT, Menezes GL, Dvortsin EP, et al. Burden of early, advanced and metastatic breast cancer in The Netherlands. *BMC Cancer* 2018; **18**(1): 262.
46. van der Waal D, Verbeek AL, den Heeten GJ, Ripping TM, Tjan-Heijnen VC, Broeders MJ. Breast cancer diagnosis and death in the Netherlands: a changing burden. *Eur J Public Health* 2015; **25**(2): 320-4.
47. Houssami N, Hunter K. The epidemiology, radiology and biological characteristics of interval breast cancers in population mammography screening. *NPJ Breast Cancer* 2017; **3**: 12.
48. Niraula S, Biswanger N, Hu P, Lambert P, Decker K. Incidence, characteristics, and outcomes of interval breast cancers compared with screening-detected breast cancers. *JAMA Netw Open* 2020; **3**(9): e2018179-e.
49. Hovda T, Hoff SR, Larsen M, Romundstad L, Sahlberg KK, Hofvind S. True and Missed Interval Cancer in Organized Mammographic Screening: A Retrospective Review Study of Diagnostic and Prior Screening Mammograms. *Acad Radiol* 2021; **29**: S180-91.
50. Lopes Cardozo JMN, Schmidt MK, van 't Veer LJ, et al. Combining method of detection and 70-gene signature for enhanced prognostication of breast cancer. *Breast Cancer Res Treat* 2021; **189**(2): 399-410.
51. Nekhlyudov L, Habel LA, Achacoso NS, et al. Adherence to long-term surveillance mammography among women with ductal carcinoma in situ treated with breast-conserving surgery. *J Clin Oncol* 2009; **27**(19): 3211-6.
52. Cardoso F, van't Veer LJ, Bogaerts J, et al. 70-gene signature as an aid to treatment decisions in early-stage breast cancer. *N Engl J Med* 2016; **375**(8): 717-29.
53. Byng D, Strauch B, Gnas L, et al. AI-based prevention of interval cancers in a national mammography screening program. *Eur J Radiol* 2022; **152**: 110321.


2

Chapter 2

Treating (Low-Risk) DCIS Patients: What can we learn from real-world cancer registry evidence?

Danalyn Byng [1,2], Valesca P. Retèl [1,2], Michael Schaapveld [1], Jelle Wesseling [3], Wim H. van Harten [1,2] and on behalf of the Grand Challenge PRECISION Consortium

[1]Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands.

[2]Health Technology and Services Research Department, Technical Medical Centre, University of Twente, Enschede, The Netherlands.

[3]Division of Molecular Pathology, The Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands.

Cancers **13(16)**, 310-320 (2021). https://doi.org/10.3390/cancers13163962

Abstract

Purpose

Results from active surveillance trials for ductal carcinoma in situ (DCIS) will not be available for > 10 years. A model based on real-world data (RWD) can demonstrate the comparative impact of non-intervention for women with low-risk features.

Methods

Multi-state models were developed using Surveillance, Epidemiology, and End Results Program (SEER) data for three treatment strategies (no local treatment, breast conserving surgery [BCS], BCS + radiotherapy [RT]), and for women with DCIS low-risk features. Eligible cases included women aged ≥ 40 years, diagnosed with primary DCIS between 1992 and 2016. Five mutually exclusive health states were modelled: DCIS, ipsilateral invasive breast cancer (iIBC) ≤ 5 years and > 5 years post-DCIS diagnosis, contralateral IBC, death preceded by and death not preceded by IBC. Propensity score-weighted Cox models assessed effects of treatment, age, diagnosis year, grade, ER status, and race.

Results

Data on n = 85,982 women were used. Increased risk of iIBC ≤ 5 years post-DCIS was demonstrated for ages 40–49 (Hazard ratio (HR) 1.86, 95% Confidence Interval (CI) 1.34–2.57 compared to age 50–69), grade 3 lesions (HR 1.42, 95%CI 1.05-1.91) compared to grade 2, lesion size ≥ 2 cm (HR 1.66, 95%CI 1.23–2.25), and Black race (HR 2.52, 95%CI 1.83–3.48 compared to White). According to the multi-state model, propensity score-matched women with low-risk features who had not died or experienced any subsequent breast event by 10 years, had a predicted probability of iIBC as first event of 3.02% for no local treatment, 1.66% for BCS, and 0.42% for BCS+RT.

Conclusion

RWD from the SEER registry showed that women with primary DCIS and low-risk features demonstrate minimal differences by treatment strategy in experiencing subsequent breast events. There may be opportunity to de-escalate treatment for certain women with low-risk features: Hispanic and non-Hispanic white women aged 50–69 at diagnosis, with ER+, grade 1 + 2, < 2 cm DCIS lesions.

Introduction

Women with asymptomatic ductal carcinoma in situ (DCIS) represent a growing proportion of women diagnosed through breast cancer screening programs.[1,2] Localized treatment strategies for DCIS demonstrate no direct survival benefit to patients.[3,4] Surgical removal of the lesion, possibly followed by radiation, is intended to lessen the risk of a subsequent ipsilateral invasive breast cancer (iIBC) and its associated mortality risk. Treatment-related adverse events following surgery and radiotherapy have a profound impact on quality of life over the first 24 months following treatment and there is concern that the active treatment of DCIS represents significant overtreatment for some individuals who will never develop invasive disease within their lifetime.[5]

As all DCIS lesions are treated, the natural disease course of DCIS remains unclear: estimates show a range of 14–53% of untreated DCIS progressing to invasive cancer over a period of 10 or more years.[6] This is a heterogeneous disease, with certain clinicopathologic characteristics known to be highly prognostic of iIBC after DCIS diagnosis, such as premenopausal status, detection by palpation, involved margins, high histologic grade, and high p16 expression.[7] Studies are ongoing to understand risk of progression from DCIS from a genomic perspective.[8] For women with a combination of low-risk clinicopathological features within the DCIS population, the risk of subsequent iIBC has not yet been quantified. Now, discussions surrounding the safe de-escalation of treatment of DCIS have taken center-stage to address this knowledge gap. An active surveillance strategy has been proposed for patients with low-risk prognostic features, including low-grade and smaller, estrogen receptor positive (ER+) lesions. This allows for the prioritization of a woman's quality of life: acknowledging that preventing breast cancer is not merely a question of tackling risk factors, but upholding the value of a life minimally affected by treatment-related morbidity. The international PRECISION (PREvent ductal Carcinoma In Situ Invasive Overtreatment Now) initiative is overseeing three clinical trials of active surveillance for low risk DCIS: Comparison of Operative to Monitoring and Endocrine Therapy (COMET), Low Risk DCIS (LORD) and Low RISk DCIS (LORIS).[9-11] These trials compare safety and clinical outcomes between patients undergoing standard interventional treatment, and those following an active surveillance strategy with regular mammographic screening.

These studies are on-going, and results will not be available for 10–20 years. Ahead of prospective data from clinical trials, real-world cancer registry data on DCIS can be used to demonstrate how women with low-risk features progress from DCIS to IBC and death. We specifically sought to identify a cohort of women

with low-grade, small (< 2 cm), ER+ lesions to who did not receive local-regional treatment to understand the potential impact of an active surveillance strategy compared to standard interventional treatment on health outcomes over a patient's lifetime. Using real-world cancer registry data from the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) program on locally treated and untreated DCIS patients, we developed a continuous time multi-state Markov model of disease progression for DCIS, integrating patient-level covariates and treatment information. The SEER database records subsequent invasive breast cancer cases after a DCIS diagnosis as new primaries, allowing for the modeling of breast cancer-specific disease progression over a patients' lifetime.

Methods

SEER patient cohort selection

Retrospective patient-level data from the SEER 18 registries database (with additional treatment fields on radiation therapy) were used for multi-state modeling of disease progression. Eligible cases included women with grade I, II, and III histologically confirmed DCIS as first primary, diagnosed between 1992 and 2016, aged ≥ 40 years at diagnosis, and with known laterality, local treatment status (surgery and radiotherapy), survival time, and cause of death. Exclusion was warranted under any of the following criteria: iIBC ≤ 2 months following DCIS as this might signify upstaging of the DCIS lesion to invasive carcinoma; death of any cause ≤6 months following DCIS diagnosis; synchronous diagnosis of contralateral invasive carcinoma (cIBC); Paget's disease; patients treated with postmastectomy radiation therapy; and patients not receiving treatment due to comorbidities or refusal (as coded in SEER). Figure 1 shows the numbers of cases excluded.

Capturing local invasive recurrences in SEER

To understand the impact of changes in SEER coding rules in 2007 which may have led to the under-reporting of subsequent iIBC following DCIS, we calculated the annual iIBC incidence density rate in the 5 years pre- and post-2007. This calculation is based on the number of iIBC events in each annual period, divided by the product of the person-time of the at-risk population during each period. This is presented for the full cohort (all risk groups), and by treatment group to account for changing treatment patterns.

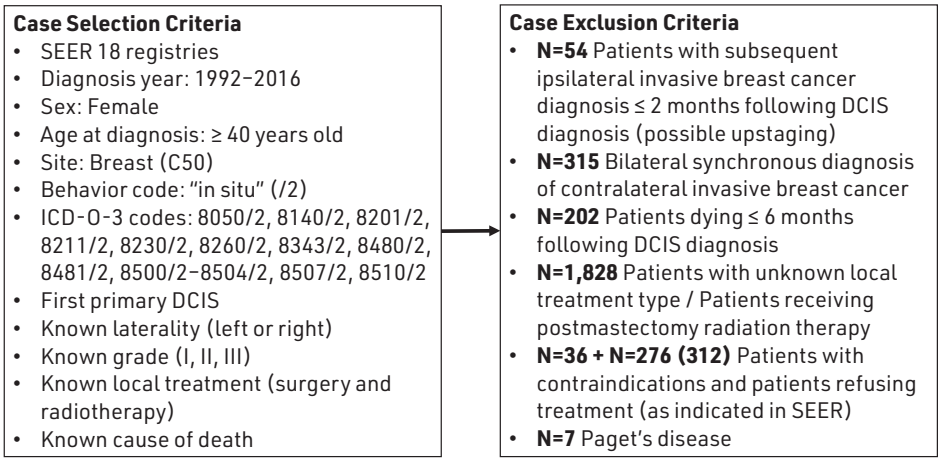


Figure 1. Surveillance, epidemiology, and end results (SEER) case selection and exclusion criteria.

Model building and statistical analysis

The multi-state model structure includes six mutually exclusive states, and the seven transitions between each state (Fig. 2). The effects of baseline patient, disease, and treatment characteristics on each transition was assessed using multivariate Cox proportional hazard regression models. The selected covariates included age at diagnosis (40–49, 50–69, 70–74, 75–79, ≥ 80 years), diagnosis year (1992–2016), race (Hispanic and non-Hispanic white, Hispanic and non-Hispanic black, other [Asian, Native American, Pacific Islander]), grade (I, II, III), lesion size (< 2 cm, ≥ 2 cm), estrogen receptor (ER) status, and local treatment strategy (no local treatment, breast conserving surgery [BCS] only, BCS followed by radiotherapy [RT], mastectomy). Complete cases were available for all variables (age, diagnosis year, treatment strategy), except for ER status, lesion size, and race. Missing observations were imputed with the substantive model compatible fully conditional specification method using co-variables with complete cases (age, diagnosis year, treatment strategy) and outcome (time, event). This method allows greater flexibility for non-linear models such as the Cox model, in that partially observed covariates are imputed based on non-linear covariate effects.[12] The R package smcfcs version 1.4.0 was used.

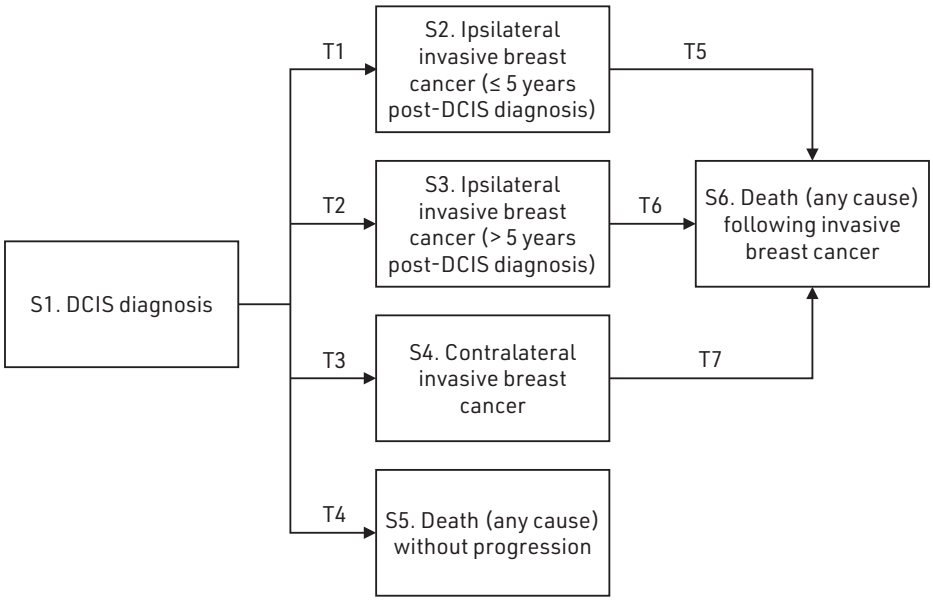


Figure 2. Multi-state model structure. The multi-state model structure includes six mutually exclusive health states (S1–S6) each represented by a box, and the seven transitions between each state (T1–T7). Arrows represent all possible transitions between states

To address possible confounding by indication, i.e. the systematic differences between patients undergoing different treatment strategies, propensity scores (PS) were calculated for each individual. The propensity score is an individual's probability of receiving treatment given their pre-treatment characteristics (i.e. age, diagnosis year, grade, race, lesion size, ER status). As there are four treatment strategies being compared, generalized boosted regression models were used to compute PS weights which balance the distribution of selected characteristics between treatment and comparison groups. The pre-treatment characteristics listed above were used to calculate PS. The mean standardized effect size and Kolmogorov-Smirnof statistic were used to choose the optimal number of iterations to establish balance. Average treatment effect (ATE) analysis was conducted to determine the relative effectiveness of no intervention, BCS, BCS+RT, and mastectomy on average in the population. For each transition-specific Cox proportional hazards model, individuals were weighted by the inverse probability of receiving the treatment they received. Doubly robust estimation controlled for any covariates with lingering imbalances. PS analysis was conducted using the R package Twang version 1.5.

To address the violation of the proportionality assumption for some predictors in the Cox model for the transition from DCIS diagnosis to iIBC and to address the Markov assumption, time to iIBC was split at 5 years post-DCIS. Therefore the following multi-state transitions were modeled: T1. DCIS diagnosis → iIBC ≤ 5 years following diagnosis; T2. DCIS diagnosis → iIBC > 5 years following DCIS diagnosis; T3. DCIS diagnosis → cIBC; T4. DCIS diagnosis → death; T5. iIBC ≤ 5 years following diagnosis → death; T6. iIBC > 5 years following diagnosis → death; T7. cIBC → death. Intermediate lesions such as a subsequent diagnosis of DCIS during follow-up after initial DCIS are not considered in the model.

Conditional transition probabilities were computed for each treatment strategy cohort (except mastectomy) and the sub-cohort of patients with low-risk features (Hispanic and non-Hispanic white women aged 50–69 at diagnosis, with ER+, grade 1 + 2, ≤ 2 cm DCIS lesions) by building Cox models stratified by transition to compute cumulative transition hazards transformed into conditional transition probabilities using the Aalen-Johansen estimator. State occupation probabilities at different time points following DCIS diagnosis could be derived from these values. Data preparation and multi-state modeling was done using the R package mstate version 0.2.11.

PS-matched groups were also created for comparison when calculating the transition probabilities derived from the multi-state models. 1:2 matching of the n=338 individuals in the low-risk non-intervention group to each of the low-risk treatment groups was carried out using the "nearest neighbour" method in the MatchIt R package version 3.0.2. Exact matching was specified by year of diagnosis, age at diagnosis, and grade. Differences in iIBC at 5 years between low-risk PS-matched treatment groups were also evaluated using hazard ratios with 95% CIs derived from Cox proportional hazards models.

All statistical analyses were performed with R version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria).

Results

Patient characteristics

Table 1 shows the patient and clinicopathologic characteristics of the N = 85,982 individuals included in the analysis set, including N = 1650 who did not receive local intervention, and N = 17,714 patients with low-risk features (Hispanic and non-Hispanic white women aged 50–69 at diagnosis, with ER+, grade 1 + 2, ≤ 2 cm DCIS lesions). Women undergoing more invasive procedures (BCS+RT, mastectomy) were generally younger with higher-risk features (high grade, large lesion sizes).

Annual iIBC incidence rate (1996–2016)

Figure 3 shows the annual iIBC incidence density rate across the 2002–2011 observation period according to the person-years at risk within our cohort during each year. With the exception of the group without local treatment, there is no obvious jump in iIBC rates post-2007. This pattern remained steady across treatment cohorts.

Transition-specific PS-weighted multivariate Cox proportional hazards models

Select baseline risk factors were shown to be highly predictive of iIBC events within the first 5 year period, with diminishing hazard for later occurring events (Table 2). Multivariate-adjusted PS-weighted models showed that women aged 40–49 at diagnosis had a statistically significantly higher risk of subsequent iIBC within 5 years compared to women aged 50–69 (Hazard ratio (HR) 1.86, 95% Confidence Interval (CI) 1.34–2.57). Grade 3 lesions also carried a higher risk compared to grade 2 (HR 1.42, 95% CI 1.05–1.91). This significant effect of high grade was not observed for events occurring after 5 years (HR 1.00, 95% CI 0.73–1.38). Lesion size ≥ 2 cm (HR 1.66, 95% CI 1.23–2.25), and black race (HR 2.52, 95% CI 1.83–3.48 compared to white race) were also predictive of subsequent iIBC events within 5 years and after 5 years (Table 2). ER+ status did not have a statistically significant association with iIBC risk for any time period. Age groups ≥ 70 years did not show a statistically significant different HR of iIBC ≤ 5 years compared to age 50–69; nor did grade 1 compared to grade 2 (Table 2).

Baseline characteristics of the primary DCIS did not demonstrate any statistically significant relationship with cIBC events, with the exception of age 70–74 which carried a higher hazard of cIBC events compared to age 50–69 (HR 1.26, 95% CI 1.11–1.42) (Table 2).

Figure 3. Ipsilateral invasive breast cancer (iIBC) incidence density rate (2002-2011)

Multi-state modeling

State occupancy probabilities for the "progression-free" state calculated from the multi-state models are visualized in Fig. 4 for the different treatment modalities for patients in the low-risk subgroup. All other transition probabilities calculated from the multi-state models are visualized in Supplementary Fig. 1; the distance between two curves represents the probability of being in a specific state at a specific time point (state occupation probability). Time-dependent transition probabilities and accompanying standard errors are listed in Supplementary Table 1–8. For low-risk women not receiving local treatment, the probability of being alive and remaining iIBC-free at 5 years was 95.5% (95% CI 87.5–98.4%) and 89.2% (95% CI 78.2–94.7%) at 10 years. The probability of experiencing an iIBC as first event at 5 years was 0.92% (95% CI 0.00–1.95%) and 3.02% (95% CI 0.00–6.05%) at 10 years. In the same cohort of low-risk women, matched according to PS and patient characteristics, the probability of experiencing an iIBC at 5 years was 0.88% (95% CI 0.10–1.66%) following BCS, and 0.35% (95% CI 0.00–0.80%) following BCS+RT. The 10 year probability was 2.48% (95% CI 0.82–4.11%) and 0.58% (95% CI 0.00–1.39%) respectively for BCS and BCS+RT. All transition probabilities in PS-matched groups are listed in Supplementary Tables 1–8. No statistically significant differences in iIBC at 5 years between low-risk PS-matched treatment groups were detected (BCS vs. AS: HR 0.83, 95% CI 0.19–3.48; BCS+RT vs. AS: HR 0.75, 95% CI 0.13–4.49).

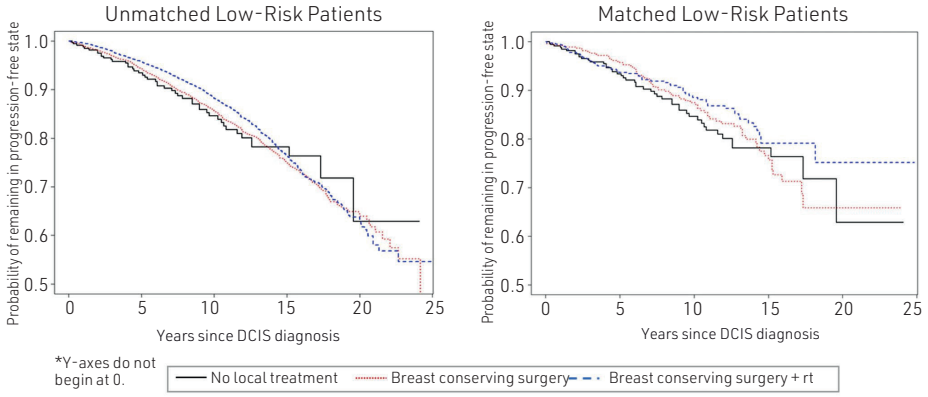


Figure 4. Progression-free state occupancy probabilities for patients with low-risk features. rt: radiotherapy.

Discussion

This analysis applied real-world cancer registry data from n=85,982 women diagnosed with primary DCIS. The excellent iIBC-free survival observed at 5 and 10 years for the women in this cohort with low-risk features is an important confirmation that an active surveillance strategy could be safe and feasible compared to standard interventional treatment. For those with low-risk features (Hispanic and non-Hispanic white women aged 50–69 at diagnosis, with ER+, grade 1 + 2, ≤ 2 cm DCIS lesions) who did not receive local treatment, their prognosis remained comparable to their matched counterparts who received surgery with or without radiotherapy. The observed 10-year probability of iIBC at 3.0%, as well as the combined risk of contralateral and ipsilateral IBC remains well within the 10-year population-wide age-specific probability of developing IBC for US women (range 2.3–3.9%).[13]

Improving the understanding of the disease process after diagnosis and treatment of primary DCIS remains an important undertaking. Through the development of multi-state models using real-world data, we were able to provide insight into how patients transition from DCIS diagnosis to iIBC or cIBC across treatment strategies. Multi-state modeling provides an advantageous approach over typical time-to-event modeling techniques as it allowed us to visualize competing event risks, and to understand what happens after an intermediate event such as an IBC. Across treatment strategies there were similar probabilities of dying without an IBC, with comparatively very low probabilities of death following IBC (Supplementary Fig. 1).

Previous studies have attempted to simulate various possibilities of the natural history of DCIS, without demarcating subgroups based on risk of subsequent breast

events.[14] This is the first study to explicitly model the disease process for women with features deemed to make them low-risk for subsequent iIBC, for whom an active surveillance strategy is targeted towards. We provide evidence beyond previously published studies which provided limited direct comparison of no locoregional treatment and standard surgical strategies. Ryser et al. recently conducted a study on cancer outcomes in DCIS patients without locoregional treatment identified in the SEER dataset. When analyzing their low-risk subgroup (non-high grade, ER/PR+, > 40 years at diagnosis) in a competing risk analysis, the 7.5-year cumulative incidence of iIBC was 5.9% (95% CI 2.3–9.5%).[15] In our analysis, the subgroup of low-risk women is further limited to women aged 50–69 at diagnosis, with small (< 2 cm) lesions. We additionally limit this selection to Hispanic and non-Hispanic White women, as our multi-state model revealed Black race to be a strong marker of iIBC ≤ 5 years post-DCIS. As cancer health disparities in racial and ethnic groups in the United States are well-established, in this analysis we do not designate race as a biological risk factor.[16] Further analysis into the systemic disadvantage and structural inequalities in screening and follow-up care which contribute to poorer health outcomes for women in minority racial and ethnic groups diagnosed with DCIS is warranted.

The SEER dataset is rich in clinico-pathological information and socio-demographic information which helps us to understand who is more likely to receive certain treatment modalities and how this impacts their health outcomes. However, despite SEER being one of the widely used cancer registries for observational research, its use is not without its possible pitfalls. The potential impact of misclassification of surgery and radiation for women who did not receive treatment should be confirmed by careful review of medical records or by patient interview. While SEER records the most invasive surgical procedure on the primary site, it is possible that some women diagnosed with DCIS at one institution sought surgical and/or radiation treatment at another institution not within the same SEER registry catchment area. Nevertheless, analyses comparing agreement between SEER data and Medicare claims for receipt of RT demonstrated that SEER reliably identified individuals who received treatment for in situ female breast events.[17] Beyond potential misclassification of treatment, the Ryser study was critiqued as having artificially low estimates of iIBC incidence, especially for cases diagnosed before changes to SEER coding of "recurrences" in 2007.[15,18] The SEER program collects data on subsequent primary cancers, but does not record information on cancer recurrences. Indeed, a diagnosis of a subsequent invasive breast cancer following DCIS can be described either as a loco-regional invasive recurrence or a new primary cancer, and language to describe this phenomena has not been consistent. In order to understand the impact of changes in SEER coding rules in 2007 which may have led to the earlier under-reporting of

subsequent iIBC following DCIS, we calculated the annual iIBC incidence density rate across the 2002–2011 observation period according to the person-time at risk within our cohort during each year. The group without local treatment showed significant variation over time, while the pattern remained steady for the cohort as a whole. This is an important observation to understand relative treatment effects (Fig. 3).

Previous studies of IBC have made attempts to distinguish new primary tumors from true recurrences after IBC, with consistent reporting that true recurrences occur sooner than new primary tumors.[19-21] We identified time dependencies for many covariates in our Cox models. This led us to splitting iIBC into two states, distinguished by events that occurred within, or following, 5 years after DCIS diagnosis. We observed a strong association between high grade and earlier ipsilateral invasive events (occurring within 5 years). The same association was not observed for events occurring after 5 years. It is possible that this is a reflection of the clonal relationship of the primary DCIS and any subsequent iIBCs; we can hypothesize that iIBC events occurring more than 5 years after the primary DCIS are likely to be unrelated, new primary tumors. Previously published information on IBC after DCIS combined with our evidence on the time-dependency of DCIS grade can inform decisions on appropriate follow-up length for future studies concerning treatment approaches for primary DCIS.

To explore the relative treatment effects on iIBC within 5 years of DCIS diagnosis for women with low-risk features, we looked at treatment-specific hazard ratios. Women with no local treatment were matched 1:2 with women treated with surgery (BCS ± RT) according to PS, and by year of diagnosis, age at diagnosis, and grade (all women considered low-risk had ER+ lesions < 2 cm and were (non)-Hispanic white). Hazard ratios showed a protective effect for surgical interventions (HR < 1) but this was non-significant in all cases.

It is well-known that the diagnosis of DCIS is associated with an increased risk of breast cancer. Retrospective observational registry studies continue to confirm this in different screen-detected DCIS populations.[22] However, for women with low-risk features, this risk is likely to be well-managed with an active surveillance strategy where bi-annual physical examinations and annual mammography allow the lesion to be closely monitored. If a woman receives local treatment for DCIS, the likelihood of a subsequent iIBC remains low. However, any subsequent loco-regional iIBC events in a previously irradiated breast will be more difficult to treat locally with re-irradiation due to increased risk of skin and subcutaneous toxicity because re-irradiation will exceed the maximum tolerable dose of radiotherapy of the skin and subcutaneous tissue. Irreversible radiation-induced fibrosis and radionecrosis

hinders the efficacy of systemic chemotherapy.[23] Treatment-related complications are further compounded by the emotional and economic toll that initial local treatment represents [5]. In a recent study on treatment preferences for screen-detected DCIS, patients valued active monitoring over standard interventional treatment.[24] This was largely influenced by the risk of progression: a 10% risk of progression at 10 years was deemed an acceptable trade-off to avoid possible side-effects from surgery or radiotherapy. Compared to the observed iIBC risk at 10 years in women with low-risk characteristics who did not receive local treatment at 3%, this provides further evidence of patients' willingness to be followed under a demonstrably safe active surveillance strategy.

Conclusion

As physicians treating women with low-risk DCIS await results from prospective trials on active surveillance, there is value in harnessing real-world evidence from cancer registries to support present-day decision-making for possible non-intervention in (low-risk) DCIS. With multi-state models, it is possible to visualize, quantify, and compare competing breast event risks for different treatment and risk groups. Evaluating time dependencies of prognostic factors in the models also allowed for the understanding of the relationship between subsequent iIBCs and the primary DCIS. Replacing conventional invasive treatment with active surveillance in this good prognosis population could improve women's well-being during the remaining (progression-free) survival time without resulting in significantly poorer disease outcome. This is an important factor to consider when making an informed treatment decision in this patient population. Capturing the full impact of possible treatment strategies over a patient's lifetime involves integrating health outcomes, health-related quality of life, patient and provider preference, as well as direct and indirect costs. In this study we provide the first set of information to help model progression outcomes and transitions between health states. This model can easily be extended to integrate cost and quality of life data points, so that researchers can model the potential cost-utility of new disease management strategies for this specific cohort of low-risk DCIS patients.

Table 1. Patient and Clinical-Pathological Characteristics

Characteristic		No Intervention (n=1,650) n (%)	BCS only (n=22,698) n (%)	BCS+ RT (n=40,265) n (%)	Mastectomy (n=21,369) n (%)	Total population (n=85,982) n (%)
Median follow-up, months (IQR)		73 (34-133)	93 (44-152)	87 (43-140)	90 (45-150)	89 (44-145)
Year of diagnosis	1992-1999	118 (7.2 %)	2,739 (12.1 %)	3,110 (7.7 %)	2,466 (11.5 %)	8433 (9.8 %)
	2000-2010	906 (54.9 %)	12,699 (55.9 %)	22,024 (54.7 %)	11,495 (53.8 %)	47,124 (54.8 %)
	2011-2016	626 (37.9 %)	7,260 (32.0 %)	15,131 (37.6 %)	7,408 (34.7 %)	30,425 (35.4 %)
Age at diagnosis	40-49	323 (19.6 %)	3,992 (17.6 %)	8,588 (21.3 %)	6,159 (28.9 %)	19,062 (22.2 %)
	50-69	826 (50.1 %)	11,375 (50.1 %)	24,386 (60.6 %)	11,268 (52.7 %)	47,855 (55.7 %)
	70-74	154 (9.3 %)	2,622 (11.6 %)	3,837 (9.5 %)	1,677 (7.8 %)	8,290 (9.6 %)
	75-79	121 (7.3 %)	2,251 (9.9 %)	2,324 (5.8 %)	1,299 (6.1 %)	5,995 (7.0 %)
	≥80	226 (13.7 %)	2,458 (10.8 %)	1,130 (2.8 %)	966 (4.5 %)	4,780 (5.6 %)
Race	White	1,169 (70.8 %)	18,037 (79.5 %)	31,284 (77.7 %)	16,503 (77.2 %)	66,993 (77.9 %)
	Black	239 (14.5 %)	2,241 (9.9 %)	4,417 (10.7 %)	2,329 (10.9 %)	9,226 (10.7 %)
	Other (American Indian/AK Native, Asian/Pacific Islander)	133 (8.1 %)	2,172 (9.6 %)	4,375 (10.9 %)	2,432 (11.4 %)	9,112 (10.6 %)
	Unknown	109 (6.6 %)	248 (1.1 %)	189 (0.5 %)	105 (0.5 %)	651 (0.8 %)
DCIS grade	I	364 (22.1 %)	5,912 (26.0 %)	6,092 (15.1 %)	2,645 (12.4 %)	15,013 (17.5 %)
	II	786 (47.6 %)	11,378 (50.1 %)	17,688 (43.9 %)	8,470 (39.6 %)	38,322 (44.6 %)
	III	500 (30.3 %)	5,408 (23.8 %)	16,485 (40.9 %)	10,254 (48.0 %)	32,647 (38.0 %)
Lesion size	<2 cm	490 (29.7 %)	14,470 (63.6 %)	25,731 (63.9 %)	9,263 (43.3 %)	49,954 (58.0%)
	2-5 cm	110 (6.7 %)	2,167 (9.5 %)	5,529 (13.7 %)	5,099 (23.9 %)	12,905 (15.0%)
	>5 cm	34 (2.1 %)	287 (1.3 %)	498 (1.2 %)	1,560 (7.3 %)	2,379 (2.8%)
	Unknown	1,016 (61.6 %)	5,774 (25.4 %)	8,507 (21.1 %)	5,447 (25.5 %)	20,744 (24.1%)
ER status	Positive/Borderline	918 (45.7 %)	11,729 (51.7 %)	25,117 (62.4 %)	11,153 (52.2 %)	48,704 (56.6 %)
	Negative	127 (6.3 %)	1,232 (5.4 %)	4,317 (10.7 %)	2,955 (13.8 %)	8,595 (10.0 %)
	Unknown	963 (48.0 %)	9,737 (42.9 %)	10,831 (26.9 %)	7,261 (34.0 %)	28,683 (33.4 %)

AK: Alaska; DCIS: ductal carcinoma in-situ; BCS: breast conserving surgery; IQR: inter-quartile range; ER: estrogen receptor; REF: reference category; RT: radiotherapy.

Table 2. Propensity Score-Weighted Cox Proportional Hazards models

		T1 (diagnosis → iIBC ≤5) HR (95% CI)	T2 (diagnosis → iIBC >5) HR (95% CI)	T3 (diagnosis → cIBC) HR (95% CI)	T4 (diagnosis → death) HR (95% CI)	T5 (iIBC ≤5 → death) HR (95% CI)	T6 (iIBC >5 → death) HR (95% CI)	T7 (cIBC → death) HR (95% CI)
Local treatment	BCS+RT	REF	REF	REF	REF	REF	REF	REF
	Mastectomy	0.50 (0.39-0.65)	0.21 (0.16-0.26)	1.18 (1.07-1.31)	1.14 (1.08-1.20)	1.30 (0.74-2.30)	1.31 (0.68-2.52)	0.96 (0.75-1.23)
	BCS	3.14 (2.70-3.67)	1.35 (1.20-1.53)	1.05 (0.95-1.16)	1.19 (1.13-1.25)	1.05 (0.70-1.58)	0.99 (0.70-1.39)	1.01 (0.78-1.23)
	No treatment	4.26 (3.12-5.81)	1.69 (1.19-2.41)	1.17 (0.87-1.57)	1.55 (1.34-1.78)	3.32 (1.31-8.45)	1.01 (0.38-2.70)	1.81 (1.06-3.09)
Year of diagnosis		1.02 (0.98-1.05)	0.97 (0.94-1.01)	1.00 (0.98-1.01)	1.00 (0.99-1.01)	1.03 (0.97-1.09)	0.98 (0.86 -1.12)	1.02 (0.97-1.07)
Age at DCIS diagnosis	50-69	REF	REF	REF	REF	REF	REF	REF
	40-49	1.86 (1.34-2.57)	1.37 (1.02-1.83)	0.76 (0.63-0.92)	0.31 (0.24-0.39)	1.29 (0.55-3.07)	1.20 (0.50-2.85)	1.23 (0.68-2.23)
	70-74	1.01 (0.59-1.74)	1.20 (0.74-1.95)	1.40 (1.09-1.80)	3.23 (2.84-3.68)	4.97 (2.08-11.88)	2.44 (0.98-6.05)	2.08 (1.19-3.64)
	75-79	0.95 (0.61-1.46)	0.97 (0.75-1.25)	1.02 (0.76-1.37)	5.96 (5.28-6.72)	4.46 (2.11-9.44)	4.25 (2.28-7.90)	2.96 (1.54-5.68)
	≥80	1.14 (0.80-1.63)	0.62 (0.44-0.87)	1.06 (0.74-1.51)	10.84 (9.78-12.02)	6.24 (3.26-11.93)	6.76 (3.12-14.64)	5.86 (3.73-9.22)
Grade	2	REF	REF	REF	REF	REF	REF	REF
	1	0.87 (0.58-1.32)	1.04 (0.74-1.44)	0.84 (0.69-1.02)	1.13 (1.02-1.24)	0.72 (0.29-1.78)	0.94 (0.41-2.14)	1.26 (0.82-1.93)
	3	1.42 (1.05-1.91)	1.00 (0.73-1.38)	0.82 (0.68-1.02)	1.10 (0.99-1.22)	1.23 (0.66-2.32)	1.29 (0.69-2.42)	0.70 (0.45-1.09)
Race	Caucasian	REF	REF	REF	REF	REF	REF	REF
	African American	2.52 (1.83-3.48)	1.79 (1.25-2.55)	1.12 (0.89-1.42)	1.37 (1.21-1.54)	1.47 (0.71-3.07)	1.51 (0.68-3.39)	0.97 (0.58-1.62)
	Other	1.46 (0.84-2.55)	1.54 (1.00-2.36)	1.07 (0.82-1.39)	0.75 (0.64-0.88)	0.46 (0.15-1.40)	1.77 (0.64-4.95)	0.66 (0.26-1.69)
Lesion size	< 2 cm	REF	REF	REF	REF	REF	REF	REF
	≥ 2 cm	1.66 (1.23-2.25)	1.38 (1.00-2.36)	1.08 (0.90-1.30)	1.18 (1.06-1.30)	1.75 (1.02-3.01)	1.80 (0.82-3.93)	1.28 (0.85-1.91)
ER status	Negative	REF	REF	REF	REF	REF	REF	REF
	Positive	0.76 (0.53-1.10)	1.51 (0.98-2.33)	1.13 (0.87-1.46)	0.94 (0.83-1.06)	0.40 (0.19-0.85)	0.51 (0.29-0.90)	0.67 (0.36-1.25)

CI: confidence interval; cIBC: contralateral invasive breast cancer; DCIS: ductal carcinoma in-situ; BCS: breast conserving surgery; HR: hazard ratio; iIBC: ipsilateral invasive breast cancer; IQR: inter-quartile range; ER: estrogen receptor; REF: reference category; RT: radiotherapy.

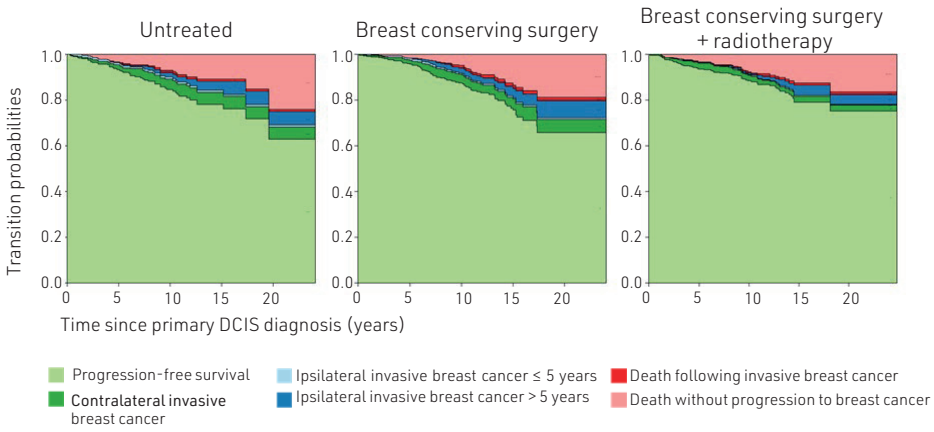
References

1. Lehman CD, Arao RF, Sprague BL, et al. National performance benchmarks for modern screening digital mammography: update from the Breast Cancer Surveillance Consortium. *Radiology* 2017; **283**(1): 49.
2. Ryser MD, Hendrix LH, Worni M, Liu Y, Hyslop T, Hwang ES. Incidence of Ductal Carcinoma In Situ in the United States, 2000–2014. *Cancer Epidemiol Biomarkers Prev* 2019; **28**(8): 1316-23.
3. Sagara Y, Mallory MA, Wong S, et al. Survival Benefit of Breast Surgery for Low-Grade Ductal Carcinoma In Situ: A Population-Based Cohort Study. *JAMA Surg* 2015; **150**(8): 739-45.
4. Correa C, McGale P, Taylor C, et al. Overview of the randomized trials of radiotherapy in ductal carcinoma in situ of the breast. *J Natl Cancer Inst Monogr* 2010; **2010**(41): 162-77.
5. King MT, Winters ZE, Olivotto IA, et al. Patient-reported outcomes in ductal carcinoma in situ: a systematic review. *Eur J Cancer* 2017; **71**: 95-108.
6. Erbas B, Provenzano E, Armes J, Gertig D. The natural history of ductal carcinoma in situ of the breast: a review. *Breast Cancer Res Treat* 2006; **97**(2): 135-44.
7. Visser LL, Groen EJ, van Leeuwen FE, Lips EH, Schmidt MK, Wesseling J. Predictors of an Invasive Breast Cancer Recurrence after DCIS: A Systematic Review and Meta-analyses. *Cancer Epidemiol Biomarkers Prev* 2019; **28**(5): 835-45.
8. Spira A, Yurgelun MB, Alexandrov L, et al. Precancer Atlas to Drive Precision Prevention Trials. *Cancer Res* 2017; **77**(7): 1510-41.
9. Hwang ES, Hyslop T, Lynch T, et al. The COMET (Comparison of Operative versus Monitoring and Endocrine Therapy) trial: a phase III randomised controlled clinical trial for low-risk ductal carcinoma in situ (DCIS). *BMJ Open* 2019; **9**(3): e026797.
10. Francis A, Thomas J, Fallowfield L, et al. Addressing overtreatment of screen detected DCIS; the LORIS trial. *Eur J Cancer* 2015; **51**(16): 2296-303.
11. Elshof LE, Tryfonidis K, Slaets L, et al. Feasibility of a prospective, randomised, open-label, international multicentre, phase III, non-inferiority trial to assess the safety of active surveillance for low risk ductal carcinoma in situ - The LORD study. *Eur J Cancer* 2015; **51**(12): 1497-510.
12. Bartlett JW, Seaman SR, White IR, Carpenter JR, Initiative* AsDN. Multiple imputation of covariates by fully conditional specification: accommodating the substantive model. *Stat Methods Med Res* 2015; **24**(4): 462-87.
13. American Cancer Society (2017) Breast Cancer Facts & Figures 2017–2018. American Cancer Society Inc., Atlanta.
14. Chootipongchaivat S, van Ravesteyn NT, Li X, et al. Modeling the natural history of ductal carcinoma in situ based on population data. *Breast Cancer Res* 2020; **22**(1): 1-12.
15. Ryser MD, Weaver DL, Zhao F, et al. Cancer Outcomes in DCIS Patients Without Locoregional Treatment. *J Natl Cancer Inst* 2019; **111**(9): 952-60.
16. Zavala VA, Bracci PM, Carethers JM, et al. Cancer health disparities in racial/ethnic minorities in the United States. *Br J Cancer* 2021; **124**(2): 315-32.
17. Noone A-M, Lund JL, Mariotto A, et al. Comparison of SEER treatment data with Medicare claims. *Med Care* 2016; **54**(9): e55.
18. Habel LA, Buist DSM. Re: Cancer Outcomes in DCIS Patients Without Locoregional Treatment. *J Natl Cancer Inst* 2019; **112**(2): 214-5.
19. Haffty BG, Carter D, Flynn SD, et al. Local recurrence versus new primary: clinical analysis of 82 breast relapses and potential applications for genetic fingerprinting. *Int J Radiat Oncol Biol Phys* 1993; **27**(3): 575-83.

20. Komoike Y, Akiyama F, Iino Y, et al. Analysis of ipsilateral breast tumor recurrences after breast-conserving treatment based on the classification of true recurrences and new primary tumors. *Breast Cancer* 2005; **12**(2): 104-11.
21. Panet-Raymond V, Truong PT, McDonald RE, et al. True recurrence versus new primary: an analysis of ipsilateral breast tumor recurrences after breast-conserving therapy. *Int J Radiat Oncol Biol Phys* 2011; **81**(2): 409-17.
22. Mannu GS, Wang Z, Broggio J, et al. Invasive breast cancer and breast cancer mortality after ductal carcinoma in situ in women attending for breast screening in England, 1988-2014: population based observational cohort study. *BMJ* 2020; **369**: m1570.
23. Delanian S, Lefaix J-L. Current management for late normal tissue injury: radiation-induced fibrosis and necrosis. *Semin Radiat Oncol* 2007; **17**: 99-107.
24. Bromley HL, Mann GB, Petrie D, Nickson C, Rea D, Roberts TE. Valuing preferences for treating screen detected ductal carcinoma in situ. *Eur J Cancer* 2019; **123**: 130-7.

Supplementary materials

Supplementary Figure 1. State occupation probabilities and transition probabilities for women with low-risk features



Transition probabilities calculated from the multi-state models are visualized for the different matched 1:2 treatment cohorts within the low-risk DCIS subgroup. These figures coincide with the data in Supplementary Tables 1-8. Each curve represents the instantaneous transition rate (or "progression") to the different possible events of interest over time. The distance between the curves represents an individual's probability of being in a specific health state at a specific time point ("state occupancy probability").

Supplementary Table 1: Probability of Remaining in DCIS State (state occupancy probability of "progression-free survival")

	No treatment		All low-risk DCIS patients						Propensity Score-Matched 1:2 low-risk DCIS patients					
			BCS		BCS+RT		MAST		BCS		BCS+RT		MAST	
	Pr.	SE	Pr.	SE	Pr.	Pr.	Pr.	SE	Pr.	SE	Pr.	Pr.	Pr.	SE
Year 0	1.000	0.000	1.000	0.000	1.000	1.000	1.000	0.000	1.000	0.000	1.000	1.000	1.000	0.000
Year 1	0.988	0.006	0.990	0.001	0.995	0.001	0.986	0.002	0.990	0.004	0.993	0.003	0.985	0.006
Year 2	0.975	0.009	0.980	0.002	0.988	0.001	0.979	0.003	0.987	0.005	0.974	0.007	0.982	0.006
Year 3	0.958	0.011	0.969	0.002	0.978	0.002	0.971	0.003	0.976	0.006	0.960	0.008	0.977	0.007
Year 4	0.954	0.012	0.960	0.003	0.967	0.002	0.962	0.004	0.970	0.007	0.948	0.009	0.972	0.008
Year 5	0.934	0.015	0.942	0.003	0.957	0.002	0.951	0.005	0.957	0.009	0.934	0.011	0.955	0.011
Year 6	0.916	0.017	0.926	0.004	0.945	0.003	0.938	0.005	0.943	0.010	0.932	0.011	0.948	0.011
Year 7	0.902	0.018	0.908	0.004	0.932	0.003	0.925	0.006	0.918	0.013	0.919	0.012	0.941	0.012
Year 8	0.882	0.021	0.892	0.005	0.918	0.004	0.914	0.006	0.898	0.015	0.916	0.013	0.937	0.013
Year 9	0.859	0.023	0.873	0.005	0.902	0.004	0.903	0.007	0.885	0.016	0.896	0.015	0.932	0.014
Year 10	0.846	0.024	0.855	0.006	0.882	0.004	0.887	0.008	0.871	0.017	0.881	0.016	0.911	0.017

BCS: breast conserving surgery; BCS+RT: breast conserving surgery + radiotherapy; MAST: mastectomy; Pr.: probability; SE: standard error.

Supplementary Table 2: Transition 1: probability of transitioning from DCIS state to ipsilateral invasive breast cancer ≤5 years post-diagnosis

	No treatment		All low-risk DCIS patients						Propensity Score-Matched 1:2 low-risk DCIS patients					
			BCS		BCS+RT		MAST		BCS		BCS+RT		MAST	
	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE
Year 0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Year 1	0.009	0.005	0.004	0.001	0.000	0.000	0.001	0.001	0.003	0.002	0.000	0.000	0.000	0.000
Year 2	0.009	0.005	0.008	0.001	0.000	0.000	0.001	0.001	0.003	0.002	0.003	0.002	0.000	0.000
Year 3	0.009	0.005	0.011	0.001	0.002	0.000	0.001	0.001	0.007	0.003	0.003	0.002	0.002	0.002
Year 4	0.009	0.005	0.013	0.002	0.003	0.001	0.002	0.001	0.007	0.003	0.003	0.002	0.002	0.002
Year 5	0.009	0.005	0.017	0.002	0.004	0.001	0.003	0.001	0.009	0.004	0.003	0.002	0.005	0.003

BCS: breast conserving surgery; BCS+RT: breast conserving surgery + radiotherapy; MAST: mastectomy; Pr.: probability; SE: standard error.

Supplementary Table 3: Transition 2: probability of transitioning from DCIS state to ipsilateral invasive breast cancer >5 years post-diagnosis

	No treatment		All low-risk DCIS patients						Propensity Score-Matched 1:2 low-risk DCIS patients					
			BCS		BCS+RT		MAST		BCS		BCS+RT		MAST	
	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE
Year 0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Year 1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Year 2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Year 3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Year 4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Year 5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Year 6	0.004	0.004	0.005	0.001	0.003	0.001	0.001	0.001	0.004	0.003	0.000	0.000	0.000	0.000
Year 7	0.004	0.004	0.010	0.002	0.005	0.001	0.001	0.001	0.015	0.006	0.003	0.003	0.000	0.000
Year 8	0.015	0.008	0.014	0.002	0.007	0.001	0.001	0.001	0.017	0.006	0.006	0.004	0.000	0.000
Year 9	0.015	0.008	0.020	0.002	0.009	0.001	0.001	0.001	0.020	0.007	0.006	0.004	0.000	0.000
Year 10	0.021	0.010	0.021	0.002	0.013	0.002	0.002	0.001	0.025	0.008	0.006	0.004	0.005	0.005

BCS: breast conserving surgery; BCS+RT: breast conserving surgery + radiotherapy; MAST: mastectomy; Pr.: probability; SE: standard error.

Supplementary Table 4: Transition 3: probability of transitioning from DCIS state to contralateral invasive breast cancer

	No treatment		All low-risk DCIS patients						Propensity Score-Matched 1:2 low-risk DCIS patients					
			BCS		BCS+RT		MAST		BCS		BCS+RT		MAST	
	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE
Year 0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Year 1	0.003	0.003	0.004	0.001	0.004	0.001	0.011	0.002	0.007	0.003	0.007	0.003	0.011	0.005
Year 2	0.009	0.005	0.006	0.001	0.006	0.001	0.013	0.002	0.007	0.003	0.010	0.004	0.011	0.005
Year 3	0.013	0.006	0.009	0.001	0.010	0.001	0.015	0.002	0.010	0.004	0.017	0.005	0.011	0.005
Year 4	0.013	0.006	0.013	0.002	0.014	0.001	0.018	0.003	0.016	0.005	0.023	0.006	0.014	0.006
Year 5	0.021	0.009	0.017	0.002	0.016	0.001	0.022	0.003	0.018	0.006	0.028	0.007	0.019	0.007
Year 6	0.025	0.010	0.021	0.002	0.020	0.002	0.024	0.003	0.026	0.007	0.028	0.007	0.023	0.008
Year 7	0.035	0.012	0.024	0.002	0.023	0.002	0.028	0.004	0.033	0.008	0.028	0.007	0.026	0.008
Year 8	0.040	0.013	0.028	0.003	0.027	0.002	0.030	0.004	0.039	0.009	0.028	0.007	0.031	0.009
Year 9	0.040	0.013	0.029	0.003	0.029	0.002	0.033	0.004	0.039	0.009	0.024	0.007	0.035	0.010
Year 10	0.046	0.014	0.032	0.003	0.034	0.002	0.037	0.004	0.039	0.009	0.024	0.007	0.035	0.010

BCS: breast conserving surgery; BCS+RT: breast conserving surgery + radiotherapy; MAST: mastectomy; Pr.: probability; SE: standard error.

Supplementary Table 5: Transition 4: probability of transitioning from DCIS state to death without progression

	No treatment		All low-risk DCIS patients						Propensity Score-Matched 1:2 low-risk DCIS patients					
			BCS		BCS+RT		MAST		BCS		BCS+RT		MAST	
	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE
Year 0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Year 1	0.000	0.000	0.002	0.001	0.000	0.000	0.002	0.001	0.000	0.000	0.000	0.000	0.005	0.003
Year 2	0.006	0.005	0.006	0.001	0.005	0.001	0.007	0.002	0.004	0.002	0.012	0.005	0.007	0.004
Year 3	0.020	0.008	0.010	0.001	0.010	0.001	0.013	0.002	0.007	0.004	0.018	0.006	0.009	0.005
Year 4	0.024	0.009	0.014	0.002	0.015	0.001	0.018	0.003	0.007	0.004	0.024	0.006	0.012	0.005
Year 5	0.032	0.010	0.023	0.002	0.022	0.002	0.024	0.003	0.016	0.006	0.032	0.008	0.021	0.008
Year 6	0.040	0.012	0.030	0.003	0.027	0.002	0.035	0.004	0.018	0.006	0.035	0.008	0.025	0.008
Year 7	0.045	0.013	0.039	0.003	0.035	0.002	0.043	0.005	0.026	0.007	0.045	0.010	0.028	0.009
Year 8	0.050	0.014	0.048	0.003	0.042	0.003	0.053	0.005	0.035	0.009	0.045	0.010	0.028	0.009
Year 9	0.068	0.017	0.057	0.004	0.053	0.003	0.058	0.006	0.041	0.010	0.065	0.012	0.028	0.009
Year 10	0.068	0.017	0.067	0.004	0.064	0.003	0.068	0.006	0.048	0.011	0.080	0.014	0.044	0.013

BCS: breast conserving surgery; BCS+RT: breast conserving surgery + radiotherapy; MAST: mastectomy; Pr.: probability; SE: standard error.

Supplementary Table 6: Transition 5: probability of transitioning from ipsilateral invasive breast cancer (≤5 years post-DCIS diagnosis) to death

	No treatment		All low-risk DCIS patients						Propensity Score-Matched 1:2 low-risk DCIS patients					
			BCS		BCS+RT		MAST		BCS		BCS+RT		MAST	
	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE
Year 0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.00	0.00	0.000	0.000	0.000	0.000
Year 1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.00	0.00	0.000	0.000	0.000	0.000
Year 2	0.000	0.000	0.036	0.034	0.000	0.000	0.000	0.000	0.00	0.00	0.000	0.000	0.000	0.000
Year 3	0.000	0.000	0.036	0.034	0.111	0.099	0.000	0.000	0.00	0.00	0.500	0.250	0.000	0.000
Year 4	0.000	0.000	0.052	0.037	0.111	0.099	0.000	0.000	0.00	0.00	0.500	0.250	0.000	0.000
Year 5	0.000	0.000	0.081	0.041	0.111	0.099	0.000	0.000	0.00	0.00	0.500	0.250	0.000	0.000
Year 6	0.000	0.000	0.137	0.047	0.111	0.099	0.000	0.000	0.00	0.00	0.500	0.250	0.000	0.000
Year 7	0.000	0.000	0.165	0.050	0.111	0.099	0.000	0.000	0.00	0.00	0.500	0.250	0.000	0.000
Year 8	0.000	0.000	0.194	0.052	0.111	0.099	0.000	0.000	0.20	0.16	0.500	0.250	0.000	0.000
Year 9	0.000	0.000	0.194	0.052	0.111	0.099	0.500	0.250	0.20	0.16	0.500	0.250	0.000	0.000
Year 10	0.000	0.000	0.211	0.053	0.111	0.099	0.500	0.250	0.20	0.16	0.500	0.250	0.000	0.000

BCS: breast conserving surgery; BCS+RT: breast conserving surgery + radiotherapy; MAST: mastectomy; Pr.: probability; SE: standard error.

Supplementary Table 7: Transition 6: probability of transitioning from ipsilateral invasive breast cancer (>5 years post-DCIS diagnosis) to death

	No treatment		All low-risk DCIS patients						Propensity Score-Matched 1:2 low-risk DCIS patients					
			BCS		BCS+RT		MAST		BCS		BCS+RT		MAST	
	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE
Year 0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Year 1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Year 2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Year 3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Year 4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Year 5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Year 6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Year 7	0.000	0.000	0.000	0.000	0.045	0.043	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Year 8	0.000	0.000	0.023	0.023	0.045	0.043	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Year 9	0.000	0.000	0.057	0.032	0.073	0.050	0.000	0.000	0.125	0.109	0.000	0.000	0.000	0.000
Year 10	0.000	0.000	0.088	0.038	0.099	0.055	0.000	0.000	0.250	0.142	0.000	0.000	0.000	0.000

BCS: breast conserving surgery; BCS+RT: breast conserving surgery + radiotherapy; MAST: mastectomy; Pr.: probability; SE: standard error.

Supplementary Table 8: Transition 7: probability of transitioning from contralateral invasive breast cancer to death

	No treatment		All low-risk DCIS patients						Propensity Score-Matched 1:2 low-risk DCIS patients					
			BCS		BCS+RT		MAST		BCS		BCS+RT		MAST	
	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE	Pr.	SE
Year 0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Year 1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Year 2	0.000	0.000	0.040	0.038	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Year 3	0.000	0.000	0.040	0.038	0.017	0.017	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Year 4	0.000	0.000	0.040	0.038	0.029	0.021	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Year 5	0.250	0.188	0.040	0.038	0.058	0.026	0.025	0.024	0.000	0.000	0.000	0.000	0.000	0.000
Year 6	0.250	0.188	0.040	0.038	0.075	0.028	0.025	0.024	0.000	0.000	0.000	0.000	0.000	0.000
Year 7	0.250	0.188	0.086	0.043	0.075	0.028	0.025	0.024	0.000	0.000	0.000	0.000	0.000	0.000
Year 8	0.357	0.185	0.096	0.044	0.098	0.030	0.025	0.024	0.077	0.071	0.000	0.000	0.000	0.000
Year 9	0.357	0.185	0.145	0.046	0.113	0.031	0.066	0.037	0.077	0.071	0.125	0.109	0.000	0.000
Year 10	0.357	0.185	0.203	0.049	0.120	0.032	0.107	0.045	0.077	0.071	0.125	0.109	0.000	0.000

BCS: breast conserving surgery; BCS+RT: breast conserving surgery + radiotherapy; MAST: mastectomy; Pr.: probability; SE: standard error.


3

CHAPTER 3

Surveillance imaging after primary diagnosis of ductal carcinoma in situ

Danalyn Byng[1,2], Samantha M. Thomas[3,4], Christel N. Rushing[3], Thomas Lynch[5], Anne McCarthy[6], Amanda B. Francescatti[6], Elizabeth S. Frank[7], Ann H. Partridge[7], Alastair M. Thompson[8], Valesca P. Retèl[1,2], Wim H. van Harten[1,2], Lars J. Grimm[9], Terry Hyslop[3,4], E. Shelley Hwang[5], Marc D. Ryser[10,11]

[1]Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands

[2]Health Technology and Services Research Department, Technical Medical Centre, University of Twente, Enschede, The Netherlands

[3]Duke Cancer Institute Biostatistics Shared Resource, Duke University, Durham NC, USA

[4]Department of Biostatistics & Bioinformatics, Duke University Medical Center, Durham, NC, USA

[5]Division of Surgical Sciences, Department of Surgery, Duke University Medical Center, Durham, NC, USA

[6]Cancer Programs, American College of Surgeons, Chicago, IL, USA

[7]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA

[8]Department of Surgery, Baylor College of Medicine, Houston, TX, USA

[9]Department of Radiology, Duke University Medical Center, Durham, NC, USA

[10]Department of Population Health Sciences, Duke University Medical Center, Durham, NC, USA

[11]Department of Mathematics, Duke University, Durham NC, USA

Accepted for publication in Radiology (in press)

Key Results

- Among 12,559 women treated with breast conserving surgery for ductal carcinoma in situ, 1 in 2 did not consistently adhere to guideline-recommended annual surveillance imaging over 5 years.
- Compared with White women, surveillance was lower in Black women (adjusted odds ratio [OR], 0.80; P<.001) and Hispanic women (OR, 0.82; P=.004).
- The 6-year rate of detection of invasive recurrence was higher in women who received surveillance imaging in the first year after treatment (absolute rate difference: 0.5%, P=.03).

Summary Statement

Approximately half of eligible women with primary ductal carcinoma in situ did not adhere to imaging surveillance guidelines over 5 years following treatment, with significant racial disparities in adherence rates.

Abstract

Background

Guidelines recommend annual imaging surveillance after diagnosis with ductal carcinoma in situ (DCIS). Guideline adherence has not been characterized in a contemporary cohort.

Purpose

Identify uptake and determinants of surveillance imaging in a large cohort of women treated for DCIS.

Materials and Methods

A stratified random sample of women treated with breast conserving surgery for primary DCIS between 2008-2015 was retrospectively selected from 1,330 U.S. facilities. Imaging exams were recorded from date of diagnosis until first distant recurrence, death, loss to follow-up or end of study (November 2018). Imaging following treatment was categorized into ten 12-month periods, starting 6 months after diagnosis. Primary outcome was per-period receipt of asymptomatic surveillance imaging (mammography, MRI, or US). Secondary outcome was diagnosis with ipsilateral invasive breast cancer. Multivariable logistic regression with repeated measures and generalized estimating equations was used to model receipt of imaging. Cumulative rates of diagnosis with ipsilateral invasive breast cancer were compared between women who did and did not receive imaging in the 6- to 18-month period following diagnosis using inverse probability weighted Kaplan-Meier estimators.

Results

12,559 women (median age, 60 years; IQR: 52, 69) were evaluated. Uptake of surveillance imaging was 75% in the first period and decreased over time (P<.001). Over the first five years, 51% of women received consistent annual surveillance. Compared with White women, surveillance was lower in Black women (adjusted odds ratio [OR], 0.80; 95% CI: 0.74, 0.88; P<.001) and Hispanic women (OR, 0.82; 95% CI: 0.72, 0.94; P=.004). Women who received surveillance in the first period had a higher 6-year rate of diagnosis of invasive breast cancer (1.6%; CI: 1.3%, 1.9%) than those who did not (1.1%; CI: 0.7%, 1.4%; difference: 0.5%; CI: 0.1%, 1.0%; P=.03).

Conclusions

Half of women did not consistently adhere to imaging surveillance guidelines over 5 years following treatment, with racial disparities in adherence rates.

Introduction

Ductal carcinoma in situ (DCIS) is a primarily screen-detected and non-obligate precursor lesion of invasive breast cancer.[1,2] Current standard of care options for primary DCIS are breast conserving surgery (BCS) with or without radiation treatment, or mastectomy, with approximately two-thirds of women electing BCS.[3] Despite excellent disease-specific survival,[3] women who receive BCS for DCIS have a recurrence rate exceeding 20% by 10 years.[4-6]

Following BCS for DCIS or invasive cancer, national guidelines from the American College of Radiology (ACR),[7,8] American Society of Clinical Oncology (ASCO),[9] and the National Comprehensive Cancer Network (NCCN)[10] recommend bilateral annual mammography for women with any history of breast cancer who have been treated with BCS, with the first surveillance screen recommended within six to twelve months after completion of initial locoregional treatment, and no end date specified.[7-10] For patients with BCS with additional breast cancer risk factors, it is further recommended that mammography be supplemented by MRI or breast US.[11]

Prior studies of surveillance imaging among contemporary cohorts of women with DCIS have been either limited in size or focused on specific patient groups.[12,13] The most comprehensive study to date reported patterns of surveillance mammography in a cohort of 3,037 women diagnosed between 1990 and 2001;[14] however, this study was limited to findings from two health systems who had a high rate of insured women. In addition, given the study's historic nature and specific focus on mammography screening, there is a need for more contemporary, real-world assessments of surveillance imaging, including MRI and US.[15,16]

To assess the clinical utility of current surveillance guidelines, it is important to understand how surveillance behaviors impact cancer outcomes and survival. Such knowledge has the potential to inform personalized screening approaches that balance the benefit of early detection of recurrence against the potential harms and costs incurred by frequent screening.[17] Because disease-specific mortality after primary DCIS is very low,[3,18] a common endpoint for such analyses is diagnosis with ipsilateral invasive breast cancer.[19]

The aim of our study was to identify the uptake and determinants of surveillance imaging after treatment for DCIS in a contemporary U.S.-based cohort of women. To this end, we collected data through a National Cancer Database Special Study, where all 1,330 participating sites locally reviewed primary source documentation and uploaded data to a central database. This study design also allowed us to

determine potential differences in the rate of diagnosis of ipsilateral invasive breast cancer according to adherence to clinical surveillance guidelines.

Methods

Study Cohort

Through a National Cancer Database (NCDB) Special Study, in collaboration with the American College of Surgeons, we conducted a retrospective study of a stratified random sample of women diagnosed with biopsy-confirmed primary DCIS in 2008-2015 from 1,330 Commission on Cancer-accredited facilities in the United States. To construct the sampling frame, we used the 2015 NCDB Patient User File (PUF). Key eligibility criteria included a primary DCIS diagnosis (sequence number 00 or 01) and known modality and timing of first course of locoregional treatment. Based on the PUF, we assigned women to one of two groups: breast conserving surgery (BCS) or mastectomy within six months of diagnosis, or no locoregional treatment within six months of diagnosis. In-depth longitudinal data on imaging exams, treatments, and cancer outcomes was collected by local cancer registrars for up to 20 women per facility, including 10 (or as many as available) women without initial locoregional treatment. For the 25,817 women included in the sampling frame, we received 21,167 abstracted records of sufficient date accuracy. This study was approved by the Duke University Health System institutional review board and is compliant with the Health Insurance Portability and Accountability Act. Written informed consent was waived.

For the current analysis we included only women who received BCS within 6 months. Women who received a mastectomy within 6 months or had incomplete records on locoregional treatment, subsequent breast events or imaging were excluded. Patient characteristics as obtained from the NCDB PUF [20] included age at diagnosis; year of diagnosis; race; Hispanic ethnicity; Charlson-Deyo comorbidity index; insurance status; summaries of educational attainment and household income in each patient's area of residence; facility type; and metropolitan area indicator. Tumor characteristics as abstracted through the NCDB Special Study included method of DCIS detection; hormone receptor status (positive if estrogen and/or progesterone- positive; negative if estrogen and progesterone-negative); nuclear grade; presence of comedonecrosis; pathologic tumor size, and surgical margin status on first BCS. Treatment characteristics included receipt of re-excision surgery; receipt of radiation treatment within six months of surgery, and initiation of endocrine therapy within 1 year of diagnosis.

In the final study cohort, the only baseline covariate with any missing entries was pathologic tumor size (n=6). Missing entries as well as NCDB PUF entries labeled as "not available" or "unknown" were grouped into a single "unknown" category.

Imaging Exams

Imaging exams were recorded starting on the date of diagnosis and until the date of the first distant recurrence, death, loss to follow-up or end of the study period (November 2018), whichever came first. For each imaging exam, we collected the following: time since diagnosis; reason for the exam: asymptomatic surveillance (i.e., imaging in the absence of new symptoms), evaluation of new symptoms, and evaluation of established diagnosis; laterality; and imaging modality (mammography, MRI, and US).

Follow-up after diagnosis was categorized into discrete 12-month intervals, starting at 6 months after diagnosis and up to the last period throughout which the patient was free of a new breast cancer diagnosis. For each surveillance period, we recorded the number and modality of asymptomatic surveillance exams received.

Ipsilateral invasive breast cancer

Time to initial diagnosis of ipsilateral invasive breast cancer was abstracted from patient records and right-censored at the time of death, loss of follow up or end of study, whichever came first. For consistency, the following were included as events: ipsilateral lymph node metastasis without preceding ipsilateral invasive breast cancer, and distant metastasis without preceding invasive cancer or lymph node metastasis (of either laterality). No censoring was applied at the occurrence of ipsilateral DCIS, contralateral DCIS, or contralateral invasive breast cancer.

Outcome measures

The primary outcome measure was per-period receipt of any asymptomatic surveillance imaging, including mammography, MRI and US. The secondary outcome measure was diagnosis with ipsilateral invasive breast cancer.

Statistical Analysis

We reported patient, tumor and treatment characteristics with and without weighting by survey design weights and we used the reverse Kaplan-Meier method[21] to compute unweighted median (IQR) follow-up. The time difference between surveillance imaging exams was visualized using Gaussian kernel density estimators. In each of the ten 12-month surveillance periods, we characterized the cross-sectional uptake of the different imaging modalities. The Cochrane-Armitage test was used to ascertain trends in surveillance imaging during follow-up.[22]

Women were classified based on their longitudinal imaging uptake over the first five surveillance periods: *consistent screeners* if they had at least one surveillance imaging study (of any modality) during each available period before an event; *non-screeners* if they didn't receive any screening before an event; and *inconsistent screeners* otherwise.

We modeled the longitudinal uptake of any asymptomatic surveillance imaging using repeated measures multivariable logistic regression with generalized estimating equations.[23] The models were adjusted for patient, tumor, and treatment characteristics and weighted by the survey design weights; where applicable, the "unknown" categories were included in the regression. We used a first order autoregressive correlation structure to account for the clustering of repeated visits over subsequent surveillance periods for each patient. We evaluated several clinically plausible interactions, e.g., between radiation treatment and tumor size and margin status. Because none of the examined interactions reached statistical significance, the final model included main effects only.

To explore the relationship between surveillance imaging and a subsequent invasive cancer diagnosis, we used Kaplan-Meier estimators to estimate the weighted and unweighted cumulative incidence of ipsilateral invasive breast cancer diagnosis starting 12 months after receipt of BCS. Women with unknown endocrine therapy status within the first year of diagnosis were excluded from the analysis. To compare the diagnosis rates between women who did and did not receive surveillance imaging within 12 months of BCS, we used inverse probability of treatment weighting-adjusted Kaplan-Meier estimators and log-rank tests.[24] The corresponding propensity scores were calculated using a multivariable logistic regression model, adjusted for patient, tumor, and treatment characteristics, and clustered by reporting facility. We assessed covariate imbalances between the groups before and after weighting using standardized mean differences (Supplementary Figure S1). Absolute differences in cumulative rates of ipsilateral invasive breast cancer diagnosis were calculated every 6 months, starting at 18 months after surgery and up to 10 years; 95% CIs were calculated by bootstrap.[25]

All statistical analyses were performed with R version 4.1.0 (R Core Team 2021), using the geepack (v1.3-2), longCatEDA (v0.31), PSweight (v1.15), and RISCA (v0.9) packages. A P-value <.05 was considered statistically significant.

Results

Characteristics of Study Cohort

The analytic cohort was comprised of 12,559 women (Figure 1); 8,989 of 12,559 (72%, unweighted) received radiation treatment in addition to BCS (Table 1). Median age at diagnosis was 60 years (IQR 52–69) and median follow up was 5.8 years (IQR 4.0–7.9). Most women were of White (10,564 of 12,559, 84%) or Black (1,371 of 12,559, 11%) race and had either government (5,155 of 12,559, 41%) or private (7,042 of 12,559, 56%) insurance. Most DCIS was screen-detected (92%), of non-high grade (55%) and hormone-receptor positive (82%). Weighted and unweighted proportions were comparable across covariates (Table 1).

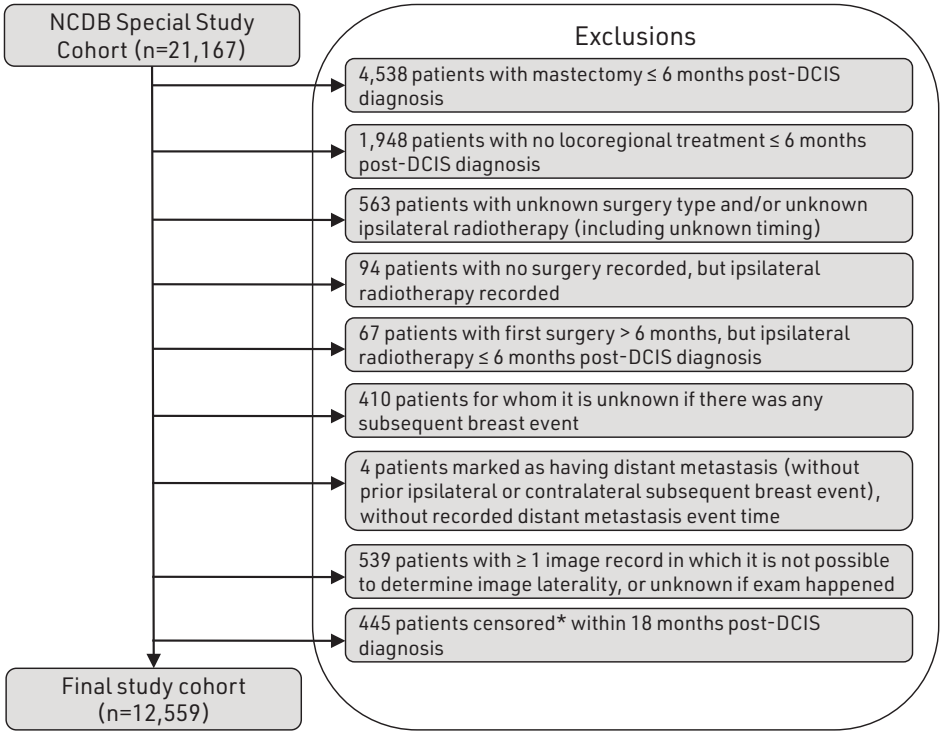


Figure 1. Curation of final study cohort

DCIS: ductal carcinoma in situ, NCDB: National Cancer Database.

*Women were censored from a surveillance period if they died or had a new breast cancer diagnosis before the end of the period.

A total of 85,057 imaging exams were abstracted (Supplementary Table S1). Mammography was the most prevalent imaging modality (82.8%), followed by US (10.6%) and MRI (6.1%). Most imaging studies were coded as asymptomatic surveillance imaging (78.0%), and most were bilateral (68.0%).

As illustrated by the distribution of inter-screen intervals for asymptomatic surveillance exams of any modality (Figure 2A), both annual and biannual screening patterns were identified.

Uptake of Surveillance Imaging

Only 9,394 of 12,559 (75%) of women underwent asymptomatic surveillance imaging of any modality in the first follow-up period between 6 and 18 months after diagnosis (Figure 2B). The dominant screening modality in this period was mammography alone (65.6%), followed by combination screening of mammography with either US or MRI (7.4%). Throughout subsequent follow-up periods, mammography remained the dominant screening modality. Uptake of surveillance imaging by any modality decreased over time (P<.001), dropping to 5,562 of 8,199 (68%) women in the fourth and 1,231 of 2,095 (59%) women in the eighth follow-up period, respectively.

Overall, 6,469 of 12,559 (52%) of women were classified as consistent screeners, 4,185 (33%) as inconsistent screeners, and 1,905 (15%) as non-screeners (Figure 2C). Among the 9,373 women who received any surveillance imaging during the first period, 6,458 (69%) were classified as consistent screeners over the first five periods.

Factors Associated with Surveillance Imaging

Several sociodemographic factors were associated with uptake of surveillance imaging, including race, ethnicity, and insurance status in the multivariable analysis (Figure 3). Uptake was lower among Black women than White women (odds ratio [OR], 0.80; 95% CI: 0.74, 0.88; P<.001) and among Hispanic women than non-Hispanic women (OR, 0.82; 95% CI: 0.72, 0.94; P=.004), respectively. Women with private insurance were more likely to receive imaging than women with government insurance (OR, 1.31; 95% CI: 1.22, 1.41; P<.001).

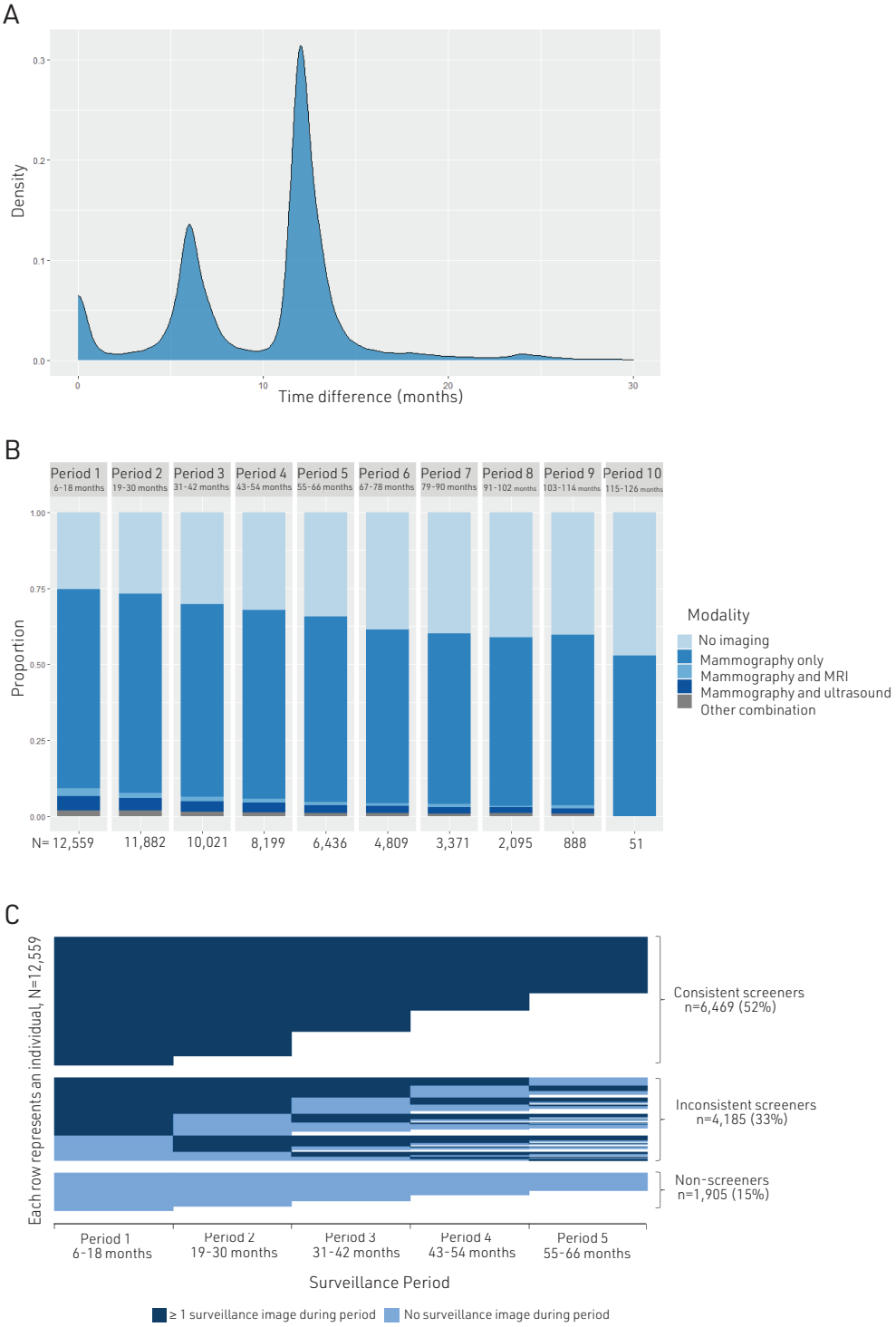

A
Density
0.3
0.2
0.1
0.0
0
10
20
30
Time difference (months)
B
Period 1
6-18 months
Period 2
19-30 months
Period 3
31-42 months
Period 4
43-54 months
Period 5
55-66 months
Period 6
67-78 months
Period 7
79-90 months
Period 8
91-102 months
Period 9
103-114 months
Period 10
115-126 months
Proportion
1.00
0.75
0.50
0.25
0.00
N= 12,559
11,882
10,021
8,199
6,436
4,809
3,371
2,095
888
51
Modality
No imaging
Mammography only
Mammography and MRI
Mammography and ultrasound
Other combination
C
Each row represents an individual, N=12,559
Consistent screeners
n=6,469 (52%)
Inconsistent screeners
n=4,185 (33%)
Non-screeners
n=1,905 (15%)
Period 1
6-18 months
Period 2
19-30 months
Period 3
31-42 months
Period 4
43-54 months
Period 5
55-66 months
Surveillance Period
≥ 1 surveillance image during period
No surveillance image during period

Figure 2. Surveillance imaging.

Panel A: Distribution of elapsed time between surveillance imaging exams for the entire cohort, represented with a kernel density plot; Panel B: Surveillance imaging by modality. For each period, the number (N) of women with available follow-up is shown below the bar chart; Panel C: Screening group classification based on surveillance imaging uptake in first five surveillance periods.

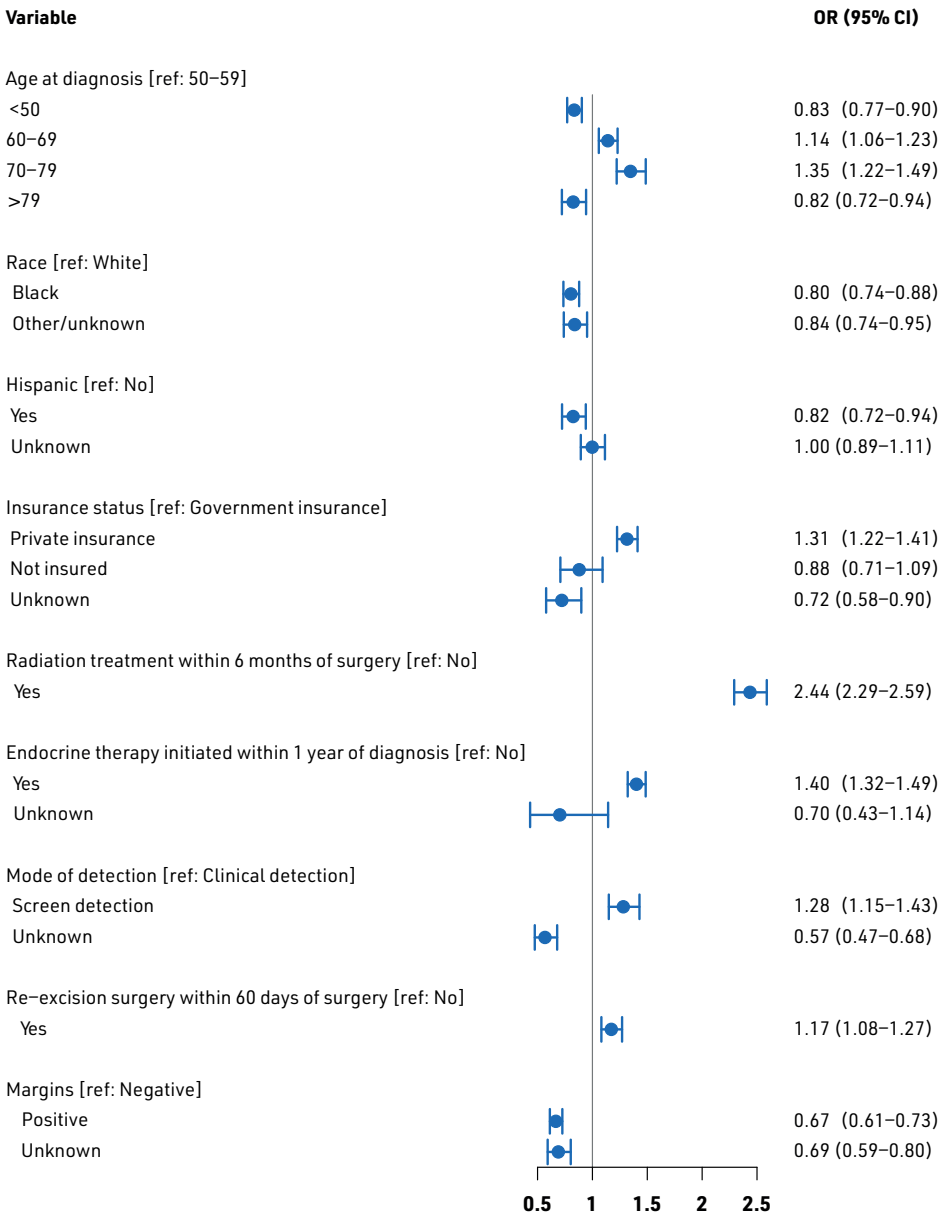


Figure 3. Predictors of surveillance screening over time. See Supplementary Figure S2 for full model result.

Receipt of adjuvant therapy after BCS was the strongest predictor for surveillance imaging, despite lower rates of recurrence anticipated for those who were treated with adjuvant therapy. Women who received adjuvant radiation treatment had higher odds of receiving surveillance screening (OR, 2.44; 95% CI: 2.29, 2.59; P<.001) than women treated with BCS alone. Similarly, women who initiated endocrine therapy

within one year of diagnosis were more likely to undergo surveillance screening than those who did not (OR, 1.40; 95% CI: 1.32, 1.49; P<.001).

Clinical variables associated with increased uptake of surveillance imaging included screen-detection of DCIS (OR, 1.28; 95% CI: 1.15, 1.43; P<.001; compared with non-screen-detection of DCIS) and receipt of re-excision surgery within 60 days (OR, 1.17; 95% CI: 1.08, 1.27; P<.001). The complete model results are found in Supplementary Figure S2.

Early Surveillance Imaging and the Rate of Diagnosis with Ipsilateral Invasive Breast Recurrence

Among the 12,519 women included in the recurrence analysis, 190 (2%) were diagnosed with an ipsilateral invasive breast cancer during follow-up (Supplementary Table 2). Among these 190 women, median time to event was 3.9 years (IQR: 2.5, 6.0) among women who received early surveillance imaging, and 4.2 years (IQR: 3.0, 6.6) among those who did not. In the unadjusted analysis there was no evidence of a difference in the rate of diagnosis between women who did (n=8,821) or did not (n=3,694) receive surveillance imaging within 12 months of BCS (log-rank test: *P*=.64; Figure 4A). In the inverse probability of treatment weighting (IPTW)-adjusted analysis, the cumulative diagnosis rates diverged over time, with a higher rate of diagnosis among women who did receive surveillance screening in the first period than those who did not (Figure 4B). At 6 years after diagnosis, the cumulative rates of diagnosis with ipsilateral invasive breast cancer were 1.6% (95% CI: 1.3, 1.9%) and 1.1% (95% CI: 0.7, 1.4%) in those who did and did not receive early surveillance imaging (rate difference: 0.5%; 95% CI: 0.1, 1.0%, P=.03). Across the extended follow-up there was no detectable difference between the two groups (IPTW-adjusted log-rank test: *P*=.08) (Fig 4C).

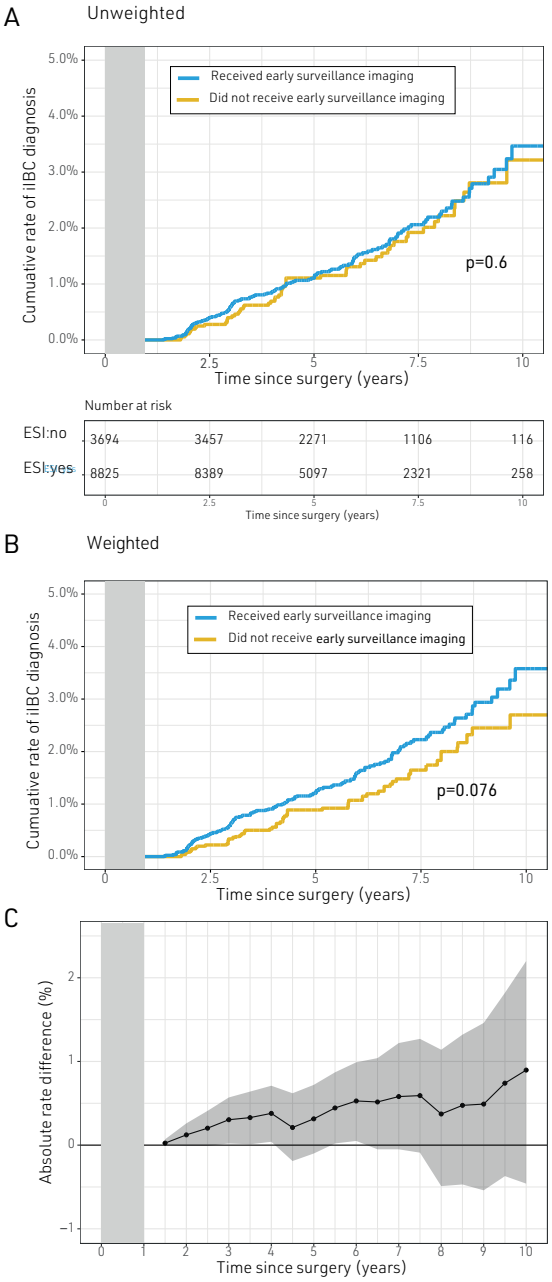


Figure 4. Cumulative rate of ipsilateral invasive breast cancer (iIBC) diagnosis in women who did and did not receive early surveillance imaging. (A) The Kaplan-Meier curve indicates the unweighted cumulative rate of iIBC diagnosis for women who did and did not undergo surveillance imaging within 12 months of surgery (shaded period). (B) As in A, but with adjustment for inverse probability of treatment weights. (C) Absolute rate difference between the Kaplan-Meier curves in B, with pointwise 95% confidence intervals (shaded).

Discussion

Adherence to recommended annual surveillance imaging after diagnosis with ductal carcinoma in situ (DCIS) has not been characterized in a contemporary cohort. Here we retrospectively identified uptake and determinants of surveillance in 12,559 women treated with breast conserving surgery (BCS) for primary DCIS. Within 6 to 18 months after diagnosis, 75% of women received surveillance imaging of any modality, and adherence to annual imaging decreased over time (P<.001). Compared to White women, surveillance uptake was lower in Black (OR, 0.80, P<.001) and Hispanic (OR, 0.82, P=.004) women. Six years after diagnosis, the rate of detection with ipsilateral invasive cancer was lower in women who adhered to annual screening in the first period (rate difference: 0.5%; P=.03).

Uptake of surveillance imaging in this cohort was lower than that reported in women undergoing BCS for DCIS between 1990 and 2001,[14] and also lower than that reported in women treated for invasive breast cancer.[26,27] While the disparities in surveillance imaging in our study mirror previous findings in women with a history of invasive breast cancer,[27,28] these disparities have been rarely documented in women with a history of DCIS. In a previous study among private health plan members diagnosed with DCIS between 1990 and 2001, Nekhlyudov and colleagues found similar differences in surveillance uptake by age, but not by race and ethnicity.[14] In a small study comparing surveillance mammography between Latina and non-Latina white women, Lopez and colleagues found lower uptake among Spanish-speaking but not English-speaking Latinas compared with non-Latinas.[12]

Surveillance disparities by race and ethnicity are particularly concerning because they may reflect limited access to care, a long-standing and systemic inequity in the U.S. healthcare system.[29] This is further complicated by racial disparities in access to breast cancer diagnosis and treatment, and breast cancer screening recommendations. These racial disparities put some women, particularly Black women, at a disadvantage.[30-32] The complicated landscape of inequitable access to care in the US means that not all women diagnosed with DCIS will be able to adhere to recommended annual surveillance imaging.[33]

Given the heterogeneous uptake of surveillance imaging across patient groups, it is critical to understand the downstream consequences of non-adherence to clinical guidelines. Our findings suggest that invasive in-breast recurrences may be found earlier in women who adhere to the guidelines. Indeed, thanks to successful balancing of measured confounders, the differences in 6-year detection rates of

ipsilateral invasive cancer were less likely due to differences in underlying risk factors, and more likely due to more intense surveillance, which in turn increases the chance of detecting an invasive recurrence earlier. Early screening was a good proxy for adherence to annual screening (69% of early screeners were consistent screeners thereafter), which suggests that the time to detection of invasive recurrences may be shorter among women with DCIS who adhere to the screening guidelines.

Interestingly, the differences in ipsilateral invasive breast cancer detection rates only emerged when adjusting for potential confounders through inverse probability weighting. A possible explanation for this may be that the underlying degree of healthcare access drives both primary treatment and subsequent surveillance. For example, while radiation treatment is associated with a lower risk of invasive recurrence,[34] it was also strongly associated with more intense surveillance imaging in our study, which in turn increased the rate of ipsilateral invasive breast cancer detection. Due to these masking effects, adjustment for covariates related to health care access was necessary to uncover the true relationship between surveillance imaging and the detection of ipsilateral invasive breast cancer.

Our study has limitations. First, data fields collected through the National Cancer Database Special Study were abstracted at participating facilities, where local registrars performed in-depth review of medical records. This may have introduced heterogeneity in accuracy and completeness of records. Second, we were unable to abstract known risk factors for invasive breast cancer that may have played a role in surveillance imaging uptake, including BRCA status, use of hormone replacement therapy, and family history of breast cancer. Consequently, we were unable to determine whether the use of supplemental MRI and US were medically indicated. Third, a complete characterization of the relationship between surveillance imaging and detection of ipsilateral invasive breast cancer is beyond the scope of the current study.

In conclusion, our findings highlight individual- and system-level opportunities to devise more targeted surveillance recommendations that reduce variability, maximize health outcomes, and ultimately increase the value of care. In general, over-utilization of cancer imaging has not been well documented,[35] and may constitute an important consideration in addition to identifying under-utilization of surveillance, as we consider a more risk-based approach to surveillance imaging. While the role of treatment in mitigating the risk of ipsilateral invasive breast cancer in women with DCIS is well documented,[34,36-41] the interplay between treatment, surveillance and outcomes is complex, and further analysis is needed to determine

why patients at a low risk for recurrence appear to have high adherence to imaging surveillance. As such, our study underscores the need for more intentional measures to provide equitable systems-based approaches to imaging surveillance in women following BCS for DCIS.

Table 1. Demographic and Clinical Characteristics of Study cohort. Patient characteristics are shown unweighted and weighted by sampling weights

Characteristic	Unweighted N (%)	Weighted %
Follow-up years (median, IQR)	5.8 (4.0-7.9)	-
Radiation treatment within 6 months of surgery		
Yes	8,989 (72 %)	67 %
No	3,570 (28 %)	33 %
Endocrine therapy initiated within 1 year of diagnosis		
Yes	6,431 (51 %)	48 %
No	6,093 (49 %)	51 %
Unknown	35 (0.3 %)	0.3 %
Age at diagnosis		
<50	2,318 (19 %)	20 %
50-59	3,642 (29 %)	30 %
60-69	3,743 (30 %)	29 %
70-79	2,171 (17 %)	16 %
>79	685 (6 %)	5 %
Year of diagnosis		
2008-11	6,140 (49 %)	51 %
2012-15	6,419 (51 %)	50 %
Insurance status		
Government insurance	5,155 (41 %)	38 %
Private insurance	7,042 (56 %)	58 %
Not insured	190 (2 %)	2 %
Unknown	172 (1 %)	2 %
Charlson-Deyo Comorbidity Index		
0	9,639 (77 %)	78 %
1	2,048 (16 %)	16 %
≥2	872 (7 %)	6 %
Race		
White	10,564 (84 %)	81 %
Black	1,371 (11 %)	13 %
Other*/Unknown	624 (5 %)	6 %
Hispanic		
No	11,354 (90 %)	90 %
Yes	590 (5 %)	5 %
Unknown	615 (5 %)	5 %
Fraction of adults in residing in zip code who did not graduate high school		
<21%	10,867 (87 %)	87 %

Characteristic	Unweighted N (%)	Weighted %
≥21%	1,692 (14 %)	13 %
Median household income in residing zip code		
<$48,000	4,725 (38 %)	33 %
≥$48,000	7,834 (62 %)	67 %
Metropolitan area with >250,000 residents		
Yes	8,357 (67 %)	77 %
No	4,202 (34 %)	23 %
Facility type		
Comprehensive Community Cancer Program	5,326 (42 %)	49 %
Academic/Research Program	2,064 (16 %)	30 %
Community Cancer Program	3,570 (28 %)	9 %
Integrated Network Cancer Program	1,398 (11 %)	11 %
Unknown	201 (2 %)	2 %
US geographic location		
Midwest	1,985 (16 %)	15 %
North East	3,869 (31 %)	25 %
South	3,859 (31 %)	35 %
West	2,645 (21 %)	22 %
Unknown	201 (2 %)	2 %

*Other race category includes: American Indian, Aleutian, or Eskimo, Asian Indian, Chinese, Fiji Islander, Filipino, Guamanian, Hawaiian, Japanese, Kampuchean, Korean, Micronesian, Pacific Islander, Pakistani, Samoan, Thai, Tongan, and Vietnamese. Information on NCDB PUF data fields is found in [20].

Table 2. Breast Cancer Characteristics. Patient characteristics are shown unweighted and weighted by sampling weights

Characteristic	Unweighted N (%)	Weighted %
Hormone receptor status*		
Positive	10,244 (82 %)	82 %
Negative	1,430 (11 %)	11 %
Not ordered/done/unknown	885 (7 %)	6 %
Mode of DCIS detection		
Screening	11,550 (92 %)	92 %
Palpation	519 (4 %)	4 %
Other (nipple discharge, skin or nipple changes)	286 (2 %)	2 %
Unknown	204 (1 %)	2 %
DCIS grade		
Grade I	2,296 (18 %)	17 %
Grade II	4,591 (37 %)	38 %
Grade III	4,679 (37 %)	38 %
Unknown	993 (8.0 %)	7 %
Presence of comedonecrosis		
Present	4,020 (32 %)	31 %
Absent	6,846 (55 %)	56 %
Unknown	1,693 (14 %)	13 %
DCIS size		
≤2 cm	5,219 (42 %)	43 %
2-5 cm	3,153 (25 %)	24 %
>5 cm	1,878 (15 %)	16 %
Unknown	2,309 (18 %)	188 %
Margins		
Negative	10,458 (83 %)	81 %
Positive	1,770 (14 %)	16 %
Unknown	331 (3 %)	3 %
Re-excision surgery within 60 days of surgery		
Yes	2,237 (18 %)	20 %
No	10,322 (82 %)	81 %

DCIS: ductal carcinoma in situ, cm: centimeter, mm: millimeter

*Positive: estrogen and/or progesterone receptor-positive, negative: estrogen and progesterone receptor-negative.

References

1. Erbas B, Provenzano E, Armes J, Gertig D. The natural history of ductal carcinoma in situ of the breast: a review. *Breast Cancer Res Treat* 2006; **97**(2): 135-44.
2. Jones JL. Overdiagnosis and overtreatment of breast cancer: progression of ductal carcinoma in situ: the pathological perspective. *Breast Cancer Res* 2006; **8**(2): 204.
3. Worni M, Akushevich I, Greenup R, et al. Trends in treatment patterns and outcomes for ductal carcinoma in situ. *J Natl Cancer Inst* 2015; **107**(12).
4. Mannu GS, Wang Z, Broggio J, et al. Invasive breast cancer and breast cancer mortality after ductal carcinoma in situ in women attending for breast screening in England, 1988-2014: population based observational cohort study. *BMJ* 2020; **369**: m1570.
5. Van Zee KJ, Subhedar P, Olcese C, Patil S, Morrow M. Relationship Between Margin Width and Recurrence of Ductal Carcinoma In Situ: Analysis of 2996 Women Treated With Breast-conserving Surgery for 30 Years. *Ann Surg* 2015; **262**(4): 623-31.
6. Solin LJ, Gray R, Hughes LL, et al. Surgical Excision Without Radiation for Ductal Carcinoma in Situ of the Breast: 12-Year Results From the ECOG-ACRIN E5194 Study. *J Clin Oncol* 2015; **33**(33): 3938-44.
7. Monticciolo DL, Newell MS, Hendrick RE, et al. Breast cancer screening for average-risk women: recommendations from the ACR commission on breast imaging. *J Am Coll Radiol* 2017; **14**(9): 1137-43.
8. Mainiero MB, Moy L, Baron P, et al. ACR Appropriateness Criteria® breast cancer screening. *J Am Coll Radiol* 2017; **14**(11): S383-S90.
9. Runowicz CD, Leach CR, Henry NL, et al. American Cancer Society/American Society of Clinical Oncology Breast Cancer Survivorship Care Guideline. *CA Cancer J Clin* 2016; **66**(1): 43-73.
10. Gradishar WJ, Anderson BO, Balassanian R, et al. Breast cancer, version 4.2017, NCCN clinical practice guidelines in oncology. *JNCCN* 2018; **16**(3): 310-20.
11. Monticciolo DL, Newell MS, Moy L, Niell B, Monsees B, Sickles EA. Breast cancer screening in women at higher-than-average risk: recommendations from the ACR. *J Am Coll Radiol* 2018; **15**(3): 408-14.
12. López ME, Kaplan CP, Nápoles AM, et al. Ductal carcinoma in situ (DCIS): posttreatment follow-up care among Latina and non-Latina White women. *J Cancer Surviv* 2013; **7**(2): 219-26.
13. Jones T, Duquette D, Underhill M, et al. Surveillance for cancer recurrence in long-term young breast cancer survivors randomly selected from a statewide cancer registry. *Breast Cancer Res Treat* 2018; **169**(1): 141-52.
14. Nekhlyudov L, Habel LA, Achacoso NS, et al. Adherence to long-term surveillance mammography among women with ductal carcinoma in situ treated with breast-conserving surgery. *J Clin Oncol* 2009; **27**(19): 3211-6.
15. Henderson LM, Ichikawa L, Buist DSM, et al. Patterns of Breast Imaging Use Among Women with a Personal History of Breast Cancer. *J Gen Intern Med* 2019; **34**(10): 2098-106.
16. Wernli K, Brandzel S, Buist D, et al. Is breast MRI Better at finding second breast cancers than mammograms alone for breast cancer survivors? [Internet] Washington (DC): Patient-Centered Outcomes Research Institute (PCORI); 2019 May. Available from: https://www.ncbi.nlm.nih.gov/books/NBK554228/ doi: 10.25302/5.2019.CE.13046656
17. Kwan J, Croke J, Panzarella T, et al. Personalizing post-treatment cancer care: a cross-sectional survey of the needs and preferences of well survivors of breast cancer. *Current Oncology* 2019; **26**(2): 138-46.

18. Sagara Y, Mallory MA, Wong S, et al. Survival Benefit of Breast Surgery for Low-Grade Ductal Carcinoma In Situ: A Population-Based Cohort Study. *JAMA Surg* 2015; **150**(8): 739-45.
19. Solin LJ. Management of Ductal Carcinoma In Situ (DCIS) of the Breast: Present Approaches and Future Directions. *Curr Oncol Rep* 2019; **21**(4): 33.
20. National Cancer Database. PUF data dictionary items. http://ncdbpuf.facs.org/node/259. Accessed March 14, 2018.
21. Shuster JJ. Median follow-up in clinical trials. *J Clin Oncol* 1991; **9**(1): 191-2.
22. Armitage P. Tests for linear trends in proportions and frequencies. *Biometrics* 1955; **11**(3): 375-86.
23. Yan J, Fine J. Estimating equations for association structures. *Stat Med* 2004; **23**(6): 859-74.
24. Xie J, Liu C. Adjusted Kaplan–Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Stat Med* 2005; **24**(20): 3089-110.
25. Efron B, Tibshirani RJ. An introduction to the bootstrap: CRC press; 1994.
26. Ruddy KJ, Sangaralingham L, Freedman RA, et al. Adherence to guidelines for breast surveillance in breast cancer survivors. *JNCCN* 2018; **16**(5): 526-34.
27. Adesoye T, Schumacher JR, Neuman HB, et al. Use of breast imaging after treatment for locoregional breast cancer (AFT-01). *Ann Surg Oncol* 2018; **25**(6): 1502-11.
28. Advani P, Advani S, Nayak P, et al. Racial/ethnic disparities in use of surveillance mammogram among breast cancer survivors: a systematic review. *J Cancer Surviv* 2021: 1-17.
29. Patel MM, Parikh JR. Patient Diversity in Breast Imaging: Barriers and Potential Solutions. *Journal of Breast Imaging* 2020; **3**(1): 98-105.
30. Mootz A, Arjmandi F, Dogan BE, Evans WP. Health Care Disparities in Breast Cancer: The Economics of Access to Screening, Diagnosis, and Treatment. *J Breast Imaging* 2020; **2**(6): 524-9.
31. Rebner M, Pai VR. Breast cancer screening recommendations: African American women are at a disadvantage. *J Breast Imaging* 2020; **2**(5): 416-21.
32. Chapman CH, Schechter CB, Cadham CJ, et al. Identifying Equitable Screening Mammography Strategies for Black Women in the United States Using Simulation Modeling. *Ann Intern Med* 2021; **174**(12): 1637-46.
33. Strong EA, Stark A, Newman LA. Disparities in DCIS Detection and Outcomes Related to Race/Ethnicity. Ductal Carcinoma In Situ and Microinvasive/Borderline Breast Cancer: Springer; 2015: 161-6.
34. Wapnir IL, Dignam JJ, Fisher B, et al. Long-term outcomes of invasive ipsilateral breast tumor recurrences after lumpectomy in NSABP B-17 and B-24 randomized clinical trials for DCIS. *J Natl Cancer Inst* 2011; **103**(6): 478-88.
35. Baxi SS, Kale M, Keyhani S, et al. Overuse of Health Care Services in the Management of Cancer: A Systematic Review. *Med Care* 2017; **55**(7): 723-33.
36. McCormick B, Winter K, Hudis C, et al. RTOG 9804: a prospective randomized trial for good-risk ductal carcinoma in situ comparing radiotherapy with observation. *J Clin Oncol* 2015; **33**(7): 709-15.
37. Wärnberg F, Garmo H, Emdin S, et al. Effect of radiotherapy after breast-conserving surgery for ductal carcinoma in situ: 20 years follow-up in the randomized SweDCIS Trial. *J Clin Oncol* 2014; **32**(32): 3613-8.
38. Allred DC, Anderson SJ, Paik S, et al. Adjuvant tamoxifen reduces subsequent breast cancer in women with estrogen receptor-positive ductal carcinoma in situ: a study based on NSABP protocol B-24. *J Clin Oncol* 2012; **30**(12): 1268-73.

39. Forbes JF, Sestak I, Howell A, et al. Anastrozole versus tamoxifen for the prevention of locoregional and contralateral breast cancer in postmenopausal women with locally excised ductal carcinoma in situ (IBIS-II DCIS): a double-blind, randomised controlled trial. *Lancet* 2016; **387**(10021): 866-73.
40. Margolese RG, Cecchini RS, Julian TB, et al. Anastrozole versus tamoxifen in postmenopausal women with ductal carcinoma in situ undergoing lumpectomy plus radiotherapy (NSABP B-35): a randomised, double-blind, phase 3 clinical trial. *Lancet* 2016; **387**(10021): 849-56.
41. McCormick B, Winter KA, Woodward W, et al. Randomized Phase III Trial Evaluating Radiation Following Surgical Excision for Good-Risk Ductal Carcinoma In Situ: Long-Term Report From NRG Oncology/RTOG 9804. *J Clin Oncol* 2021: JCO. 21.01083.

Supplementary materials

Supplementary Table S1: Characteristics of Imaging exams

Characteristic	N (%)
Reason for imaging	
Asymptomatic surveillance imaging	66,314 (78.0)
Evaluation of new sign/symptom	10,647 (12.5)
Evaluation of established diagnosis	7,740 (9.1)
Unable to determine	356 (0.4)
Imaging modality	
Mammography	70,413 (82.8)
Breast US	8,985 (10.6)
Breast MRI	5,192 (6.1)
Other breast imaging study	468 (0.6)
Imaging laterality	
Bilateral	57,837 (68.0)
Ipsilateral	21,158 (24.9)
Contralateral	6,063 (7.1)

N=607 women did not have any imaging exams recorded.

Supplementary Table S2: Characteristics of recurrent ipsilateral invasive breast cancer

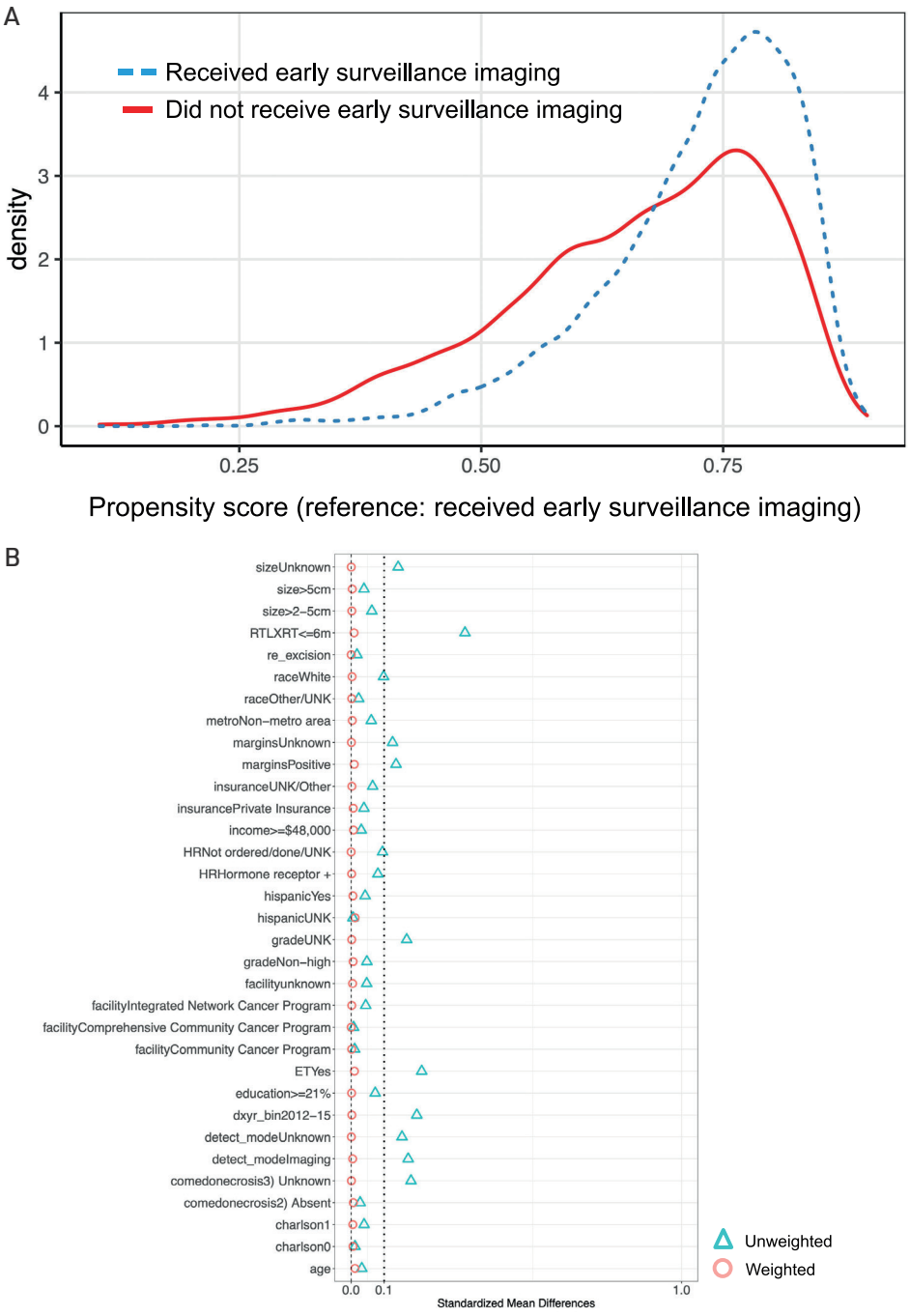
Characteristic	With early surveillance imaging[1] (N=8,825) n (%)	Without early surveillance imaging[1] (N=3,694) n (%)
Total number of events	135 (1.5)	55 (1.5)
Time to recurrence in years (median, IQR)[2]	4.4 (1.4-9.7)	4.7 (1.8-9.6)
Clinical characteristics		
Breast cancer death without SLBE/DM recorded[3]	1 (0.7)	2 (3.6)
Distant Metastasis[4]	13 (9.6)	3 (5.5)
Grade I invasive breast cancer	14 (10.3)	12 (21.8)
Grade II invasive breast cancer	59 (43.7)	14 (25.5)
Grade III invasive breast cancer	31 (23.0)	17 (30.9)
Grade IV invasive breast cancer	2 (1.5)	0 (0.0)
Cell type undetermined	15 (11.1)	7 (12.7)
Mode of detection		
Patient (or partner or other) detected sign or symptom that prompted non-routine doctor visit	19 (14.1)	6 (11.0)
Physician detected during scheduled, routine visit	7 (5.2)	4 (7.3)
Detected on routine imaging study for cancer follow-up in the absence of symptoms	88 (65.2)	35 (63.6)
Incidental finding on unrelated other imaging	2 (1.5)	0 (0.0)
Unable to determine	19 (14.1)	10 (18.2)

[1]Early surveillance imaging was defined as the receipt of surveillance imaging of any modality within 6 to 18 months after diagnosis.

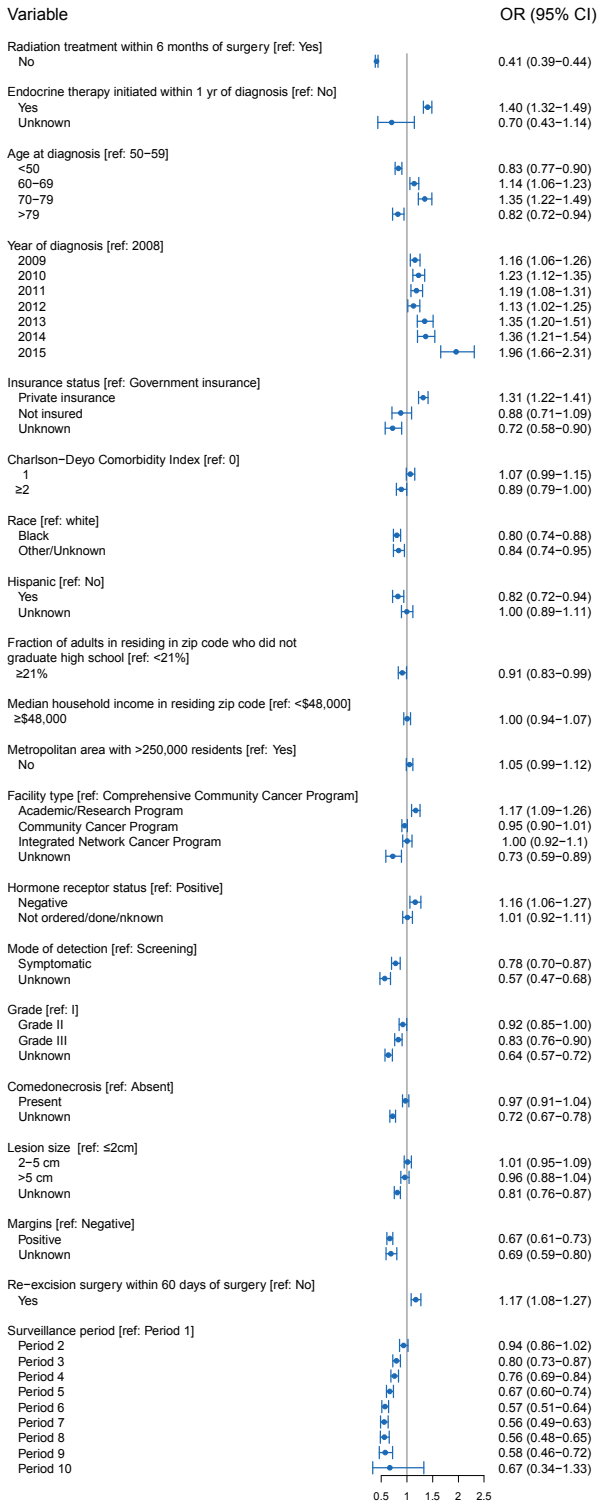
[2]Among women who had an event

[3]If a breast cancer death or distant metastasis was recorded in the absence of a subsequent locoregional breast event (SLBE) of either laterality or distant metastasis (DM), it was considered an event.

[4]If a DM was recorded in the absence of a subsequent locoregional breast event (SLBE) of either laterality, it was considered an event.



Supplementary Figure S1. Inverse probability weighting (IPW). (A) Kernel density estimate of the propensity score distribution among women who did and did not receive early surveillance imaging (ESI) between 6 and 18 months. (B) Standardized mean difference (SMD) of covariates between the two groups (ESI: Yes, and ESI: No), before and after IPW. After weighting, all SMDs were below 0.1.



Supplementary Figure S2.
Predictors of surveillance screening (complete model results)


4

Chapter 4

Preferences of Treatment Strategies among Women with Low-Risk DCIS and Oncologists

Danalyn Byng [1,2], Valesca P. Retèl [1,2], Ellen G. Engelhardt [1,3], Catharina G. M. Groothuis-Oudshoorn [2], Janine A. van Til [2], Renée S. J. M. Schmitz [3], Frederieke van Duijnhoven [4], Jelle Wesseling [3,5], Eveline Bleiker [1,6], Wim H. van Harten [1,2] and on behalf of the Grand Challenge Precision Consortium

[1] Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands

[2] Department of Health Technology and Services Research, Technical Medical Centre, University of Twente, Enschede, The Netherlands

[3] Division of Molecular Pathology, The Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands

[4] Division of Surgical Oncology, The Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands

[5] Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands

[6] Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands

Cancers **13(16)**, 310-320 (2021). https://doi.org/10.3390/cancers13163962

Simple Summary

Preferences for treatment strategies for low-risk ductal carcinoma in situ (DCIS), a potential precursor of invasive breast cancer (IBC) including a new active surveillance strategy, were elicited with a discrete choice experiment among recently-diagnosed women and oncologists involved in the care of women with DCIS. Patients exhibited strong preferences for active surveillance and seemed prepared to accept much higher levels of 10-year risk of developing ipsilateral invasive breast cancer than oncologists. Both patients and oncologists showed a strong aversion toward more extensive locoregional treatments (i.e., breast conserving surgery followed by radiotherapy, and mastectomy), while both groups demonstrated a strong preference toward shorter follow-up intervals.

Abstract

As ongoing trials study the safety of an active surveillance strategy for low-risk ductal carcinoma in situ (DCIS), there is a need to explain why particular choices regarding treatment strategies are made by eligible women as well as their oncologists, what factors enter the decision process, and how much each factor affects their choice. To measure preferences for treatment and surveillance strategies, women with newly-diagnosed, primary low-risk DCIS enrolled in the Dutch CONTROL DCIS Registration and LORD trial, and oncologists participating in the Dutch Health Professionals Study were invited to complete a discrete choice experiment (DCE). The relative importance of treatment strategy-related attributes (locoregional intervention, 10-year risk of ipsilateral invasive breast cancer (iIBC), and follow-up interval) were discerned using conditional logit models. A total of n = 172 patients and n = 30 oncologists completed the DCE. Patient respondents had very strong preferences for an active surveillance strategy with no surgery, irrespective of the 10-year risk of iIBC. Extensiveness of the locoregional treatment was consistently shown to be an important factor for patients and oncologists in deciding upon treatment strategies. Risk of iIBC was least important to patients and most important to oncologists. There was a stronger inclination toward a twice-yearly follow-up for both groups compared to annual follow-up.

Introduction

An active surveillance strategy has been proposed as a new treatment strategy for women with grade I or II primary ductal carcinoma in situ (DCIS), considered a potential precursor of invasive breast cancer (IBC). Between 2014 and 2017, three international, multicenter prospective randomized controlled trials (RCT) evaluating the safety and feasibility of an active surveillance strategy as an alternative to surgical intervention for low-risk DCIS began. Women recruited to the LORD trial in the Netherlands (NCT02492607),[1] the LORIS trial in the United Kingdom (NCT02766881),[2,3] and the COMET trial in the United States (NCT02926911)[4,5] are allocated evenly between the active surveillance arm, and the surgical intervention arm. Women in both arms are followed in the same fashion, with annual mammography (bi-annual in COMET) for a period of up to 10-years post-diagnosis, with ipsilateral invasive breast cancer (iIBC)-free rate as the primary endpoint. All trials have a non-inferiority design, which specifies a clinically meaningful margin for which active surveillance can be considered safe, in terms of the iIBC-free rate, compared to surgical intervention.

Enrolment into these trials was difficult due to strong treatment preferences among eligible woman. Despite public awareness and communication workshops to tackle informational asymmetries in the target population and improve enrolment into the LORIS trial, by the date of the study's closing in March 2020, only 181 of the targeted 932 women were recruited.[6] Women eligible for the LORD trial demonstrated strong treatment preferences, declining enrolment when randomized to their non-preferred arm. This phenomenon is widely reported for trials with randomization; a systematic review and meta-analysis of partially randomized patient preference trials revealed that more than 50% of refusal of randomization was due to patient preference.[7] This challenge to recruitment may be especially true when no novel treatment option is being offered that potentially improves survival such as in the context of de-escalation trials. If active surveillance is the novel strategy, eligible patients can always de-escalate their own treatment on their own accord and in agreement with their treating oncologists. It is not necessary to enroll into a trial to gain access to the desired treatment and follow-up strategy, unless selecting a de-escalation strategy is informed by risk-stratification using biomarkers not available outside a trial.

The active surveillance trials for DCIS are part of a growing trend toward de-escalation of locoregional and systemic treatment for early breast cancer and DCIS.[8] Given the context of already excellent long-term survival for treated DCIS,

numerous studies have evaluated the role and added benefit of radiotherapy and endocrine therapy following surgery.[9-15] While all studies have reported a reduction of local recurrence following use of radiotherapy or adjuvant endocrine therapy, not one has demonstrated survival benefits from these treatments. Unlike the ongoing active surveillance trials, these previous studies have focused on DCIS without differentiating groups by future risk of iIBC, with the exception of the randomized RTOG 9804 trial. This is the only trial that was restricted to low-risk DCIS, defined by lesion size ≤ 2.5 cm, low or intermediate grade, and negative margins ≥ 3 mm, and aimed to estimate the effect of omitting radiotherapy.[10,16] This trial was open for recruitment between July 1996 and July 2006 and was closed with less than 40% of the originally planned 1790 women accrued.

To tackle difficulties recruiting women, the LORD trial changed from a RCT design to a preference-based design in July 2020. In the COMET trial, the study design allows for 'crossover' if a patient randomized to one arm opts for the other (e.g., if a patient randomized to active surveillance opts for surgery in the absence of invasive breast cancer, or vice versa). Rate of crossover is included as a study endpoint.[4]

Within the context of low-risk DCIS and apart from trial enrollment, there is a need to explain why particular treatment choices are made and what factors enter into the decision process to better inform shared decision-making processes between patients and physicians. Furthermore, if an active surveillance strategy is deemed safe and effective based on the findings of these studies in the future, incorporating the patients' preferences in treatment decision making will serve to improve treatment compliance and satisfaction. A woman's preference for treatment strategy may not only be informed by the extensiveness of the procedure itself, but also what happens afterward: the follow-up regimen and possible outcomes including risk of progression of disease or other impacts on self-image.

In light of the challenges posed by strong preferences that women with recently-diagnosed low-risk DCIS seem to have regarding their treatment strategy, a discrete choice experiment (DCE) was designed to discern their preferences for treatment and follow-up strategies, while capturing the relative importance of treatment characteristics and the acceptable trade-offs that they make between them. An active surveillance strategy may only be deemed acceptable to bring into clinical practice if it can be demonstrated in prospective trials that low-risk DCIS can safely be monitored without causing excess iIBC rates compared to conventional treatment. Therefore, part of this aim was to measure how women weigh the importance of risk of iIBC, relative to other aspects of a treatment strategy. By having health

care professionals involved in the care of these women also complete the DCE, a comparison of preferences can also be made between patients and oncologists.

Materials and Methods

Study Population: Patients and Oncologists

Between June 2019 and June 2020, women with low-risk DCIS who declined enrolment into the LORD trial due to strong preferences for either study arm were invited to participate in the Dutch prospective CONTROL DCIS Registration study. As part of the study, participating women completed a baseline questionnaire including DCE within one month of their enrollment, before possible active treatment. In January 2021, the CONTROL DCIS Registration study was subsequently closed because the LORD trial initially randomizing between active surveillance and conventional treatment was amended to a patient-preference design, similar to the CONTROL study. At closing, 28 (78%) of the women registered to the CONTROL study had selected an active surveillance strategy instead of surgical intervention. From July 2020 onward, newly recruited women to the amended LORD trial have been invited to complete the new version of the baseline questionnaire including the same DCE used for the CONTROL DCIS Registration. Women with completed baseline questionnaires up to February 2021 were included in this analysis.

All eligible women were over 45 years old, diagnosed with primary low or intermediate grade DCIS detected on screening mammography, residing in the Netherlands. At the time of completion of the questionnaire, the women would have chosen either an active surveillance strategy or surgical intervention, but may not yet have undergone the full procedure in the latter option at the time of the baseline questionnaire. The patient respondents included in this study were recruited from 30 hospitals across the Netherlands.

Between October 2019 and December 2020, health care professionals involved in the care of women with DCIS in the Netherlands were invited to participate in the online Health Care Professionals Study questionnaire.

Questionnaire Design for Patients and Health Care Professionals

Patients completing the baseline questionnaire of the CONTROL DCIS Registration and LORD trial (preference-based design) were offered a paper or digital copy of the questionnaire (in Dutch) that comprises questions about socio-demographic characteristics, DCE questions, and health-related quality of life (HRQoL) items. An information letter was also included, outlining the purpose of the study and procedures.

Health care professionals involved in the LORD trial received a personal invitation email to participate in the online Health Professionals Study, and to recruit those not involved in the LORD trial, an email invitation was distributed via the Dutch society for surgical oncology, the Dutch society for radiation oncology and all regional breast cancer working groups affiliated with the Netherlands Comprehensive Cancer Organization. The questionnaire investigated the participants' own preference for treatment, the impact of clinical characteristics on treatment preference, and need for decision support tools. Surgical oncologists and radiation oncologists completing the online questionnaire were invited to complete the same DCE as the patients.

Approval for the CONTROL DCIS study and Health Care Professionals study was obtained from the Netherlands Cancer Institute's institutional review board. The Medical Ethics Committee of the Netherlands Cancer Institute approved the LORD patient preference trial.

Intolerance of Uncertainty

In the questionnaire presented to patients, a series of socio-demographic and health-related quality of life (HRQoL) items were included. The responses to the Dutch version of the 12-item Intolerance of Uncertainty Scale (IUS-12) were used for this analysis.[17,18] The IUS-12 assesses self-reported responses to uncertainty, ambiguous situations, and future events. Twelve items are rated on a 5-point Likert scale ranging from 1 (not at all characteristic of me) to 5 (entirely characteristic of me), summing a total score with a maximum of 60 (higher scores indicate greater uncertainty). Intolerance of uncertainty represents a predisposition toward overestimating the chance of possible (but unlikely) undesirable outcomes in uncertain circumstances, while also finding this chance threatening.[19] A threshold to demarcate women with "low" and "high" uncertainty based on the total score was set at the median value resulting from the respondent population in order to compare responses between women with "low" and "high" uncertainty.

Sesign of the Discrete Choice Experiment

DCEs provide a format to elicit choices in a structured way, making it possible to statistically model binary ("either-or") choices. They are one of many stated-preference methods in which respondents choose between alternatives in a repeated series of choice tasks. Within each choice, a selection of attributes (e.g., features of the treatment) where varying possible "levels" are provided. The DCE uses an experimental design that determines the presentation of specific attribute-level combinations, out of many possible combinations. This makes it possible to

compute the influence of changes in attribute-levels on choice for a treatment strategy.[20]

Attributes and their associated levels that capture relevant features of treatment strategies were identified through a review of the literature and expert elicitation. Experts (psychosocial oncology experts (EB, EGE); oncology nurse (VS); pathologist (JW); DCE experts (CGMGO, JAVT)) were asked to comment on and complete the list of attributes and possible levels. The final selection was confirmed via a series of interviews with health care professionals.[21] Treatment attributes included locoregional treatment (levels: no surgery, breast conserving surgery ± radiotherapy, mastectomy), interval between follow-up mammography screening appointments (levels: two years, one year, six months), and chance of ipsilateral invasive breast cancer at 10 years (levels: 5%, 10%, 15%) (Table 1).

For each DCE question (i.e., "choice task"), respondents would choose between two hypothetical treatment strategy alternatives ("Option 1" and "Option 2") that consist of a unique combination of different attribute levels, determined through an experimental design. All participants were provided educational content on the purpose of the DCE, emphasizing that the treatment strategies and outcomes presented were hypothetical situations. The combination of the strategy alternatives, attributes, and their levels resulted in 90 hypothetical scenarios, derived from a fractional main effects experimental design. Unrealistic combinations of different attribute levels were removed from the experimental design. Presenting all scenarios to respondents would be too burdensome, so a subset of scenarios was used. The R package AlgDesign version 1.2.0 was used to generate a D-efficient design consisting of 36 hypothetical scenarios, divided into three versions of the DCE consisting of 12 choice tasks each. Participants were randomized to receive one of the three versions. An example of a choice task has been provided in the online Supplementary Materials (Methods S1); design restrictions are described in Methods S2.

Table 1. Attributes and their respective levels in the discrete choice experiment.

Attributes	Levels
Locoregional treatment strategy	No surgery; breast conserving surgery; breast conserving surgery followed by radiotherapy; mastectomy
10-year risk of ipsilateral invasive breast cancer (iIBC)	5%; 10%; 15%
Surveillance mammography follow-up interval	6 months; 1 year; 2 years

The minimum required sample size was determined to be $n = 84$, based on the rule of thumb for conditional logit models proposed by Johnson and Orme, taking into consideration the number of choice tasks, alternatives, and analysis cells.[22,23]

Conditional Logit Model and Comparing Patient and Oncologist Preferences

To estimate the relative importance of treatment-related features across all respondents, separate conditional logit models for binary choice were built for patients and oncologists. This technique is informed by random utility theory, where a regression model is used to relate choice (i.e., choice of treatment strategy) as a function of the features of the choice (i.e., the attributes and respective levels).[24] The attributes locoregional treatment, follow-up interval, and risk of subsequent iIBC were included in the model as covariates using dummy coding. Resulting co-efficients (β) were exponentialized to derive odds ratios. *p*-values < 0.05 are considered as statistically significant. In a model with pooled data, an interaction term for respondent type was included for all dummy-coded attribute-levels to determine where preferences differed between patients and oncologists. To test the significance of the overall interaction between different attributes and respondent type, the likelihood ratio test was conducted comparing models with and without the interaction terms, with two degrees of freedom.

To understand the relative contribution of the attribute-level to the utility that the respondent assigns to an alternative, importance weights were calculated separately for patients and oncologists. Utility can be understood as the measure of value or importance, and consequently important weights represent the relative importance of each level. These importance weights are the resulting coefficients (β) of the conditional logit models. The overall importance weight (*OIW*) of each attribute (*i*) was calculated by dividing the range in regression coefficients of each attribute *i* (i.e., the difference between the least and most preferred attribute levels, $maxC_i - minC_i$), by the sum of the coefficient ranges of the three attributes ($maxC_j - minC_j$).

$$OIW_{Attribute\ i} = \frac{maxC_i - minC_i}{\Sigma_k \left(maxC_j - minC_j\right)} \quad (1)$$

Scaled overall importance weights (as a fraction of 100) were then derived for each attribute, together summing 100. Overall importance weights were calculated separately for oncologists, patients, and for patient subgroups (women with "high" and "low" uncertainty intolerance, as determined by the IUS-12 cohort median value, patients undergoing active surveillance, patients undergoing conventional treatment, and women with "high" and "low/intermediate" educational attainment).

A description of an effect-modifier analysis to study the extent to which certain patient characteristics impacted the preferences of respondents is described in the online Supplementary Materials (Methods S3).

Maximum acceptable risk was calculated for patients based on the resulting coefficients from the conditional logit model. This can be understood as what extra risk of ipsilateral invasive breast cancer at 10 years patients are willing to take for getting no treatment compared to breast conserving surgery. This is calculated by dividing coefficients to determine the change in risk of ipsilateral invasive breast cancer that would offset the utility gain of the most preferred locoregional treatment strategy.

All statistical analyses were performed with R version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria). The R package mlogit version 1.1-1 was used for the conditional logit models.

Results

Respondents

A total of 202 individuals completed the questionnaire including DCE by March 2021; 37 patients from the CONTROL DCIS registration, 135 patients from the LORD trial, and 30 radiation and surgical oncologists from the Health Professionals Study (Table 2). Patients had a mean age of 59 (range 45–77), and 95 (55.2%) were engaged in paid labor (part-time or full-time). A total of 76.7% opted for no surgical intervention for their primary low-risk DCIS. Responses on the IUS-12 uncertainty intolerance scale ranged between 20 and 51, with the median value at 30. A total of 70% of the oncologist respondents were female. More than 50% treated more than 15 women with DCIS per year. Twenty oncologists (67%) specialized in surgical oncology and the remaining in radiation oncology; all were employed at a range of hospitals across the Netherlands (Table 2).

Importance of Treatment Characteristics

Table 3 shows the aggregate results of the discrete choice experiment for patients and oncologists separately based on the conditional logit models. Model coefficients are also plotted in Figure 1. The preferred locoregional treatment option for patients and oncologists was no surgery, then breast conserving surgery, followed by breast conserving surgery and radiotherapy. The least preferred option was mastectomy. A follow-up interval of six months was preferred by all respondents. Patients did not assign large relative importance to any of the possible levels of iIBC risk whereas for

Table 2. Patient and oncologist characteristics.

Characteristics	Patients (*n* = 172) *N* (%)	Oncologists (*n* = 30) *N* (%)
Age, years (median, range)	59 (45–77)	N.A.
Sex		
Female	172 (100%)	21 (70.0%)
Male	0	9 (30.0%)
Actual treatment selected		
Active surveillance	132 (76.7%)	N.A.
Conventional treatment	38 (22.1%)	N.A.
Unknown	2 (1.2%)	N.A.
Educational level		
Low	37 (21.5%)	0
Intermediate	78 (45.3%)	0
High	57 (33.1%)	30 (100%)
Employment status		
Employed (part-time or full-time)	95 (55.2%)	30 (100%)
Unemployed/pension	77 (44.8%)	0
Hospital type		
Academic medical center	3 (1.7%)	8 (26.7%)
General teaching hospital	105 (61.0%)	14 (46.7%)
Specialized oncology hospital	25 (14.5%)	5 (16.7%)
General hospital	39 (22.7%)	3 (10.0%)
Region of the Netherlands		
North	3 (1.7%)	3 (10.0%)
East	60 (34.9%)	5 (16.7%)
West	98 (57.0%)	17 (56.7%)
South	11 (6.4%)	5 (16.7%)
Subspecialty		
Surgical oncology	N.A.	20 (66.7%)
Radiation oncology	N.A.	10 (33.3%)
Number of patients with DCIS treated per year		
2–5 patients	N.A.	1 (3.3%)
6–10 patients	N.A.	7 (23.3%)
11–15 patients	N.A.	3 (10.0%)
16–20 patients	N.A.	11 (36.7%)
>20 patients	N.A.	8 (26.7%)
Years' experience treating patients with DCIS		
2–5 years	N.A.	1 (3.3%)
6–10 years	N.A.	9 (30.0%)
>10 years	N.A.	20 (66.7%)

oncologists, iIBC risk was a very important factor. There was a statistically significant difference between the oncologists and patients in their preference of attributes (likelihood ratio test on interaction between respondent type and attribute, $p < 0.001$) (Table 3). The test of interaction between respondent type and the attribute-level 15% risk of iIBC was also statistically significant ($p = 0.02$). Oncologists and patients were not statistically significantly different in their preference for the other attribute-levels for follow-up interval or locoregional treatment.

We determined what extra risk of ipsilateral invasive breast cancer at 10 years patients were willing to take for getting no surgery compared to breast conserving surgery. This calculation of maximum acceptable risk found that the additional increase in risk (from the reference level of 5%) that exactly offsets the increase in utility of having no surgery (i.e., not experiencing the side-effects) would be 11.2%.

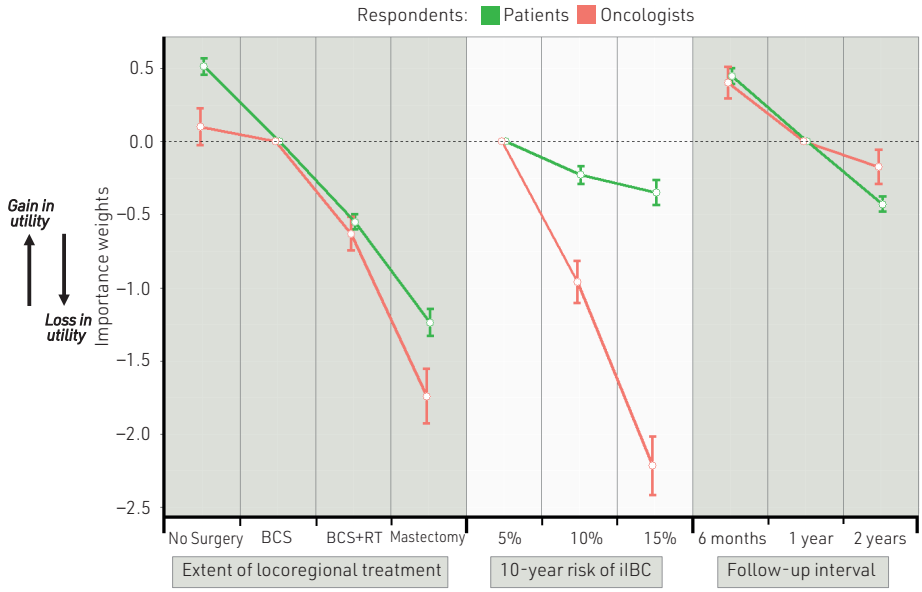


Figure 1. Importance weights derived from the conditional logit model. Standard error bars shown are an indication uncertainty in respondents' preferences. Importance weights for each attribute-level shown are a measure of relative preference. Moving from one attribute-level to an adjacent attribute-level is an indication of the relative gain or loss in utility, where utility is a representation of the strength of preferences. BCS: breast conserving surgery; iIBC: ipsilateral invasive breast cancer; RT: radiotherapy.

Table 3. Stated preferences across all respondents based on the conditional logit model with dummy coding.

Attribute Levels	**Patients**		
	Coefficient (β)	**SE**	**Exp (β)**
Locoregional treatment			
Breast conserving surgery	(ref.)		
No surgery	0.513 *	0.111	1.67
Breast conserving surgery + radiotherapy	−0.551 *	0.102	0.58
Mastectomy	−1.239 *	0.185	0.29
10-year risk of ipsilateral invasive breast cancer			
5%	(ref.)		
10%	−0.229	0.122	0.79
15%	−0.350 *	0.174	0.70
Interval surveillance follow-up			
1 year	(ref.)		
6 months	0.448 *	0.110	1.56
2 years	−0.429 *	0.103	0.65
Interaction Terms [a]			**Coefficient (β)**
Attribute: Locoregional treatment * respondent type			
Level: No surgery			
Patient			(ref.)
Oncologist			−0.413
Level: Breast conserving surgery + radiotherapy			
Patient			(ref.)
Oncologist			−0.082
Level: Mastectomy			
Patient			(ref.)
Oncologist			−0.504
Attribute: 10-year risk of ipsilateral invasive breast cancer * respondent type			
Level: 10% risk of iIBC			
Patient			(ref.)
Oncologist			−0.734
Level: 15% risk of iIBC			
Patient			(ref.)
Oncologist			−1.870
Attribute: follow-up interval * respondent type			
Level: 6mo follow-up interval			
Patient			(ref.)
Oncologist			−0.045
Level: 2 year follow-up interval			
Patient			(ref.)
Oncologist			0.254

	Oncologists		
	Coefficient (β)	**SE**	**Exp (β)**
	(ref.)		
	0.100	0.251	1.11
	−0.632 *	0.229	0.53
	−1.743 *	0.371	0.18
	(ref.)		
	−0.962 *	0.290	0.38
	−2.219 *	0.399	0.11
	(ref.)		
	0.403	0.218	1.50
	−0.175	0.235	0.84
SE	**Exp (β)**	***p*-Value**	
		<0.001 [b]	
0.275	0.66	0.13	
0.251	0.92	0.75	
0.414	0.60	0.22	
		<0.001 [b]	
0.314	0.48	0.02	
0.435	0.15	<0.001	
		<0.001 [b]	
0.245	0.96	0.85	
0.258	1.29	0.32	

iIBC: ipsilateral invasive breast cancer; ref.: reference level; * Statistically significant p-value < 0.05; [a] Computed from model using pooled data from all respondents, with interaction term for respondent type. [b] Based on the likelihood ratio test comparing models with and without the interaction terms, with 2 degrees of freedom.

Influence of the Attributes on Patients' and Oncologists' Preference

Scaled overall importance weights for each attribute are shown in Figure 2. These weights represent the relative influence of an attribute on the respondents' preference for a treatment strategy. For patients, 10-year risk of iIBC was the least important attribute dictating preference, whereas this was the most important for oncologists (representing 14% vs. 50% of importance). For patients, the locoregional treatment was the most important attribute dictating preference, followed by follow-up interval. Heterogeneity in preferences exists among patient subgroups. Women who chose a conventional treatment strategy in real life assigned a higher overall importance weight to 10-year risk of iIBC compared to women who chose active surveillance (38% vs. 11%). Models were built separately for women split by their scores in the bottom and top half of the intolerance of uncertainty scale. Women with higher uncertainty intolerance scores seemed to attach higher importance to follow-up interval slightly more than their counterparts on the other side of the scale (20% vs. 17%). The relative importance of iIBC risk was the same for both groups. When inspecting relative importance of attributes by education level, the importance of locoregional treatment and follow-up interval was shown to be nearly equal.

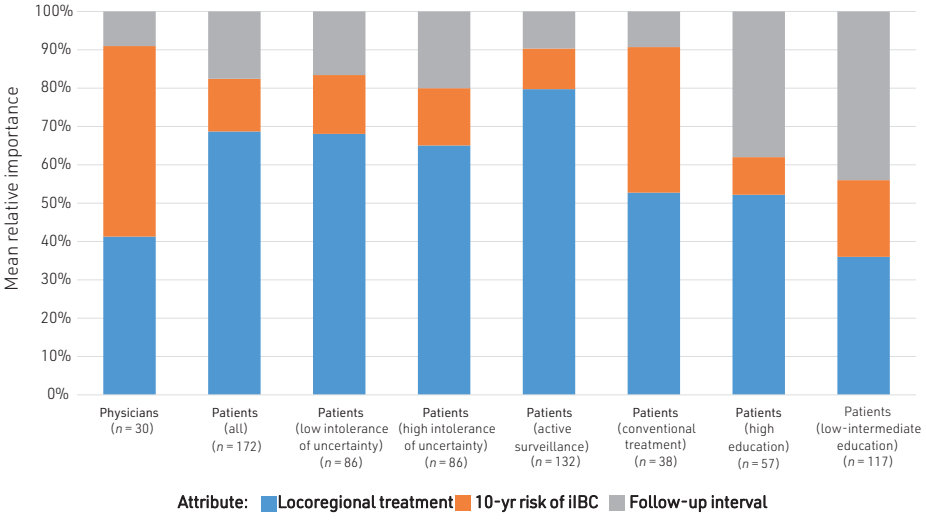


Figure 2. Overall relative importance weights for attributes (features) of treatment strategy. Exploration of further patient subgroup stratifications are described in the online Supplementary Materials (Methods S3). iIBC: ipsilateral invasive breast cancer.

Discussion

The extensiveness of the locoregional treatment was consistently shown to be an important factor for patients and their care providers in deciding upon treatment strategies for low-risk primary DCIS. In our analyses, risk of ipsilateral invasive breast cancer was least important to patients and most important to oncologists. There was a stronger inclination toward a twice-yearly follow-up for both oncologists and patients compared to annual follow-up.

We found that women in the Netherlands had very strong preferences for an active surveillance strategy with no surgery, irrespective of the 10-year risk of iIBC. This was also the case for our respondents who scored higher than the cohort's average score (30) on the IUS-12 uncertainty intolerance scale. These women, known to have a higher intolerance of uncertainty, were not dissimilar to their counterparts with lower intolerance in assigning a comparatively small overall importance weight to the risk of iIBC. For these women, the locoregional treatment, followed by the interval between follow-up mammograms, were more important. It is possible that the IUS-12 uncertainty intolerance scale does not capture future breast cancer risk tolerance in comparison to tolerance of risk attributed to other attributes (e.g., risk of infection or post-operative complications).[25,26] Furthermore, the risk of iIBC already remains rather low among these women with good-prognosis DCIS, and they are being asked to evaluate a risk far in the future at 10 years. The women in our study not only attached lower importance to future risk of breast cancer, but also attached higher importance to breast conservation through having no surgery. This can be aligned with prospect theory, popularized by Kahneman and Tversky, which posits that "people underweigh outcomes that are merely probable in comparison with outcomes that are obtained with certainty".[27] It is not yet understood how the dimension of temporal distance to the risk in question factors into decision making and preferences measured in DCEs, particularly for DCIS.[28] A study of intolerance of uncertainty among men undergoing active surveillance for prostate cancer found that intolerance of uncertainty had a significant relationship with the experience of cancer-related symptoms[29]. The women in our study were asymptomatic and their DCIS was detected through the national breast cancer screening program, so they remain physically unaffected by their diagnosis.

An important related consideration that likely factors into a patient's choice is the understanding of one's personal risk of upstaging to invasive breast cancer; this was not explicitly captured in the DCE design. Uncertainty still remains over the proportion of patients with a core needle biopsy showing DCIS with "low-risk" clinicopathological characteristics who actually have concurrent invasive carcinoma

in the breast. This uncertainty is now understood to have an impact on participation in trials studying active surveillance. A retrospective series based on a small sample of women who would have met eligibility criteria for active surveillance trials found low upstaging rates (6–10%).[30] All upstaged cases were good-prognosis invasive carcinomas: all were node negative and HER2 negative. Furthermore, a proportion of women with DCIS will have complete removal of the lesion at biopsy, and subsequently experience a low upgrade rate (8.2%).[31] A study in the Netherlands addressed the issue of the reliability of preoperative biopsy, and identified several factors that can aid in further risk stratification of women being considered for non-operative management.[32] An important takeaway from these studies is that even with possible upstaging, overall survival should not be significantly compromised. Access to high-quality annual mammography is readily available, and invasive carcinomas can be treated on time. Nevertheless, the prediction of upstaging of DCIS to invasive disease remains an important area of ongoing research, and will serve to identify the lowest achievable upstaging rate among women eligible for clinical trials of active surveillance. This may in turn address some of the challenges with trial accrual, and better inform the understanding of risk of upstaging.

We used a discrete choice experiment as a "stated preference" method where respondents were asked to choose between alternatives from among a set of hypothetical scenarios generated from an experimental design.[33] This can be contrasted with the concept of "revealed preference" in which we observe actual choices made by respondents in real life. The women included in our study were participants in studies (the CONTROL DCIS Registration and LORD trial) that had a preference-based design. Sixty-eight percent of our patient respondents selected active surveillance as an alternative to surgical intervention in real life. Active surveillance is not yet an accepted treatment strategy according to European clinical guidelines. Conventional treatment for DCIS mimics that of early breast cancer, with breast conserving surgery being the preferred local treatment option.[34] Results from the ongoing prospective clinical trials for active surveillance will not be available for at least five to 10 years, and recruitment into these trials remains challenging. However, in the Dutch context, our study demonstrated that women diagnosed with low- and intermediate-grade DCIS have already established strong preference and desire to undergo active surveillance ahead of the results about safety and 10-year risk of ipsilateral invasive breast cancer.

The non-inferiority design of the Dutch LORD trial is based on the assumption that the 10-year iIBC-free rate is 95% in the surgery group. The non-inferiority margin was chosen at 3.168 on the hazard-ratio scale, corresponding to a 10-year iiBC-free probability of 85% in the active surveillance group.[35] As DCE scenarios presenting

active surveillance were always associated with an increased risk of 5% or 10% compared to the surgical treatment option, we found that these differences were deemed acceptable by the patient respondents. Even when two surgical treatment options were compared, patients had much stronger preferences for strategies with less extensive procedures, irrespective of an associated increased risk. This pattern was not seen among oncologist respondents; a difference of 10% in risk was not deemed acceptable by oncologists on average.

This DCE is the first published study evaluating treatment preferences in women with a recent diagnosis of DCIS. A DCE evaluating patient preferences for outcomes following DCIS treatment was conducted in a healthy cohort of women in the United States attending a comprehensive cancer screening mammography clinic.[36] These women were not diagnosed with DCIS, nor did they have a personal history of breast cancer. That study found that respondents weighed breast cancer risk as the most important factor, but this was closely followed by chronic pain and infection. Again, this is in contrast with the patient respondents in our study who demonstrated that 10-year iIBC risk was the least important factor. The extent to which women without the experience of the disease in question respond similarly to women with the disease is known to be affected by scale heterogeneity, explained by differences between groups due to familiarity with the disease.[37] In the online Supplementary Materials (Methods S4, Figure S1), we provide an evaluation of scale heterogeneity between the patients and oncologists in our sample to understand how similarly these groups respond. We also note that differences in sample size between patients and oncologists may have impacted the difference observed between these two groups. These considerations are necessary to draw comparisons between preferences of any two groups of individuals including women who have been diagnosed with DCIS and those not, to understand the influence of psychological distance on accepting treatment strategies with possible higher risk of a future iIBC event.[38]

Conclusions

This study provided insights into the treatment strategy preferences of a large cohort of women participating in a preference-based prospective study for low-risk DCIS. These women, recently diagnosed with DCIS, assigned the greatest importance to extensiveness of locoregional treatment and surveillance follow-up interval. In stark contrast, risk of iIBC was the most important factor for oncologists involved in the care of DCIS. The responses to the DCEs are reflected in the women's actual treatment choices: the vast majority (68%) chose an active surveillance strategy to manage their low-risk DCIS. The insights gained through

this study about the concordant and discordant preferences for treatment strategies between women and their oncologist may help to inform treatment decision making processes as prospective trials aim to recruit more women. Finally, if an active surveillance strategy is found to be a safe alternative to surgery, incorporating patients' preferences in treatment decision making will serve to improve strategy compliance, satisfaction, and shared-decision making processes.

References

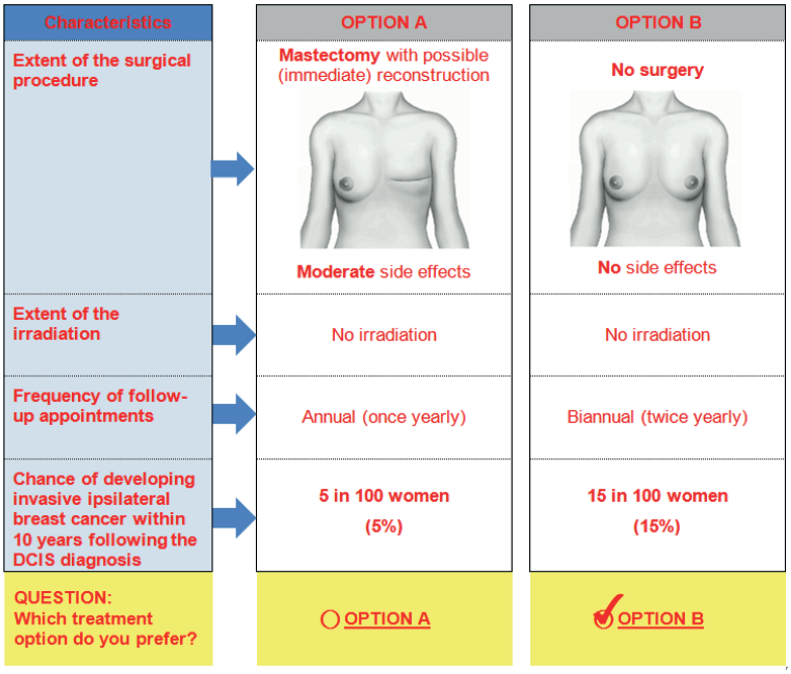
1. Elshof LE, Tryfonidis K, Slaets L, et al. Feasibility of a prospective, randomised, open-label, international multicentre, phase III, non-inferiority trial to assess the safety of active surveillance for low risk ductal carcinoma in situ - The LORD study. *Eur J Cancer* 2015; 51(12): 1497-510.
2. Francis A, Thomas J, Fallowfield L, et al. Addressing overtreatment of screen detected DCIS; the LORIS trial. *Eur J Cancer* 2015; **51**(16): 2296-303.
3. Wallis M, Bartlett J, Billingham L, et al. The LORIS trial: Randomising patients with lowor low intermediate-grade ductal carcinoma in situ (DCIS) to surgery or active monitoring. *Breast Cancer Res Treat* 2018; **167**:325–326.
4. Hwang ES, Hyslop T, Lynch T, et al. The COMET (Comparison of Operative versus Monitoring and Endocrine Therapy) trial: a phase III randomised controlled clinical trial for low-risk ductal carcinoma in situ (DCIS). *BMJ Open* 2019; **9**(3): e026797.
5. Youngwirth LM, Boughey JC, Hwang ES. Surgery versus monitoring and endocrine therapy for low-risk DCIS: The COMET trial. *Bull Am Coll Surg.* 2017; **102**(1): 62–3.
6. Cuzick J, Sestak I, Forbes JF, et al. UK Interdisciplinary Breast Cancer Symposium 2020. *Breast Cancer Res Treat* 2020; **180**: 527–96.
7. Wasmann KA, Wijsman P, van Dieren S, et al. Partially randomised patient preference trials as an alternative design to randomised controlled trials: Systematic review and meta-analyses. *BMJ Open.* 2019; **9**: e031151.
8. Piccart MJ, Hilbers FS, Bliss JM, et al. Road map to safe and well-designed de-escalation trials of systemic adjuvant therapy for solid tumors. *J Clin Oncol* 2020; **38**(34): 4120–9.
9. Correa C, McGale P, Taylor C, et al. Overview of the randomized trials of radiotherapy in ductal carcinoma in situ of the breast. *J Natl Cancer Inst Monogr* **2010**(41); 162–77.
10. McCormick B, Winter K, Hudis C, et al. Rtog 9804: A prospective randomized trial for good-risk ductal carcinoma in situ comparing radiotherapy with observation. *J Clin Oncol Off J Am Soc Clin Oncol* 2015; **33**(7): 709–15.
11. Wapnir IL, Dignam JJ, Fisher B, et al. Long-term outcomes of invasive ipsilateral breast tumor recurrences after lumpectomy in NSABP B-17 and B-24 randomized clinical trials for DCIS. *J Natl Cancer Inst* 2011; **103**(6): 478–88.
12. Wärnberg F, Garmo H, Emdin S, et al. Effect of radiotherapy after breast-conserving surgery for ductal carcinoma in situ: 20 years follow-up in the randomized swedcis trial. *J Clin Oncol* 2014; **32**(32): 3613–18.
13. Allred DC, Anderson SJ, Paik S, et al. Adjuvant tamoxifen reduces subsequent breast cancer in women with estrogen receptor-positive ductal carcinoma in situ: A study based on NSABP protocol b-24. *J Clin Oncol* 2012; **30**(12): 1268–73.
14. Forbes JF, Sestak I, Howell A, et al. Anastrozole versus tamoxifen for the prevention of locoregional and contralateral breast cancer in postmenopausal women with locally excised ductal carcinoma in situ (ibis-ii dcis): A double-blind, randomised controlled trial. *Lancet.* 2016; **387**(10021): 866–73.
15. Margolese RG, Cecchini RS, Julian TB, et al. Anastrozole versus tamoxifen in postmenopausal women with ductal carcinoma in situ undergoing lumpectomy plus radiotherapy (NSABP B-35): A randomised, double-blind, phase 3 clinical trial. *Lancet.* 2016; **387**(10021): 849–856.
16. McCormick B. Randomized trial evaluating radiation following surgical excision for "good risk" DCIS: 12-year report from NRG/RTOG 9804. *Int J Radiat Oncol Biol Phys* 2018; **102**(5): 1603.

17. Carleton RN, Norton MA, Asmundson GJ. Fearing the unknown: A short version of the intolerance of uncertainty scale. *J Anxiety Disord* 2007; **21**(1): 105–17.
18. Helsen K, Van den Bussche E, Vlaeyen JW, Goubert L. Confirmatory factor analysis of the Dutch intolerance of uncertainty scale: Comparison of the full and short version. *J Behav Ther Exp Psychiatry* 2013; **44**(1): 21–9.
19. Boswell JF, Thompson-Hollands J, Farchione TJ, Barlow DH. Intolerance of uncertainty: A common factor in the treatment of emotional disorders. *J Clin Psychol* 2013; **69**(6): 630–45.
20. Hauber AB, González JM, Groothuis-Oudshoorn CG, et al. Statistical methods for the analysis of discrete choice experiments: A report of the ISPOR conjoint analysis good research practices task force. *Value Health* 2016; **19**(4): 300–15.
21. Engelhardt E, Byng D, Klaver K, et al. Women diagnosed with ductal carcinoma in situ (DCIS) and healthcare providers' views on active surveillance for DCIS. Results from focus groups and in-depth interviews. *Eur J Cancer* 2020; **138**:S36.
22. Johnson R, Orme B. *Getting the Most from CBC.* Sawtooth Software; Sequim, WA, USA: 2003. (Sawtooth Software Research Paper Series).
23. Orme B. *Sample Size Issues for Conjoint Analysis Studies.* Sawtooth Software; Sequim, WA, USA: 1998. Sawtooth Software Technical Paper.
24. McFadden D. Conditional Logit Analysis of Qualitative Choice Behavior. In: Zarembka P., editor. *Frontiers in Econometrics.* Academic Press; New York, NY, USA: 1973. pp. 105–142.
25. Jonczyk MM, Fisher CS, Babbitt R, et al. Surgical predictive model for breast cancer patients assessing acute postoperative complications: The breast cancer surgery risk calculator. *Ann Surg Oncol* 2021; **28**(9): 5121-31.
26. Rose M, Manjer J, Ringberg A, Svensson H. Surgical strategy, methods of reconstruction, surgical margins and postoperative complications in oncoplastic breast surgery. *Eur J Plast Surg* 2014; **37**(4): 205–14.
27. Kahneman D, Tversky A. *Handbook of the Fundamentals of Financial Decision Making: Part I.* World Scientific; Singapore: 2013. Prospect theory: An analysis of decision under risk; pp. 99–127.
28. Lloyd AJ. Threats to the estimation of benefit: Are preference elicitation methods accurate? *Health Econ* 2003; **12**(5): 393–402.
29. Tan H-J, Marks LS, Hoyt MA, et al. The relationship between intolerance of uncertainty and anxiety in men on active surveillance for prostate cancer. *J Urol* 2016; **195**(6): 1724–30.
30. Grimm LJ, Ryser MD, Partridge AH, et al. Surgical Upstaging Rates for Vacuum Assisted Biopsy Proven DCIS: Implications for Active Surveillance Trials. *Ann Surg Oncol* 2017; **24**(12): 3534–40.
31. Nicosia L, di Giulio G, Bozzini AC, et al. Complete Removal of the Lesion as a Guidance in the Management of Patients with Breast Ductal Carcinoma In Situ. *Cancers* 2021; **13**(4): 868.
32. Mannu GS, Groen EJ, Wang Z, et al. Reliability of preoperative breast biopsies showing ductal carcinoma in situ and implications for non-operative treatment: A cohort study. *Breast Cancer Res Treat* 2019; **178**(2): 409–18.
33. Bridges JF, Hauber AB, Marshall D, et al. Conjoint analysis applications in health—A checklist: A report of the ISPOR good research practices for conjoint analysis task force. *Value Health*. 2011; **14**(4): 403–13.
34. Cardoso F, Kyriakides S, Ohno S, et al. Early breast cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up† *Ann Oncol* 2019; **30**(8): 1194–1220.

35. The LORD study. [Accessed on 14 February 2021]; Available online: https://clinicaltrials.gov/ct2/show/NCT02492607
36. Chapman BM, Yang JC, Gonzalez JM, Havrilesky L, Reed SD, Hwang ES. Patient preferences for outcomes following DCIS management strategies: A discrete choice experiment. *JCO Oncol Pract* 2021; **17**(11): e1639-48.
37. Vass CM, Wright S, Burton M, Payne K. Scale heterogeneity in healthcare discrete choice experiments: A primer. *Patient* 2018; **11**(2): 167–13.
38. Raue M, Streicher B, Lermer E, Frey D. How far does it feel? Construal level and decisions under risk. *J Appl Res Mem Cogn* 2015; **4**(3): 256–64.

Supplementary materials

Methods S1: Example of A Choice Task



In the experiment, respondents are presented with a series of questions in which they are asked to choose a preferred alternative from a set of hypothetical treatment profiles. These treatment profiles vary by levels of treatment factors, shown here and in Figure 1 of the main manuscript.

Our experimental design resulted in 36 choice tasks, and we divided these into three blocks. Each individual was randomized to complete one of the three blocks, containing 12 choice tasks. One example choice task is shown above.

Methods S2: Design Restrictions

To ensure the presentation of hypothetical scenarios with a closer representation of what would be seen in the real-world, this study used restrictions in the DCE design. In the presentation of treatment strategies, the option with a more invasive local intervention was always associated with a lower chance of iIBC at 10 years. Choice tasks always compared different locoregional treatments; as such, the "no surgery" level of the locoregional treatment attribute was never associated with the "5%" level of the risk of iIBC attribute.

Methods S3: Assessing Patient Heterogeneity with Effect-Modifier Analyses

A series of multivariable mixed logit models were built including responses from patients only. To study the extent to which certain patient characteristics impacted the stated preferences of respondents, an effect-modifier analysis was conducted by including interaction terms for age, highest level of education completed, and employment status. Nested random effects for respondent ID and hospital were included to account for correlation among the multiple questions answered per individual, as well as the possible correlation for respondents treated in the same hospital. We found no statistically significant interactions for age (< vs >50 years) or employment status. A multivariable mixed logit with random effects only for respondent ID included an interaction term for hospital type ("Specialized oncology hospital, Y/N"). This was also not statistically significant.

We did however find that compared to women with low/intermediate educational attainment, women with a high level of education demonstrated aversion to mastectomy (coefficient –0.38, p value 0.04), and preference towards breast conserving surgery with radiotherapy (coefficient 0.27, p value 0.004), compared to breast conserving surgery alone. Overall importance weights stratified by high and low/intermediate educational attainment are shown in Figure 2.

Methods S4: Analysis of the Scale Factor

The assumption of homogeneous utility weights requires that unobservable components of utilities should be mutually independent and homoscedastic.[1] The potential for preference and scale heterogeneity in responses for the total sample and by subgroups (i.e., oncologists and patients) should be therefore measured and accounted for. As it is likely that variances differ between datasets derived from women with DCIS and oncologists, attribute-level estimates (i.e., preferences) between both groups cannot be directly compared without first considering scale factor differences using the Swait and Louviere test.[2] Preference differences between patients and oncologists may differ due to real difference in preference or due to scaling. The latter comes from more or less certainty one has over their preferences, in other words, the scale factor can be understood as a measure of the psychological distance that individuals from different groups (e.g., patients and oncologists) have towards given events. This is due to the perfect confound between the mean and variance of the betas.

To compute the Swait and Louviere test, the log likelihoods derived from the conditional logit models for both groups' datasets were collected (L1 and L2). The

attribute-level codes for one dataset were multiplied by a possible scale factor, then the two datasets were combined to derive a pooled log likelihood (Lμ). These steps were repeated for a range of possible scale factors, until a log likelihood representing the model with the best fit was found. This is compared with the log likelihoods derived from the separate models from each group, then compared with the chi-square value of the number of parameters in the model (K) plus 1 (representing degrees of freedom), as outlined in the following formula[3,4]:

$$\lambda_A = -2^*\left[L_\mu - (L_1 + L_2)\right] < x^2$$

Following the steps of the Swait & Louviere test, the hypothesis of equal attribute level estimates was rejected ($p < 0.05$). When varying the scale parameter from 0 to infinity, the corresponding maximal log likelihood differed significantly with the sum of the separate log likelihoods of the two models (patients and oncologists). Therefore it can be concluded that irrespective of the value of the scale parameter, there is always a difference between patients and oncologists in their preferences. Due to the contrast in sample sizes between groups, it is important to consider that the results of the Swait & Louviere test may produce different findings with a larger sample of oncologists.

Coefficients from the patient and oncologists models are also plotted in Figure S1 (referred to as the Swait and Louviere plot). Corresponding to the slope of the line fitted through the points the figure suggests that coefficients between the models differ by a scalar of approximately 0.45.

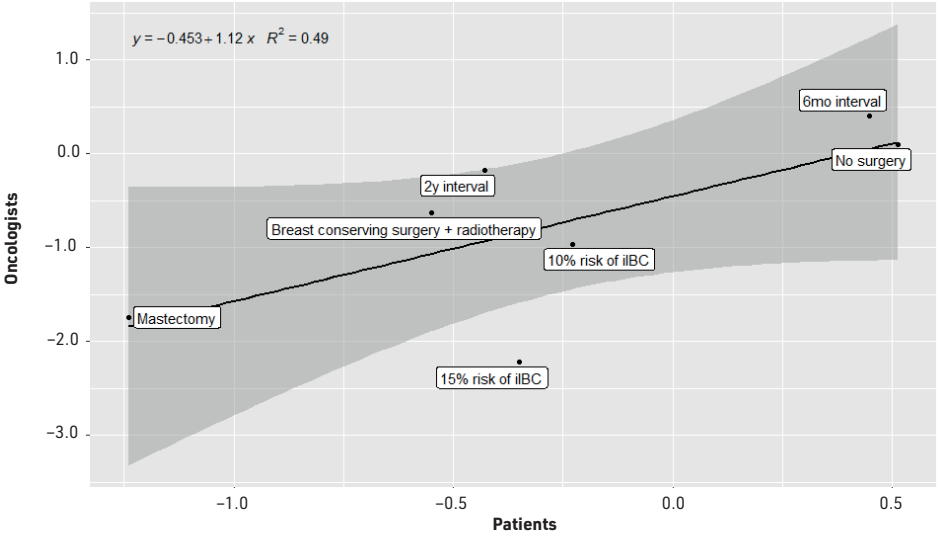


Figure S1. 1 Swait and Louviere plot of coefficients derived from conditional logit models for patients and oncologists. iIBC: ipsilateral invasive breast cancer; mo: month; y: year.

References

1. McFadden D. Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka (ed.), *Frontiers in Econometrics;* Academic Press: New York: 1973; pp. 105-142.
2. Swait J, Louviere J. The role of the scale parameter in the estimation and comparison of multinomial logit models. *J Mark Res* 1993; **30**(3): 305–14.
3. Vass CM, Wright S, Burton M, Payne K. Scale heterogeneity in healthcare discrete choice experiments: A primer. *Patient* 2018; **11**(2): 167–73.
4. Veldwijk J, Groothuis-Oudshoorn CGM, Kihlbom U, et al. How psychological distance of a study sample in discrete choice experiments affects preference measurement: A colorectal cancer screening case study. *Patient Prefer Adherence* 2019; **13**: 273–82.


5

Chapter 5

An early economic evaluation of active surveillance for low-risk DCIS

D. Byng[1,2], M. Schaapveld,[1] E. H. Lips,[3] F.H. van Duijnhoven,[4] J. Wesseling,[3,5] W.H. van Harten,[1,2*] V.P. Retèl,[1,2*] Grand Challenge PRECISION Consortium

[1]Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands

[2]Health Technology and Services Research Department, Technical Medical Centre, University of Twente, Enschede, The Netherlands

[3]Division of Molecular Pathology, The Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands

[4]Division of Surgical Oncology, The Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands

[5]Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands

Submitted

Abstract

Aims: Perform early economic evaluation comparing active surveillance (AS) to surgery for women with low-risk DCIS, a precursor of invasive breast cancer.

Materials & Methods: 10-year incremental costs (€) and quality-adjusted life years (QALYs) were compared between a simulated cohort of women undergoing breast conserving surgery ± radiotherapy, and a cohort with a low-risk subgroup undergoing AS using a semi-Markov model. Scenario and headroom analyses evaluated a better-performing biomarker to select low-risk women for AS.

Results: AS resulted in lower costs and survival, but higher QALYs (+0.40). Scenario analyses maintained survival outcomes and maximized QALYs.

Conclusions: AS for low-risk DCIS is cost-effective, but a better-performing biomarker to select low-risk women can maximize quality-adjusted outcomes.

Future Perspective: Women with primary DCIS who have a low-risk of experiencing subsequent invasive breast cancer will benefit from an active surveillance strategy instead of active locoregional treatment. Accurate biomarkers for DCIS surgery de-escalation are still in the exploration phase, but early economic evaluations have revealed promise in terms of cost-effectiveness and in the willingness of women to undergo active surveillance.

Summary Points:

- We characterize the costs and quality-adjusted health outcomes associated with (non)-interventional strategies for women with primary low-risk DCIS using a semi-Markov model for early cost-effectiveness modeling.
- We explore two bio-marker based strategies for selecting low-risk women who could opt for an active surveillance strategy. Firstly, we use standard pathological information on DCIS grade (low-to-intermediate) and estrogen-receptor-positive status, similar to the eligibility criteria of ongoing prospective clinical trials for active surveillance. In the scenario analysis, models used information on COX-2 protein expression and breast adipocyte size to select low-risk women to forgo surgery.
- Forgoing surgery among these women resulted in significant gains in quality of life, despite an expected elevated rate of ipsilateral invasive breast cancer and somewhat reduced life years on average.
- This early economic evaluation demonstrated that introducing an active surveillance option to select women with low-risk features can be a cost-effective alternative to immediate surgery and adjuvant radiotherapy.

Background and objectives

Active surveillance (AS) is a disease management strategy that allows for the routine monitoring of a given condition for signs of progression. The aim is to uphold quality of life and avoid interventional treatment and related side-effects unless warranted following a progression of the condition. A now widely accepted strategy for some men with low-risk prostate cancer,[1] AS is currently being evaluated as an alternative to surgical resection in (ongoing) clinical trials for women diagnosed with low-risk primary ductal carcinoma in situ (DCIS), a potential precursor of invasive breast cancer (IBC), in the LORD (LOw Risk DCIS), LORIS (LOw Risk dcIS), and COMET (Comparing an Operation to Monitoring, With or Without Endocrine Therapy) studies.[2-5]

The motivation to bring AS into clinical practice for DCIS stems from an understanding about the heterogeneous risk of subsequent ipsilateral IBC (iIBC) following a diagnosis of primary DCIS.[6] For women with low-risk DCIS features, including low-to-intermediate grade and estrogen receptor [ER]-positive status, their risk of progression to iIBC remains low[7] enough to make them an ideal group to consider the option of AS. As these women make up approximately 50% of screen-detected DCIS,[8] the impact on the healthcare system could be considerable.

With a high prevalence of DCIS (representing 20–25% of all screen-detected 'breast cancers') the present costs of treatment to the healthcare system are substantial.[9] The current standard of care dictated by professional guidelines in Europe and the United States recommend surgical resection of the lesion, possibly followed by radiotherapy and endocrine therapy for all DCIS.[10,11] Treatment-related morbidity and the related impact on health-related quality of life among the many women now living with a diagnosis of primary DCIS has already motivated the movement towards de-escalation strategies for adjuvant therapy.[12] Multigene assays such as the Oncotype DX DCIS score and DCISionRT are used to select women who could forgo radiotherapy after BCS.[13,14] While these have been clinically validated, neither option has been found to be cost-effective.[15-17] For "good-risk" patients defined by the Radiation Therapy Oncology Group (RTOG) 9804 study (≥ 60 years, ER-positive, tumor extent 2.5 cm, low-to-intermediate grade, and margins ≥ 3 mm), adjuvant radiotherapy and tamoxifen use were associated with reduced ipsilateral breast recurrence over the long-term follow-up period.[18] However, despite the risk-reducing effect of both radiotherapy and tamoxifen, observation after BCS was found to be the most cost-effective option for women with these "good-risk" features.[19,20]

Recent DCIS cohort studies have identified further promising prognostic factors which have strong associations with developing subsequent iIBC: human epidermal growth factor receptor (HER2) overexpression (odds ratio (OR) 1.56; 95% confidence interval (95% CI), 1.05-2.31), high cyclooxygenase (COX)-2 protein expression (OR 2.97; 95% CI, 1.72-5.10), presence of periductal fibrosis (OR 1.44; 95% CI, 1.01-2.06), and large breast adipocyte size (OR 2.75; 95% CI, 1.25-6.05).[21,22] DCIS with both high COX-2 expression and large breast adipocytes was associated with a 12-fold higher risk (OR 12.0; 95% CI, 3.10-46.3) for subsequent iIBC.[22]

With an increasing understanding of the heterogeneous nature of DCIS, the possibility to further select women who can safely forgo locoregional treatment based on these features or markers is promising. While these prognostic factors await specification and validation in larger clinical studies, BCS will remain the minimum accepted approach for all women with DCIS.

In the Netherlands, women with newly diagnosed DCIS with low-risk features participating in a discrete choice experiment demonstrated strong preferences to forgo BCS and adjuvant therapy altogether, opting instead for AS despite the current minimal clinical evidence to support the safety and feasibility of such a strategy.[23] Results from prospective AS trials will however not be available for at least 10 years. Using mathematical modelling techniques, real-world cancer registry data, and DCIS patient-derived quality of life and preference information, the objective of this paper is to simulate possible patient- and health-system level impacts of introducing an AS strategy. In addition, we can simulate scenarios using different possible biomarkers to improve the selection of women eligible for AS. This early economic model can inform future research and policy and foreshadow the likely drivers of cost-effectiveness associated with presumed trial outcomes. Furthermore, the maximum possible cost for a hypothetical perfect biomarker solution can be modeled.

Methods

Comparators

The base-case model compares two strategies for all women with screen-detected primary DCIS (Figure 1). The first strategy (Strategy A) consists of standard immediate surgical treatment for all DCIS following European clinical guidelines, consisting of breast conserving surgery with (75%) or without (25%) radiotherapy, followed by annual surveillance mammography for 5 years post-diagnosis.[8] The comparator strategy (Strategy B) similarly focuses on all women with screen-detected DCIS, while using standard pathological information to identify a subset of women at low-risk for progression to ipsilateral invasive breast cancer. Standard pathological information on the DCIS grade (low-to-intermediate) and ER-positive status, similar to the eligibility criteria in the LORD trial, is used to identify low-risk women. These women are considered eligible to forgo surgery and opt for an active surveillance strategy consisting of clinical follow-up and surveillance mammography for 10 years post-diagnosis. Based on observed patterns of enrolment into the LORD trial and previously reported Dutch registry data, we assumed 50% of all women with screen-detected primary DCIS have low-risk features and would be eligible for the active surveillance strategy.[8]

Setting and location

The decision model is set in the Netherlands and takes on a Dutch healthcare perspective considering only direct medical costs. A 10-year time-horizon was chosen to limit the dependency on assumptions, given lack of availability of head-to-head trial or real-world data on the comparative effectiveness of surgery vs. active surveillance. This time horizon was further substantiated by the evidence that the use of adjuvant RT is only associated with lower risk of iIBC in the first decade after DCIS diagnosis, with lower risks of second breast events over time.[24,25] The persistence of the treatment effect of surgery on survival beyond 10 years is also not known.[26] It is expected that in 10-years time, technological advancements will have been made which can more accurately select women for active surveillance.

Choice of model

A multi-state modeling approach was used to simulate the disease process after diagnosis with primary DCIS. This approach is based on a continuous-time semi-Markov model.[27,28] The use of Markov models to conduct economic evaluations in healthcare settings is widespread, but may be more appropriate to model disease processes for chronic, long-term illnesses given the use of constant unvarying transition probabilities and the memoryless "clock-forward" property.[29] For our purposes modeling DCIS, we use a semi-Markov model which employs a clock-

reset approach.[30] History, or time spent in a given state is measured, allowing transition probabilities between states to be time-dependent. Survival times from which transition probabilites are derived are treated as continuous variables. Such a model also allows for the consideration of population heterogeneity and competing event and mortality risks.[31]

The semi-Markov multi-state model was built with five health states: (1) iIBC-free; (2) iIBC within five years of DCIS diagnosis; (3) iIBC more than five years post-DCIS diagnosis; (4) death after iIBC; and (5) death without experiencing iIBC (Figure 2). The iIBC health state was split into two to capture the different possible biological processes relating to a subsequent iIBC. This decision was due to the observed difference in frequency of events (hazard rate) in the first 5 years, compared to after 5 years.[25,32] Two all-cause death states are modeled, occurring either after an intermediate iIBC event, or without any intermediate event.

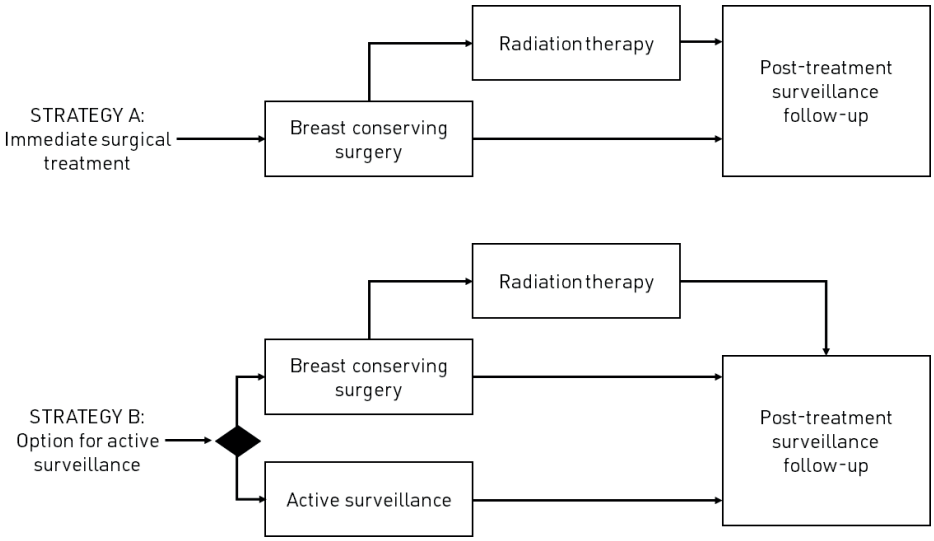


Figure 1. Strategies explored in the economic evaluation

The base-case model compares two strategies for all women with screen-detected primary DCIS. Strategy A consists of standard immediate surgical treatment for all DCIS, consisting of breast conserving surgery with (75%) or without (25%) radiotherapy, followed by annual surveillance mammography for 5 years post-diagnosis. The comparator strategy (Strategy B) similarly focuses on all women with screen-detected DCIS, while using a biomarker to identify a subset of women at low-risk for progression to ipsilateral invasive breast cancer. 50% of the cohort would be eligible for active surveillance based on low-risk characteristics (estrogen receptor positive DCIS grade I/II). 75% of women treated with breast conserving surgery will undergo adjuvant radiation therapy.

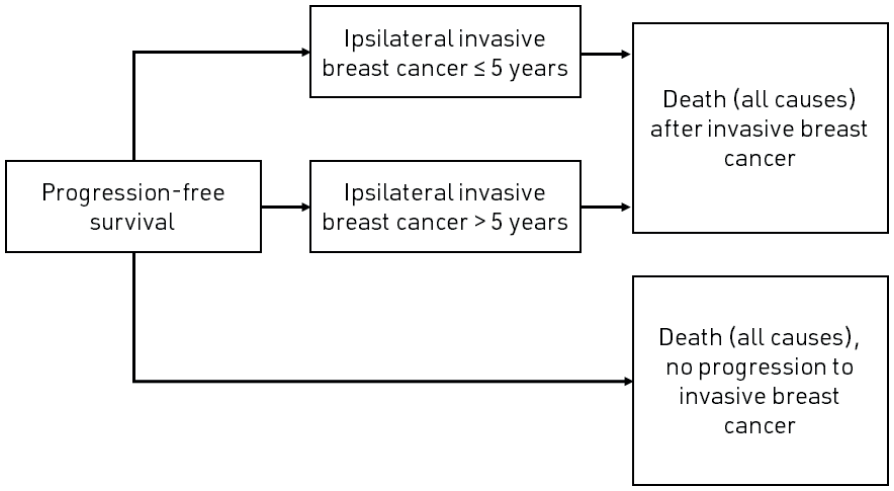


Figure 2. Multi-state model

A graphic representation of the semi-Markov model, also known as a multi-state model, used to simulate the disease process after a primary DCIS diagnosis. Women begin at the progression-free survival state, and can transition to ipsilateral invasive breast cancer (iIBC) within, or after 5 years, or eventually transition to death without experiencing an iIBC.

Target population and subgroups

The source population of this study was a retrospective cohort of screening-age women diagnosed with primary DCIS derived from the Surveillance, Epidemiology, and End Results Program (SEER) cancer registry database. The purpose of using such a population-based cancer registry for information on treatment uptake, DCIS clinicopathological characteristics, and outcomes was to ensure a representative cohort of women who would be considered for a AS strategy, with a representative distribution of characteristics. Furthermore, a large subgroup of untreated women was only feasible to derive from SEER, and not from Dutch registry sources. Using the SEER cancer registry database, we identified N=31,068 women aged 45–75 at the time of DCIS diagnosis. Women with low-risk characteristics (low-to-intermediate grade, ER+) were identified as the low-risk subgroup.

In order to ensure that subsequent iIBCs were abstracted correctly as new primaries, we selected women who were diagnosed with DCIS from 2007 onwards, given the changes in SEER coding rules that required registrars to record subsequent invasive breast cancer following DCIS as a new primary cancer and not a locoregional

invasive recurrence.[33] In the selected cohort, women were diagnosed up to and including 2016. The cohort included women who were treated for their DCIS with breast conserving surgery, with or without radiotherapy, as well as women who did not undergo immediate treatment with surgery.

This cohort is a subset of a larger previously-reported DCIS cohort; covariate selection and missing data imputation steps are described elsewhere.[32] DCIS treatment is modeled based on patterns of care for surgery and radiotherapy observed in SEER for this cohort.

Treatment strategy outcomes

In this decision model, treatment outcomes for women undergoing surgery are based on patient-level outcomes observed directly in the SEER cohort. Conditional transition probabilities were computed by building Cox proportional hazards models stratified by transition to compute cumulative transition hazards transformed into conditional transition probabilities using the Aalen-Johansen estimator.[34] State occupation probabilities at different time points following DCIS diagnosis could be derived from these values. State occupation and transition probabilities are derived separately for women with low-risk (low-to-intermediate grade, ER+) and normal/high risk DCIS characteristics who underwent surgery. To model outcomes for the subset of women opting for the active surveillance strategy, patient-level iIBC outcomes observed in SEER for the subgroup of untreated women with low-risk characteristics were modeled. The remaining transition probabilities from the iIBC states and for all-cause death were taken from the surgery group. Data preparation and multi-state modeling was done using the R package mstate version 0.2.11.

Health-related quality of life

Health-related quality of life (HRQoL) and utility measurements were derived from two studies evaluating health-state and treatment preferences among women with DCIS. Utilities were derived from a study which quantified preferences for managing (low-risk) screen detected DCIS using the EQ-5D-5L approach among women with a personal history of DCIS.[35] This is the only study published to-date which has derived utility values direcly from women with DCIS. A utility decrement for salvage mastectomy was derived from a cost-effectiveness study on adjuvant treatment for low-risk DCIS reported by Ward et al.[19] It is assumed that the utility weights associated with progression remain the entire occupancy time in a given state, up to the full time horizon of 10 years.[35]

Estimating resources and costs

In this model, we assumed that all women surgically-treated for their DCIS will undergo breast conserving surgery and 75% of these women will undergo adjuvant radiotherapy. Women who experience an invasive breast cancer event after their DCIS diagnosis, who were previously treated with breast conserving surgery with or without radiotherapy, are treated with salvage mastectomy, (immediate) breast reconstruction with implant, and adjuvant chemotherapy in line with to European and American clinical guidelines on locoregional recurrence following DCIS.[11,36-38] Women following the active surveillance strategy who experience a subsequent invasive breast cancer are primarily treated with breast conserving surgery, adjuvant radiotherapy and chemotherapy.

Associated healthcare utilization and surgical costs were based on Dutch costs derived from a previously published population-based cost-utility analysis of four common surgical treatment pathways for breast cancer in the Netherlands.[39] This analysis followed a large, representative cohort of women, and included the cost of the surgical intervention, outpatient visits, admission days, diagnostics-related resources, and costs of complications during the treatment. Chemotherapeutic costs for women who experience an invasive breast cancer event are based on a previously published Dutch cost-effectiveness model for women with early-stage breast cancer.[40] All costs were reported in 2020 Euros, with any adjustments made using the Consumer Price Index.

Deterministic analysis

Total costs and quality-adjusted life years (QALYs) were discounted at 4.0% and 1.5%, respectively, according to the Dutch Guidelines for the Conduct of Economic Evaluations in Health Care.[41] The incremental cost-effectiveness ratio (ICER) was calculated as the difference in costs divided by the difference in QALYs between the cohort including an active surveillance strategy, and the cohort following standard interventional treatment.

Sensitivity analyses

One-way sensitivity analyses were performed to assess robustness of model outcomes. Cost parameters were individually assessed at the minimum and maximum values of their range (Table 1) to identify those most influential on incremental costs. Utility parameters were similarly assessed based on the accompanying upper and lower bounds of the reported 95% confidence interval to identify influence over incremental QALYs.

Table 1. Base-case model parameters

Cost	Base Cost (EUR)	Range (EUR)	Distribution	References
Breast conserving surgery, including re-excision	9,636	7,227–12,045	Gamma	Kouwenberg et al. 2021
Whole breast radiotherapy	7,606	5,704–9,508	Gamma	Kouwenberg et al. 2021
Salvage mastectomy	9,553	7,165–11,941	Gamma	Kouwenberg et al. 2021
Immediate implant-based reconstruction following mastectomy	19,554	14,666–24,442	Gamma	Kouwenberg et al. 2021
Chemotherapeutic treatment*	16,600	12,450–20,750	Gamma	Retèl et al. 2020
Mammography	91.97	68.98–114.96	Gamma	CZ Healthcare Insurance. Tarieventool. Accessed November 28, 2021. https://www.cz.nl/service-en-contact/zoek-tarieven
Utility	**Base Estimate**	**Range**	**Distribution**	**References**
Breast conserving surgery alone	0.768	0.696–0.848	Beta	Bromley et al. 2019
Breast conserving surgery with radiotherapy	0.729	0.592–0.837	Beta	Bromley et al. 2019
Active surveillance	0.879	0.848–1.000	Beta	Bromley et al. 2019
Progressed DCIS to invasive breast cancer	0.622	0.388–0.745	Beta	Bromley et al. 2019
Utility decrement of receiving salvage mastectomy	0.180	0.080–0.250	Beta	Ward et al. 2021

All oncologic surgery costs include non-OR, outpatient, admission, diagnostics, and plastic surgery costs.
*Includes chemotherapy costs, infection-prevention medication, outpatient stay costs.

Probabilistic analyses

Probabilistic analyses were used to simultaneously assess the uncertainty of all inputs by randomly drawing cost and utility parameter values from assigned distributions. Beta distributions were used for utilities, and gamma distributions were used for costs. Five thousand Monte Carlo simulation iterations were used. The results of the simulations are illustrated in an incremental cost-effectiveness plane, which visualizes the extent to which including an option for active surveillance is more or less effective and expensive compared to standard immediate interventional treatment. The effectiveness of active surveillance is shown separately as incremental life years and as QALYs gained/lost among all women with DCIS.

Cost-effectiveness acceptability curves (CEACs) were plotted to illustrate the impact of uncertainty on the outcomes, given a range of possible willingness to pay (WTP) thresholds per QALY gained. In the Netherlands, a threshold of €80,000 per QALY is standard for severe diseases, for preventive strategies a lower WTP

level of € 20.000 is used.[42] Across the Monte Carlo simulation iterations, we report incremental discounted costs, life years, QALYs and accompanying 95% credential intervals (CI) for each strategy.

Scenario analysis

Scenario analysis models used information on COX-2 protein expression and breast adipocyte size to select low-risk women to forgo surgery, based on the case-control study of Almekinders et al.[22] According to the study, women with low COX-2 protein expression and with lower relative area of breast adipose tissue (adipocyte area^{75th}) had a cumulative incidence of iIBC similar to the general population. While the study population received BCS only, following expert consultation, the scenario analysis assumes no added value of surgery, and that iIBC incidence remains similar to the general population.

In this scenario analysis, these women (representing ~10% of the screen-detected DCIS population) would be eligible to forgo surgery. Under an active surveillance strategy their risk of iIBC is assumed to remain the same (i.e. BCS is assumed to have no effect), and are therefore considered low-risk. Normal-risk women would be treated with BCS only. Using data on a a subset of screening-age women from the study,[22] we derived transition probabilities for the DCIS to iIBC states for the low-risk subgroup (those with low COX-2 protein expression and adipocyte area^{75th}) and for the remaining normal-risk DCIS population. All other transition probabilities to the remaining health states were derived from the base-case model, using only data on women treated with BCS. A comparison is made to a standard-care strategy, where all women are treated with BCS regardless of risk, and 75% receive adjuvant radiotherapy.

Comparison of low-risk patients by biomarker

In order to better visualize and compare the variability across Monte Carlo simulations, probabilistic results and cost-effectiveness planes are reported separately for the two subsets of women eligible for AS: (1) low-risk women based on strategy B in the base-case model (low-to-intermediate grade, ER+) and (2) low-risk women based on the scenario analysis (low COX-2 protein expression and adipocyte area^{75th}). Incremental QALYs per patient in each biomarker scenario are reported.

Headroom analysis for hypothetical perfect biomarker to select low-risk women

A headroom analysis[43] was conducted to estimate the maximum cost for which a (new) biomarker can be brought to market while maintaining the cost-effectiveness

of its use. This analysis focuses on the biomarkers defined in the scenario analysis: COX-2 protein expression and breast adipocyte size. They are considered 'hypothetically perfect' because outcomes in the low-risk group are assumed to mirror the treated population, as illustrated by Almekinders et al.[22] As this scenario analysis simulated a biomarker with optimal diagnostic accuracy, we used the results to conduct a headroom analysis to determine the maximum possible cost for a hypothetical perfect biomarker solution that would accurately select women who could forgo surgery (assumed to be 10% of entire cohort). For the cost of the new technology to remain cost-effective, the headroom is assessed as the net reduction of costs associated with the strategy, plus the societal WTP threshold for a QALY multiplied by the maximum QALY gain across the entire cohort.[43,44]

$$\begin{aligned} Headroom = &(Net\ reduction\ of\ healthcare\ costs) \\ &+ (WTP\ threshold) \times (Additional\ QALYs\ generated) \end{aligned}$$

Analyses were performed using Microsoft Excel, version 2019 (Microsoft, Redmond, WA) and R (R Project, Vienna, Austria). This report conforms to the Consolidated Health Economic Evaluation Reports standards statement.[45]

Results

Base-case model

Table 2 presents results from the base-case model with 100,000 simulated DCIS patients for each strategy. When an active surveillance strategy is introduced for all low-risk (low-to-intermediate grade and ER+) patients as an alternative to immediate surgery, a QALY gain of 0.4 is achieved for the entire DCIS cohort, with an accompanying average per patient cost saving of 6,353 € over the time horizon. Without adjustment for utilities, this strategy results on average in a limited loss of life years (-0.06, 95% CI -0.26 to 0.16).

The Monte Carlo probabilistic sensitivity analysis demonstrated that introducing an active surveillance strategy produced higher net benefits (i.e. was cost-effective) in 95.7% of the 5,000 model simulations at the 80,000 € WTP threshold, and 99.3% at the 20,000 € WTP threshold (Figure 3). This is similarly illustrated in the cost-effectiveness plane in Figure 4A. All model simulations showed a decrease in incremental costs when introducing AS, while 93% of the simulations showed an increase in incremental QALYs. The trend of improved effectiveness was less apparent when considering life years alone: 69% of the simulations showed a decrease in incremental life years (Figure 4B).

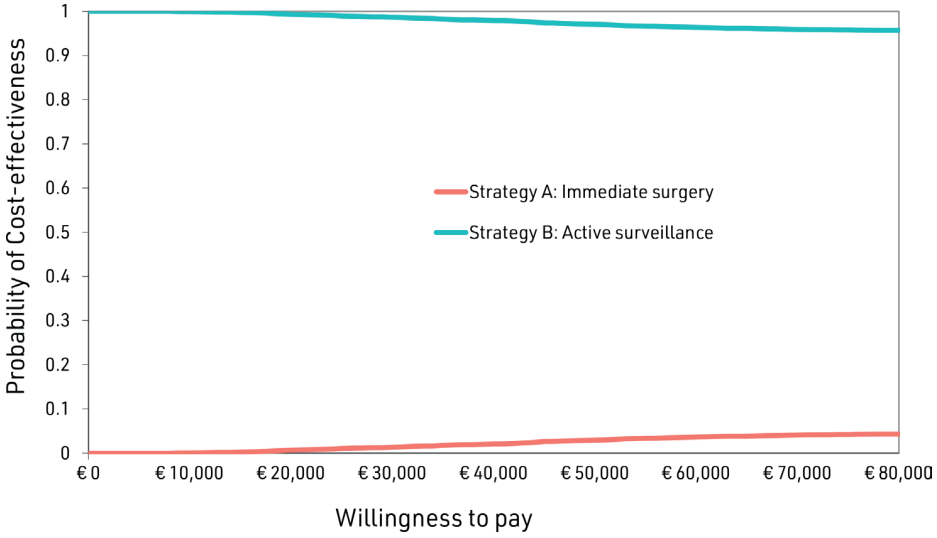


Figure 3. Cost-effectiveness acceptability curve (base-case analysis)

Percentage of time each strategy is cost-effective at varying willingness to pay thresholds are represented. The red line represents the percentage of time the standard immediate surgery (Strategy A) is cost-effective, the blue line for the active surveillance strategy (Strategy B).

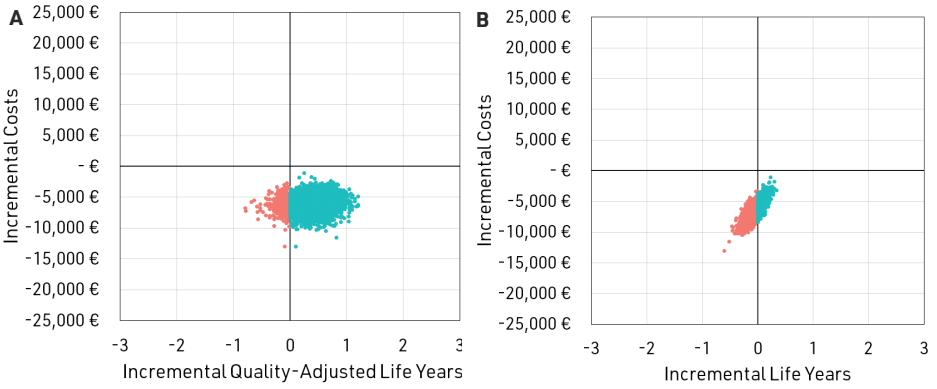


Figure 4. Cost-effectiveness planes (base-case analysis)

Cost-effectiveness planes in the base-case analysis. Each dot on the graph represents the results of the 5,000 Monte Carlo simulation iterations for the probabilistic sensitivity analysis of introducing an active surveillance strategy for women with low-risk DCIS (ER+ and low-/intermediate-grade) compared to standard surgical intervention. In panel A, effectiveness is represented by incremental *quality-adjusted* life years, and panel B by incremental life years. Points shown in blue represent simulations in which introducing an active surveillance strategy resulted in overall absolute health benefits across the simulated cohort.

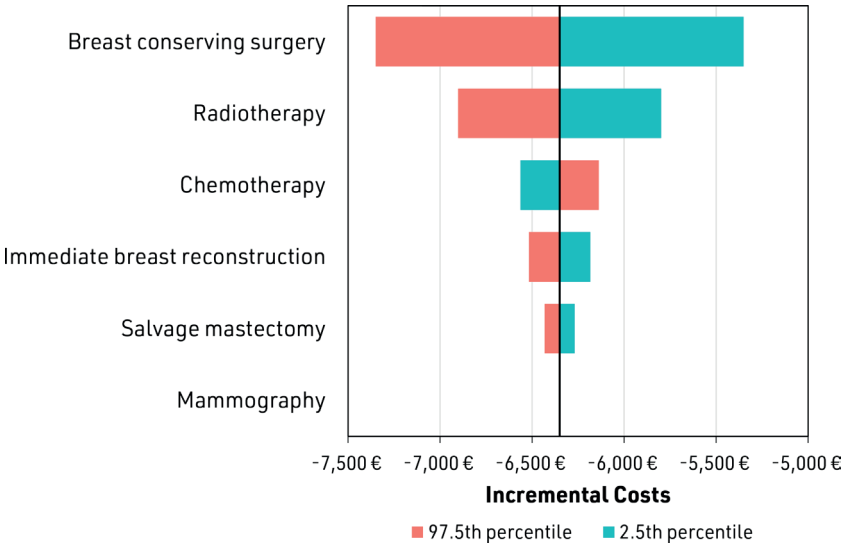


Figure 5A. One-way sensitivity analysis for costs
Tornado diagram showing one-way sensitivity analysis of base-case incremental costs to variations in parameters. Parameters are varied one at a time by the 2.5 and 97.5th percentile of their assigned distribution. Ranking is based on size of effect on incremental costs of introducing an active surveillance strategy.

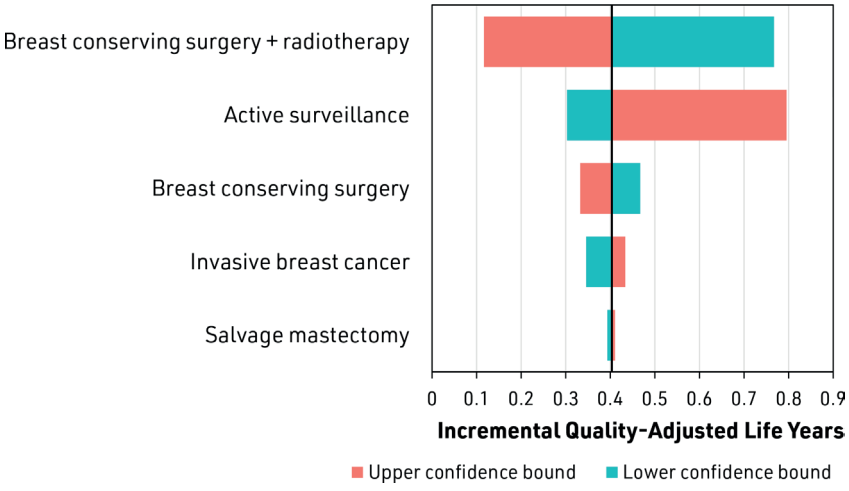


Figure 5B. One-way sensitivity analysis for utilities
Tornado diagram showing one-way sensitivity analysis of base-case incremental quality-adjusted life years (QALYs) to variations in input parameters. Parameters are varied one at a time by the 2.5 and 97.5th percentile of their assigned distribution. Inputs are ranked by size of effect on incremental QALYs of introducing an active surveillance strategy.

Figure 5A shows that for all cost inputs, not one drives cost-effectiveness of the active surveillance strategy. Breast conserving surgery, followed by radiotherapy does however have the largest influence over incremental costs. With decreasing costs of each surgical treatment type, the incremental cost savings for the strategy including active surveillance obviously lessens. Conversely, as more women would experience an iIBC in Strategy B, lowering the cost of chemotherapy would also increase the cost savings for this strategy. In Figure 5B, varying all utilities individually by their lower and upper bounds would still result in a QALY gain for Strategy B. Notably, the utility associated with invasive breast cancer did not have an impact on the cost-effectiveness of introducing an AS strategy.

Scenario Analysis

The results of the scenario analysis are visualized in Figure 6. A de-escalation strategy whereby 10% of women are selected for active surveillance based on a different definition of low-risk status (low COX-2 protein expression and adipocyte area^{75th}) and the remaining undergo BCS, is cost-effective compared to BCS with or without radiotherapy for all women. Compared to the base-case analysis, there is more apparent uncertainty in the model inputs which is indicated by the dispersion of the results of the simulation iterations across the two lower quadrants of Figure 6.

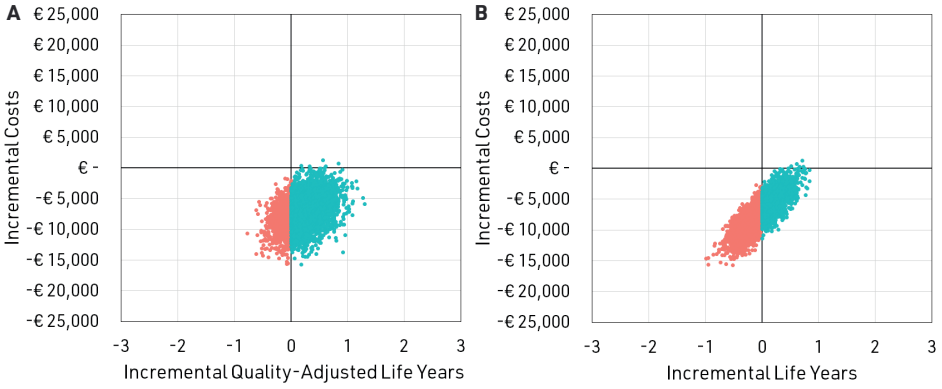


Figure 6. Cost-effectiveness planes (scenario analysis)

Cost-effectiveness planes in the scenario analysis. The graph shows the results of the 5,000 Monte Carlo simulation runs for the probabilistic sensitivity analysis of introducing an active surveillance strategy for women with low-risk DCIS (defined by low COX-2 protein expression and lower relative area of breast adipose tissue) compared to standard surgical intervention. In panel A, effectiveness is represented by incremental ***quality-adjusted*** life years, and panel B by incremental life years. Points shown in blue represent simulations in which introducing an active surveillance strategy resulted in overall absolute health benefits across the simulated cohort.

Comparison of low-risk patients by biomarkers

Figure 7 shows the results of the Monte Carlo simulation iterations for each of the low-risk subgroups explored in the base-case and scenario analyses. 92.8% of model runs resulted in incremental QALY gains for low-risk women defined by low-to-intermediate grade, ER+ DCIS. The model using low COX-2 expression and adipocyte area^{75th} to define low-risk status had greater parameter uncertainty due to the smaller available dataset.[22] However, 97.0% of model runs resulted in incremental QALY gains for this group of low-risk women. As a group, women following the scenario analysis strategy gained 1.02 QALYs (95% CI, -0.05 to 1.98), while women in the base-case strategy gained 0.81 QALYs (95% CI, -0.26 to 1.70).

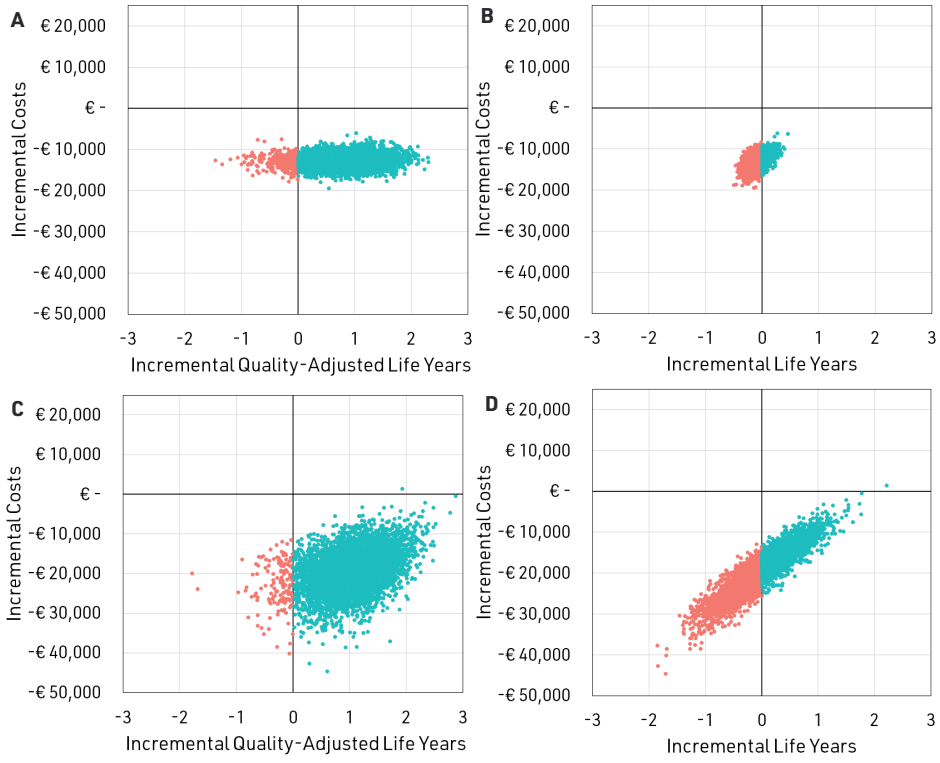


Figure 7. Cost-effectiveness planes for low-risk subgroups only

Cost-effectiveness planes showing the results of shows the results of the 5,000 Monte Carlo simulation runs for the probabilistic sensitivity analysis, using only data on women identified as low-risk. All strategies show cost-savings. Simulations for which an active surveillance strategy resulted in gains in incremental ***quality-adjusted*** life years (panels A and C) and incremental life years (panel B and D) are shown in blue. The greater the variation of simulation points is an indication of the uncertainty in the parameter inputs used.

Headroom Analysis

The use of the hypothetical perfect biomarker to select women for the active surveillance strategy was expected to yield an average cost savings of €7,076 and 0.23 incremental QALYs per patient based on the scenario analysis for the total DCIS cohort (Table 2).

$$Headroom = €7{,}076 + €20{,}000 \times 0.23$$

Assuming that this biomarker strategy could reliably select 10% of the DCIS cohort as low-risk, the headroom available for this strategy was calculated as €6,227 for a WTP threshold of €20,000. This is the maximum unit cost for which this biomarker strategy can be brought to market while remaining cost-effective.

Table 2. Probabilistic results

Strategy	Incremental costs, 95% CI	Incremental life years, 95% CI	Incremental QALYs, 95% CI	ICER costs per QALY
Base case: Whole DCIS cohort, 10-year time horizon. Simulation based on SEER data. **Strategy A:** Standard surgical intervention (BCS±RT) for all women **Vs.** **Strategy B:** Women with ER+, Grade I/II DCIS (50% of whole cohort) undergo AS; remaining undergo standard surgical intervention (BCS±RT)	-6,353 € (-8,811, -3,966 €)	-0.06 (-0.26, 0.16)	0.40 (-0.15, 0.87)	-15,981 €
Scenario analysis: Whole DCIS cohort, 10-year time horizon. Simulation based on SEER data and Almekinders et al. **Strategy A:** Standard surgical intervention (BCS±RT only) for all women **Vs.** **Strategy B:** Women with Low COX2 / Adipocyte area^{75th} Quartile 1 DCIS (10% of whole cohort) undergo AS; remaining undergo standard surgical intervention with BCS only	-7,076 € (-11,978, -2,427 €)	0.00 (-0.49, 0.48)	0.23 (-0.30, 0.76)	-31,071 €

Discussion

DCIS is a potential precursor to invasive breast cancer. At present, nearly all women with DCIS are treated in the same manner as early breast cancer as it is not possible to reliably predict who may progress to invasive disease. Due to this uncertainty many women with harmless DCIS are treated. This early economic evaluation demonstrated that introducing an active surveillance option to select women with low-risk features can be a cost-effective alternative to immediate surgery and adjuvant radiotherapy. Women with low-to-intermediate grade, ER+ DCIS make up approximately 50% of screen-detected primary DCIS. In this analysis, forgoing surgery resulted in significant gains in quality of life, despite an expected elevated rate of iIBC and somewhat reduced life years in this group.

The results of this and other published studies[19,20,46] on locoregional treatment for DCIS highlight the potential for a range of de-escalation strategies. However, the analyses presented here are limited by the use of data from different sources and different country settings. SEER data was chosen to model outcomes after DCIS given the volume of data available and the rigorous coding rules set forth in 2007 requiring subsequent iIBC to be recorded as new primaries. Despite the availability of data from retrospective DCIS cohorts in the Netherlands,[47,48] SEER data has been used extensively to model the natural disease history of DCIS[49,50] and the impact of DCIS treatment, and was therefore chosen for this study. Nevertheless, this study remains an early cost-effectiveness analysis which may have limited generalizability and applicability at this time.

We observed substantial variation in incremental QALYs across simulation iterations. This can be reflective of the variation in preferences regarding benefits to be gained from undergoing active surveillance. The overall positive impact of quality-adjustment does become particularly apparent in the one-way sensitivity analysis (Figure 5B). Even when applying the lowest value associated with the 2.5th percentile of the utilities' assigned distributions, resulting incremental QALYs always remained greater than 0.1. These results may foreshadow the forthcoming results of the Dutch LORD study if the 10-year risk of iIBC remains within the pre-specified non-inferiority threshold.

Whether active surveillance will eventually become an acceptable alternative to surgery still remains uncertain. While half of women with screen-detected primary DCIS would be eligible based on the low-risk criteria of the base-case model, women's preferences and their access to regular surveillance imaging will dictate whether such a strategy is suitable to them. A discrete choice experiment among

Dutch women with newly diagnosed low-risk DCIS found very strong preferences for an active surveillance strategy with no surgery, irrespective of the 10-year risk of iIBC.[23] This preference was also highly related to the desire for consistent follow-up with annual mammography.

A biomarker which selects a more defined group of women who stand to gain no benefit from surgical intervention could alleviate some of the uncertainty and variation around the benefits of an active surveillance strategy. We used a scenario analysis to explore how such a hypothetically perfect biomarker could be used in combination with a de-escalation strategy for all women with screen-detected primary DCIS. The biomarker is considered 'perfect' because iIBC rates in the low-risk group would match a healthy population, as demonstrated by Almekinders et al.[22] The scenario analysis showed a higher QALY gain among this group: 1.02 incremental QALYs compared to 0.81 QALYS among the low-risk group defined in the base case analysis.

In this scenario analysis, a smaller proportion of women (10%) were selected for active surveillance based on COX-2 expression and small adipocyte size, and the remaining women underwent de-escalation to BCS only. The choice of modeling this prevalent de-escalation strategy was intended to address the questions surrounding overtreatment across all women with DCIS. Incremental life years remained unaffected in the scenario analysis, reflecting results from previously published observational cohort studies and randomized controlled trials demonstrating that while RT may reduce rates of local recurrence, it provides no improvement of overall-, distant-metastasis-free, or cancer-specific-survival.[51-55] Previously published cost-effectiveness analysis of adjuvant RT for low-risk DCIS similarly demonstrated cost savings for a BCS only strategy; adding RT would result in negligible QALY gains at significant costs.[19,20]

One benefit of conducting early cost-effectiveness analyses is that it creates an opportunity to understand the potential of new strategies under the most optimistic assumptions. We performed a headroom analysis based on a simulation of the impact of a hypothetical perfect biomarker using COX-2 and adipose area information. If this biomarker information could indeed accurately select a group of women who could forgo surgery without negatively impacting life years, the potential downstream effects are significant. For such a biomarker-based strategy to remain cost-effective at a WTP threshold of €20,000, the upper ceiling price for this could be set at €6,227.

A major shift in contemporary thinking about disease management for DCIS is underway.[12,56,57] Existing literature detailing the heterogeneous nature of DCIS now supports a more individualized approach based on individual risk of progression to invasive breast cancer. But before active surveillance is brought into clinical practice, robust evidence from the LORD, LORIS, and COMET trials must first confirm the safety of this approach. This study contributes to the knowledge base on DCIS management, and supports the continuation of research on identifying and validating biomarkers that can select women to safely forgo surgery.

References

1. Klotz L, Vesprini D, Sethukavalan P, et al. Long-term follow-up of a large active surveillance cohort of patients with prostate cancer. *J Clin Oncol* 2015; **33**(3): 272-7.
2. Elshof LE, Tryfonidis K, Slaets L, et al. Feasibility of a prospective, randomised, open-label, international multicentre, phase III, non-inferiority trial to assess the safety of active surveillance for low risk ductal carcinoma in situ - The LORD study. *Eur J Cancer* 2015; **51**(12): 1497-510.
3. Francis A, Thomas J, Fallowfield L, et al. Addressing overtreatment of screen detected DCIS; the LORIS trial. *Eur J Cancer* 2015; **51**(16): 2296-303.
4. Hwang ES, Hyslop T, Lynch T, et al. The COMET (Comparison of Operative versus Monitoring and Endocrine Therapy) trial: a phase III randomised controlled clinical trial for low-risk ductal carcinoma in situ (DCIS). *BMJ Open* 2019; **9**(3): e026797.
5. Wallis M, Bartlett J, Billingham L, et al. The LORIS Trial: randomising patients with lowor low intermediate-grade ductal carcinoma in situ (DCIS) to surgery or active monitoring. *Breast Cancer Res Treat* 2018; **167**(1): 325-6.
6. Visser LL, Groen EJ, van Leeuwen FE, Lips EH, Schmidt MK, Wesseling J. Predictors of an Invasive Breast Cancer Recurrence after DCIS: A Systematic Review and Meta-analyses. *Cancer Epidemiol Biomarkers Prev* 2019; **28**(5): 835-45.
7. Pinder SE, Thompson AM, Wesserling J. Low-risk DCIS. What is it? Observe or excise? *Virchows Archiv* 2021: 1-12.
8. Luiten JD, Luiten EJT, van der Sangen MJC, et al. Patterns of treatment and outcome of ductal carcinoma in situ in the Netherlands. *Breast Cancer Res Treat* 2021; **187**(1): 245-54.
9. Yashkin AP, Greenup RA, Gorbunova G, Akushevich I, Oeffinger KC, Hwang ES. Outcomes and Costs for Women After Breast Cancer: Preparing for Improved Survivorship of Medicare Beneficiaries. *JCO Oncol Pract* 2021; **17**(4): e469-e78.
10. Gradishar WJ, Anderson BO, Balassanian R, et al. Breast cancer, version 4.2017, NCCN clinical practice guidelines in oncology. *JNCCN* 2018; **16**(3): 310-20.
11. Cardoso F, Kyriakides S, Ohno S, et al. Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2019; **30**(8): 1194-220.
12. Hwang ES, Solin L. De-Escalation of Locoregional Therapy in Low-Risk Disease for DCIS and Early-Stage Invasive Cancer. *J Clin Onocol* 2020; **38**(20): 2230-9.
13. Nofech-Mozes S, Hanna W, Rakovitch E. Molecular Evaluation of Breast Ductal Carcinoma in Situ with Oncotype DX DCIS. *Am J Clin Pathol* 2019; **189**(5): 975-80.
14. Shah C, Bremer T, Cox C, et al. The Clinical Utility of DCISionRT® on Radiation Therapy Decision Making in Patients with Ductal Carcinoma In Situ Following Breast-Conserving Surgery. *Ann Surg Oncol* 2021; **28**(11): 5974-84.
15. Raldow AC, Sher D, Chen AB, Recht A, Punglia RS. Cost Effectiveness of the Oncotype DX DCIS Score for Guiding Treatment of Patients With Ductal Carcinoma In Situ. *J Clin Oncol* 2016; **34**(33): 3963-8.
16. Raldow AC, Sher D, Chen AB, Punglia RS. Cost Effectiveness of DCISionRT for Guiding Treatment of Ductal Carcinoma in Situ. *JNCI Cancer Spectr* 2020; **4**(2): pkaa004.
17. Kim H, Vargo JA, Smith KJ, Beriwal S. Cost-Effectiveness Analysis of Biological Signature DCISionRT Use for DCIS Treatment. *Clin Breast Cancer* 2021; **21**(3): e271-e8.
18. McCormick B, Winter KA, Woodward W, et al. Randomized Phase III Trial Evaluating Radiation Following Surgical Excision for Good-Risk Ductal Carcinoma In Situ: Long-Term Report From NRG Oncology/RTOG 9804. *J Clin Oncol* 2021.

19. Ward MC, Vicini F, Al-Hilli Z, et al. Cost-Effectiveness Analysis of No Adjuvant Therapy Versus Partial Breast Irradiation Alone Versus Combined Treatment for Treatment of Low-Risk DCIS: A Microsimulation. *JCO Oncol Pract* 2021; **17**(8): e1055-e74.
20. Gupta A, Jhawar SR, Sayan M, et al. Cost-Effectiveness of Adjuvant Treatment for Ductal Carcinoma In Situ. *J Clin Oncol* 2021; **39**(21): 2386-96.
21. Visser LL, Elshof LE, Schaapveld M, et al. Clinicopathological Risk Factors for an Invasive Breast Cancer Recurrence after Ductal Carcinoma In Situ-A Nested Case-Control Study. *Clin Cancer Res* 2018; **24**(15): 3593-601.
22. Almekinders MMM, Schaapveld M, Thijssen B, et al. Breast adipocyte size associates with ipsilateral invasive breast cancer risk after ductal carcinoma in situ. *NPJ Breast Cancer* 2021; **7**(1): 31.
23. Byng D, Retèl VP, Engelhardt EG, et al. Preferences of Treatment Strategies among Women with Low-Risk DCIS and Oncologists. *Cancers (Basel)* 2021; **13**(16).
24. Rakovitch E, Sutradhar R, Hallett M, et al. The time-varying effect of radiotherapy after breast-conserving surgery for DCIS. *Breast Cancer Res Treat* 2019; **178**(1): 221-30.
25. van Seijen M, Lips EH, Fu L, et al. Long-term risk of subsequent ipsilateral lesions after surgery with or without radiotherapy for ductal carcinoma in situ of the breast. *Br J Cancer* 2021; **125**(10): 1443-9.
26. Sagara Y, Mallory MA, Wong S, et al. Survival Benefit of Breast Surgery for Low-Grade Ductal Carcinoma In Situ: A Population-Based Cohort Study. *JAMA Surg* 2015; **150**(8): 739-45.
27. Soares MO, Canto ECL. Continuous time simulation and discretized models for cost-effectiveness analysis. *Pharmacoeconomics* 2012; **30**(12): 1101-17.
28. van Rosmalen J, Toy M, O'Mahony JF. A mathematical approach for evaluating Markov models in continuous time without discrete-event simulation. *Med Decis Making* 2013; **33**(6): 767-79.
29. Briggs A, Sculpher M. An introduction to Markov modelling for economic evaluation. *Pharmacoeconomics* 1998; **13**(4): 397-409.
30. Williams C, Lewsey JD, Briggs AH, Mackay DF. Cost-effectiveness analysis in R using a multi-state modeling survival analysis framework: a tutorial. *Med Decis Making* 2017; **37**(4): 340-52.
31. Zaric GS. The impact of ignoring population heterogeneity when Markov models are used in cost-effectiveness analysis. *Med Decis Making* 2003; **23**(5): 379-86.
32. Byng D, Retèl VP, Schaapveld M, Wesseling J, van Harten WH. Treating (low-risk) DCIS patients: What can we learn from real-world cancer registry evidence? *Breast Cancer Res Treat* 2021; **187**(1): 187-96.
33. Habel LA, Buist DSM. Re: Cancer Outcomes in DCIS Patients Without Locoregional Treatment. *J Natl Cancer Inst* 2019; **112**(2): 214-5.
34. Putter H, Spitoni C. Non-parametric estimation of transition probabilities in non-Markov multi-state models: the landmark Aalen–Johansen estimator. *Stat Methods Med Res* 2018; **27**(7): 2081-92.
35. Bromley HL, Mann GB, Petrie D, Nickson C, Rea D, Roberts TE. Valuing preferences for treating screen detected ductal carcinoma in situ. *Eur J Cancer* 2019; **123**: 130-7.
36. van Huizum MA, Hage JJ, Rutgers EJ, Hoornweg MJ. Immediate breast reconstruction with a myocutaneous latissimus dorsi flap and implant following skin-sparing salvage mastectomy after irradiation as part of breast-conserving therapy. *J Plast Reconstr Aesthet Surg* 2016; **69**(8): 1080-6.

37. Cardoso F, Harbeck N, Fallowfield L, Kyriakides S, Senkus E. Locally recurrent or metastatic breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2012; **23 Suppl 7**: vii11-9.
38. Gradishar WJ, Moran MS, Abraham J, et al. NCCN guidelines® insights: Breast cancer, version 4.2021: Featured updates to the NCCN guidelines. *JNCCN* 2021; **19**(5): 484-93.
39. Kouwenberg CAE, Mureau MAM, Kranenburg LW, et al. Cost-utility analysis of four common surgical treatment pathways for breast cancer. *Eur J Surg Oncol* 2021; **47**(6): 1299-308.
40. Retèl VP, Byng D, Linn SC, et al. Cost-effectiveness analysis of the 70-gene signature compared with clinical assessment in breast cancer based on a randomised controlled trial. *Eur J Cancer* 2020; **137**: 193-203.
41. Roijen L, Linden N, Bouwmans C, Kanters T, Tan S. Kostenhandleiding: Methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg. *Zorginstituut Nederland The Netherlands: Zorginstituut Nederland* 2015.
42. Versteegh MM, Ramos IC, Buyukkaramikli NC, Ansaripour A, Reckers-Droog VT, Brouwer WBF. Severity-Adjusted Probability of Being Cost Effective. *Pharmacoeconomics* 2019; **37**(9): 1155-63.
43. Girling A, Lilford R, Cole A, Young T. Headroom approch to device development: Current and future directions. *Int J Technol Assess Health Care* 2015; **31**(5): 331-8.
44. Cosh E, Girling A, Lilford R, McAteer H, Young T. Investing in new medical technologies: a decision framework. *J Commer Biotechnol* 2007; **13**(4): 263-71.
45. Husereau D, Drummond M, Petrou S, et al. Consolidated health economic evaluation reporting standards (CHEERS) statement. *Int J Technol Assess Health Care* 2013; **29**(2): 117-22.
46. Trentham-Dietz A, Ergun MA, Alagoz O, et al. Comparative effectiveness of incorporating a hypothetical DCIS prognostic marker into breast cancer screening. *Breast Cancer Res Treat* 2018; **168**(1): 229-39.
47. Elshof LE, Schaapveld M, Schmidt MK, Rutgers EJ, van Leeuwen FE, Wesseling J. Subsequent risk of ipsilateral and contralateral invasive breast cancer after treatment for ductal carcinoma in situ: incidence and the effect of radiotherapy in a population-based cohort of 10,090 women. *Breast Cancer Res Treat* 2016; **159**(3): 553-63.
48. Elshof LE, Schmidt MK, Rutgers EJT, van Leeuwen FE, Wesseling J, Schaapveld M. Cause-specific Mortality in a Population-based Cohort of 9799 Women Treated for Ductal Carcinoma In Situ. *Ann Surg* 2018; **267**(5): 952-8.
49. Ryser MD, Worni M, Turner EL, Marks JR, Durrett R, Hwang ES. Outcomes of Active Surveillance for Ductal Carcinoma in Situ: A Computational Risk Analysis. *J Natl Cancer Inst* 2016; **108**(5).
50. Ryser MD, Weaver DL, Zhao F, et al. Cancer Outcomes in DCIS Patients Without Locoregional Treatment. *J Natl Cancer Inst* 2019; **111**(9): 952-60.
51. Li PC, Zhang Z, Cronin AM, Punglia RS. Mortality after invasive second breast cancers following prior radiotherapy for DCIS. *JNCCN* 2019; **17**(11): 1367-71.
52. Correa C, McGale P, Taylor C, et al. Overview of the randomized trials of radiotherapy in ductal carcinoma in situ of the breast. *J Natl Cancer Inst Monographs* 2010; **2010**(41): 162-77.
53. Donker M, Litiere S, Werutsky G, et al. Breast-conserving treatment with or without radiotherapy in ductal carcinoma in situ: 15-year recurrence rates and outcome after a recurrence, from the EORTC 10853 randomized phase III trial. *J Clin Oncol* 2013; **31**(32): 4054-9.
54. Wapnir IL, Dignam JJ, Fisher B, et al. Long-term outcomes of invasive ipsilateral breast tumor recurrences after lumpectomy in NSABP B-17 and B-24 randomized clinical trials for DCIS. *J Natl Cancer Inst* 2011; **103**(6): 478-88.

55. Wärnberg F, Garmo H, Emdin S, et al. Effect of radiotherapy after breast-conserving surgery for ductal carcinoma in situ: 20 years follow-up in the randomized SweDCIS Trial. *J Clin Oncol* 2014; **32**(32): 3613-8.
56. van Seijen M, Lips EH, Thompson AM, et al. Ductal carcinoma in situ: to treat or not to treat, that is the question. *Br J Cancer* 2019; **121**(4): 285-92.
57. Groen EJ, Elshof LE, Visser LL, et al. Finding the balance between over- and under-treatment of ductal carcinoma in situ (DCIS). *Breast* 2017; **31**: 274-83.


6

Chapter 6

Cost-effectiveness analysis of the 70-gene signature compared with clinical assessment in breast cancer based on a randomised controlled trial

Danalyn Byng [1,2*], Valesca P. Retèl [1,2*], Sabine C. Linn [3,4,5], Katarzyna Jóźwiak [1], Hendrik Koffijberg [2], Emiel J. Rutgers [6], Fatima Cardoso [7], Martine J Piccart [8], Coralie Poncet [9], Laura J. van 't Veer [10], Wim H. van Harten [1,2]

[1] Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands.

[2] Health Technology and Services Research Department, Technical Medical Centre, University of Twente, Enschede, The Netherlands.

[3] Division of Molecular Pathology, The Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands.

[4] Department of Medical Oncology, The Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands.

[5] Department of Pathology, University Medical Centre, Utrecht University, Utrecht, The Netherlands.

[6] Department of Surgery, The Netherlands Cancer Institute-Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands.

[7] Breast Unit, Champalimaud Clinical Center/Champalimaud Foundation, Lisbon, Portugal.

[8] Breast Cancer Translational Research Laboratory J.-C. Heuson, Institut Jules Bordet, Université Libre de Bruxelles, Brussels, Belgium.

[9] European Organisation for Research and Treatment of Cancer, Brussels, Belgium.

[10] Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, California, United States of America.

* Contributed equally.

Eur J Cancer **137,** 193–2013 (2020). https://doi.org/10.1016/j.

Highlights

- The 70-gene signature is cost-effective compared with clinical assessment alone
- Minimal survival differences are balanced against quality of life gains
- Results are consistent amongst the five studied countries

Abstract

Background

The clinical utility of the 70-gene signature (MammaPrint®) to guide chemotherapy use in T1-3N0-1M0 breast cancer was demonstrated in the Microarray in Node-Negative and 1 to 3 Positive Lymph Node Disease May Avoid Chemotherapy (MINDACT) study. One thousand four ninety seven of 3356 (46.2%) enrolled patients with high clinical risk (in accordance with the modified Adjuvant! Online clinical-pathological assessment) had a low-risk 70-gene signature. Using patient-level data from the MINDACT trial, the cost-effectiveness of using the 70-gene signature to guide adjuvant chemotherapy selection for clinical high risk, estrogen receptor positive (ER+), human epidermal growth factor 2 negative (HER2-) patients was analysed.

Patients and methods

A hybrid decision tree-Markov model simulated treatment strategies in accordance with the 70-gene signature with clinical assessment versus clinical assessment alone, over a 10-year time horizon. Primary outcomes were quality-adjusted life years (QALYs), country-specific costs and incremental cost-effectiveness ratios (ICERs) for six countries: Belgium, France, Germany, Netherlands, UK and the US.

Results

Treatment strategies guided by the 70-gene signature result in more QALYs compared with clinical assessment alone. Costs of the 70-gene signature strategy were lower in five of six countries. This led to dominance of the 70-gene signature in Belgium, France, Germany, Netherlands and the US and to a cost-effective situation in the UK (ICER £22,910/QALY). Annual national cost savings were €4.2M (Belgium), €24.7M (France), €45.1M (Germany), €12.7M (Netherlands) and $244M (US). UK budget increase was £8.4M.

Conclusion

Using the 70-gene signature to safely guide chemotherapy de-escalation in clinical high risk patients with ER+/HER2- tumours is cost-effective compared with using clinical assessment alone. Long-term follow-up and outcomes from the MINDACT trial are necessary to address uncertainties in model inputs.

Introduction

Genomic profiling is a crucial tool to inform prognosis and support treatment decisions in the adjuvant setting. The recognition that patients with early breast cancer may be overtreated necessitated reliable prognostic tools to aid in therapy de-escalation. De-escalation simultaneously addresses the prioritisation of a patient's quality of life and reduces the strain on healthcare systems by avoiding high-cost treatments offering no additional or very limited benefit to a patient's survival.

The phase III EORTC 10041/BIG 3–04 Microarray in Node-Negative and 1 to 3 Positive Lymph Node Disease May Avoid Chemotherapy (MINDACT) trial (NCT00433589) was an international, prospective, randomised study evaluating the clinical utility of the 70-gene expression signature (MammaPrint®) combined with clinical-pathological criteria for selection of patients for adjuvant chemotherapy in breast cancer.[1] The trial enrolled patients with histologically-confirmed invasive breast cancer, with operable T1-3 disease and up to three positive lymph nodes. Five-year median results demonstrated that forgoing adjuvant chemotherapy in patients with high-risk clinical-pathological features, but whom are low-risk according to the 70-gene signature, does not compromise relapse and survival outcomes. The short- and long-term treatment-related adverse events of chemotherapy could be avoided, given the rate of distant metastasis free survival (DMFS) at five years that was 94.7% (95% confidence interval: 92.5%–96.2%), which remained above the pre-determined non-inferiority threshold of 92.0%.

This study reports a cost-effectiveness analysis of treatment strategies guided by the 70-gene signature versus treatment decisions based on clinical risk assessment for a target group of patients with ER+/HER2- breast cancer considered to be clinically high risk. The use of genomic signatures is recommended for this subset of patients by national and international clinical guidelines, such as those arising from the St. Gallen Consensus Conference, European Society for Medical Oncology (ESMO) and American Society of Clinical Oncology.[2-4] Patient-level data were sourced from the MINDACT trial to model relapse and survival outcomes over a ten-year time horizon. This was the most clinically relevant horizon, given the availability of 5-year follow-up data from the MINDACT trial, and the large risk-reducing effect of adjuvant chemotherapy within the first five years.[5] As treatments, test adherence and costs vary widely across countries, analyses are conducted separately for five European countries participating in the MINDACT trial and the US.

Patients and Methods

Model description

A hybrid decision tree-Markov cohort model was constructed. Three mutually exclusive health states were defined as follows: distant metastasis free survival (DMFS), distant metastasis and death (Fig. 1). Costs for six countries (Belgium, France, Germany, Netherlands, United Kingdom and the United States) were applied separately in the model. For all countries, a healthcare perspective was adopted. The model was constructed with a ten-year time horizon and six-month cycle length. Total costs and quality-adjusted life years (QALYs) were discounted at country-specific rates (Table S1). The model compared two strategies: (1) treatment strategies guided by the 70-gene signature in combination with clinical assessment and (2) treatment strategies guided by clinical assessment alone.

The primary outcome of interest, incremental cost-effectiveness ratios (ICERs), was calculated by dividing incremental costs by incremental QALYs. The ISPOR Guidelines for Good Modelling Practices and Cost-Effectiveness Alongside Clinical Trials were used for building the model.[6] The model was programmed in Microsoft Excel, version 2010 (Microsoft, Redmond, WA).

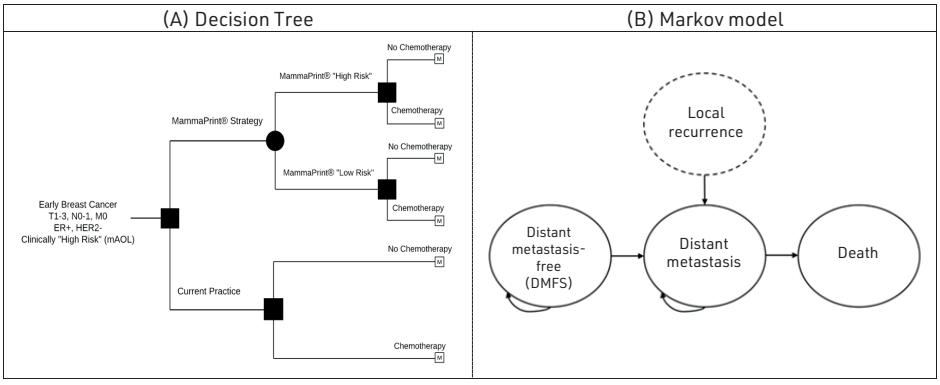


Figure 1. Hybrid decision tree-Markov model. The hybrid (A) decision tree/(B) Markov model structure used to estimate costs, clinical outcomes and quality-adusted life years of using MammaPrint® compared with current practice (clinical risk assessment using the modified Adjuvant! Online [mAOL]) in patients with ER+, HER2- early breast cancer. In the decision tree, a square node represents the decision node at entry, the filled circles are chance nodes, and the squares with the letter M represent Markov nodes.

Population

The MINDACT trial used the modified version of Adjuvant! Online as a clinical-pathological assessment for all patients enrolled in the study. Based on this assessment, all patients were assigned a binary 'high' or 'low' clinical risk score.[1] For our studied patient population and base-case cost-effectiveness analysis, we used data on all patients identified as 'high risk' with ER+/HER2- tumours (n = 2297) (Table 3). Therefore we compared two simulated populations and their associated treatment strategies: (1) patients assessed as clinically 'high risk' with ER+/HER2- tumours who do not undergo genomic profiling, with adjuvant treatment decisions based solely on clinical-pathological characteristics; (2) patients assessed as clinically 'high risk' with ER+/HER2- tumours who undergo genomic profiling with the 70-gene signature, with adjuvant treatment decisions based on their clinical and genomic risk.

Probabilities

Survival probabilities and extrapolations

Using patient-level data on clinically 'high' risk individuals with ER+/HER2- tumours from the MINDACT study, two event end points were evaluated in the model: DMFS, defined as time until the first distant metastatic recurrence or death from any cause, and overall survival (OS), defined as time until death from any cause. Patients were censored at last examination date, if no event was experienced. Survival and hazard rates for each risk and treatment allocation group, based upon the intention-to-treat population, were modelled with three different parametric survival distributions (Weibull, Gompertz, exponential) to estimate rates for the observed five-year follow-up period and for extrapolation to ten years. These survival distributions were fitted to 1000 bootstrapped samples to obtain standard errors of the survival and hazard rates. The Weibull survival distribution was selected for the full extrapolated model based upon the known treatment effect of chemotherapy beyond five years.[5] Interval-specific conditional survival probabilities and associated standard errors were used as the transition probabilities for each cycle of the model. Analyses were conducted with Stata, version 13. A detailed description and visualisation of the parametric modelling approach, including conditional survival probabilities are shown in Supplementary Methods 1, Tables S2–S3, and Fig. S1.

Other probabilities

In accordance with the ESMO guidelines, the parameters in the base-case model reflect the standard treatment pathway of patients with early breast cancer.[3] If current national treatment guidelines deviate from the ESMO guidelines, these were captured as country-specific treatment assumptions (Table 1, and Supplementary Methods 2–3). Adherence to chemotherapy recommendations guided by the 70-gene

signature was based upon real-world values for a patient cohort with predominantly ER+/HER2- tumours reported by Kuijer et al.[7] These patients demonstrated 95% adherence towards the 70-gene signature test results. Adherence following the clinical risk assessment alone was based on expert opinion used by the National Institute for Health and Care Excellence diagnostics assessment program (in 77% of clinical high-risk patients, chemotherapy is recommended).[8]

Health effects

Health-related quality of life (HRQol) was modelled by assigning utilities to the different health states. Baseline clinical-genomic subgroup-specific health state utility values were drawn from a study which captured patient well-being specific to receiving the results of their clinical and gene expression profile. These were measured with the three-level EuroQoL (EQ-5D-3L) amongst 800 enrolled patients in the MINDACT trial (Supplementary Methods 4).[9] This baseline utility value was used for the first six-month cycle for all patients in the DMFS state. In the second cycle, patients remaining in the DMFS state revert to the utility value reported by Lidgren et al.[10] For patients receiving adjuvant chemotherapy, HRQoL decrements drawn from Campbell et al.[11] were applied for the first three cycles in the DMFS state to reflect the negative impacts of chemotherapy (i.e. chemotherapy-related adverse events) on underlying HRQoL during and immediately following chemotherapy. The utility value for the experience of acute myeloid leukaemia as a rare late-effect chemotherapy-related adverse event (cumulative ten-year probability of 0.0049[12] was also applied to patients in the DMFS state and drawn from Younis et al.[13] Utility values for the distant metastasis state were also drawn from Lidgren et al.[10]

Costs

Costs of the 70-gene signature were provided by Agendia NV. This included transport, local specimen processing and value-added tax. Treatment costs were obtained from multiple sources: national drug databases, literature and from governmental white papers on coverage decisions for the 70-gene signature. All direct medical costs relevant to the treatment and disease pathway (from initial treatment to death) for ER+ patients are considered, including costs of endocrine treatment and local/regional recurrence which are not expected to differ between strategies; all are listed in detail in Table S4. European country costs are expressed in 2017/18 Euros, UK costs are expressed in 2017/18 pound sterling and US costs are expressed in 2017/18 dollars.

Probabilistic analysis

Cost and utility parameter values were randomly drawn from assigned distributions. Five thousand Monte Carlo simulation iterations were used. The results of the

simulation are illustrated in an Incremental Cost-Effectiveness plane. To show decision uncertainty, cost-effectiveness acceptability curves (CEACs) are presented.[14] CEACs show the probability that a strategy has the highest net monetary benefit, given a range of willingness-to-pay (WTP) thresholds. A strategy is deemed cost-effective depending on how much society is willing to pay for a gain in effect (i.e. per QALY gained). The World Health Organization has previously proposed a WTP threshold of one to three times annual gross domestic product per capita, although some countries follow other approaches to determining an appropriate threshold which may be more conservative.[15] We assumed an average WTP threshold of €30,000 for Europe, £20,000-£30,000 for the UK and $50,000-$100,000 for the US.

Sensitivity and scenario analyses

To test the robustness of model outcomes, a series of sensitivity analyses were performed. Cost and utility parameters were individually assessed at the 2.5 and 97.5th percentile of their assigned distribution to identify those most influential on incremental costs and incremental QALYs. Two alternative parametric distributions (exponential and Gompertz) were used in the modelling and extrapolation of DMFS and OS. To model the possibility that patients return to the same quality of life as the general population, we apply country-specific population utility norms for the distant metastasis free state, based off the values reported by Janssen et al.[16]

Finally, because adherence to guidelines can vary widely (e.g. from 40 to 99% in the Netherlands),[17] a two-way table was constructed varying the chemotherapy adherence proportions under both treatment strategies to demonstrate how this impacts costs, QALYs and ICERs. In a scenario analysis, disease-free survival was used as an alternative health state to DMFS to capture locoregional recurrences (Supplementary Methods 6).

Budget impact based on costs per population

In the countries examined in this study, early-stage breast cancer (stage I and II) comprises approximately 90% of breast cancers diagnosed, with ~70–75% of these cases being ER+/HER2- by clinical-pathological assessment.[18-23] Total costs for the 70-gene signature strategy were therefore multiplied by the current country-specific annual incidence of eligible patients in the target group. For Belgium this amounted to an incidence of 4000/year[21], for France 20,000/year[22], for Germany 24,000/year[20], for the NL 5000/year[23], for the UK 19,000 patients/year[18], and for the US 85,000/year.[19]

Model validation

The cost-effectiveness model was validated using the Assessment of the Validation Status of Health-Economic decision models[24] tool (described in Supplementary Methods 5) and evaluated by two external experts.

Results

Mean results of the base-case analysis

The 70-gene signature-guided strategies gained 0.02 QALYs for all countries, compared with strategies guided by clinical assessment alone (Table 2). The total trajectory costs per patient amounted to the following: €39,571 vs. €40,626 in Belgium; €36,002 vs. €37,237 in France; €43,483 vs. €45,361 in Germany; €41,582 vs. €44,130 in the Netherlands; £13,711 vs. £13,268 in the UK; and $104,400 vs. $107,269 in the US (Table 2).

Probabilistic analyses

The 70-gene signature-guided strategies were cost-effective compared with clinical assessment–guided strategies in all countries, given the WTP thresholds of €30,000/£30,000/$30,000 (Table 2). The probability that the 70-gene signature produced higher net benefit than clinical assessment alone using this threshold was 72% for Belgium, 75% for France, 79% for Germany, 85% for the Netherlands, 54% for the UK and 64% for the US (Fig. 2, Fig. S2).

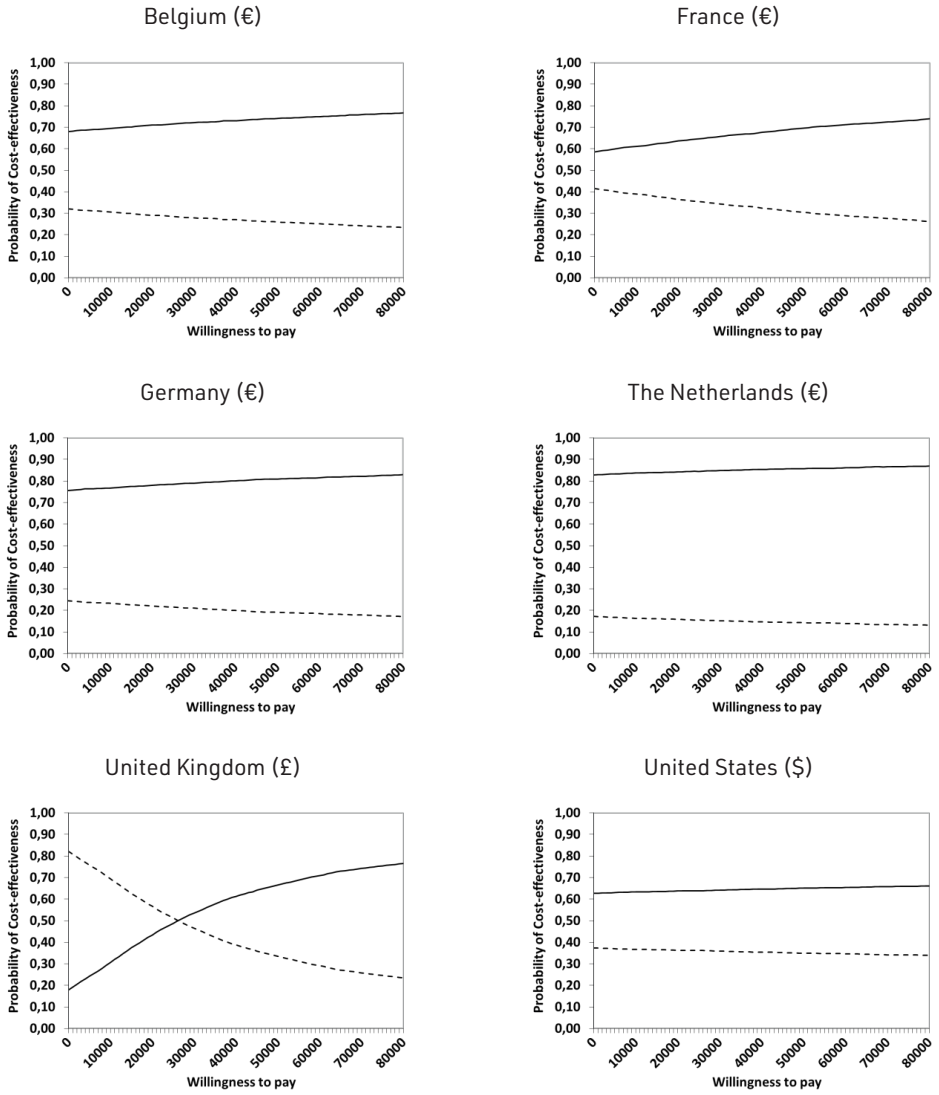


Figure 2. Cost-effectiveness acceptability curves. The solid line represents the probability of a treatment strategy guided by the 70-gene signature to be cost-effective (Y-axis), given a series of within-country willingness-to-pay thresholds as displayed on the X-axis. The dashed line represents the same but for a treatment strategy guided by clinical risk assessment only.

Sensitivity analyses

One-way sensitivity analyses did not affect the cost-effectiveness of the 70-gene signature-guided treatment strategies. Fig. 3 and S3 show the value of the test utilities, chemotherapy costs and 70-gene signature costs to be the biggest drivers

of cost-effectiveness. The different parametric distributions used to extrapolate outcomes to 10 years did not change the cost-effectiveness result. Application of country-specific population utility norms led to a lower incremental QALY for all countries, except for the UK and US which saw a higher incremental QALY (Table S8). Finally, the two-way adherence analysis which was tested on the German model revealed that when 99% of clinical high-risk cases receive chemotherapy, adherence to treatment strategies guided by the 70-gene signature among clinical high/genomic low cases should be at least 70% to remain cost-effective. If the proportion of chemotherapy given in clinical high-risk cases drops to 50%, then the adherence towards the 70-gene signature low-risk result should be at least 90% (Table S5).

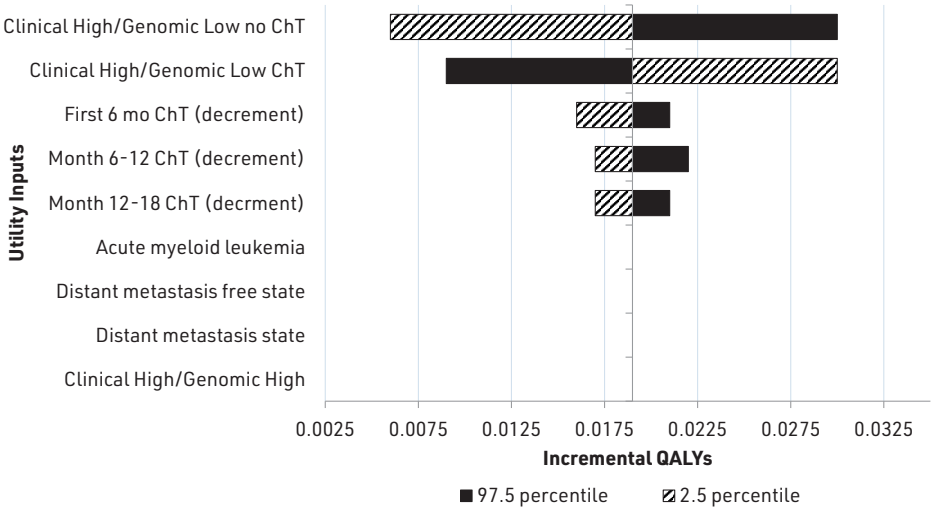


Figure 3. One-way sensitivity analysis of utility input effect on incremental quality-adjusted life years. The model for Germany is used as an example. The sensitivity of the incremental quality-adjusted life years gained under the treatment strategy guided by the 70-gene signature (X-axis) is tested by varying the utility inputs (displayed on the Y-axis) by their respective values in the 2.5th and 97.5th percentiles. ChT, chemotherapy; QALYs, quality-adjusted life years.

Budget impact based on costs per population

The 70-gene signature led to annual cost savings ranging from €4.2M in Belgium, €24.7M in France, €45.1M in Germany, €12.7M in the Netherlands and $244M in the US. A budget impact of £8.4M was seen for the UK. The variation in costs and cost savings can be attributed to the size of the target population in each country and differences in country-specific treatment guidelines and costs.

Discussion

Based on MINDACT data, for patients with ER+/HER2- tumours deemed to be clinically high risk in accordance with Adjuvant! Online, treatment strategies guided by the 70-gene signature saved costs in five of six countries, gained QALYs and were cost-effective in all six countries, given country-specific WTP thresholds. Several considerations should be made in interpreting the results of our analyses. The real-world use of and adherence to the 70-gene signature can differ widely from recommendations outlined in clinical guidelines. This has been highlighted in a range of publications.[7,25,26] Provided that a country closely follows guidelines for adjuvant chemotherapy use in clinical 'high' and 'intermediate'-risk patients, adopting the 70-gene signature and adhering to its recommendation to avoid chemotherapy in 'genomic-low' patients, this will prove to be cost-effective. To demonstrate this in a sensitivity analysis, we varied chemotherapy prescription rates (which may vary according to patient and provider preferences) and analysed the impact on cost-effectiveness. Another important consideration is chemotherapy-specific costs, which were found to be the biggest driver of incremental costs. In countries where chemotherapy costs are low (such as the UK), the 70-gene signature strategy might be cost-effective up to a WTP threshold of £30,000 but no longer dominates. The UK was the only country in our study where the 70-gene signature strategy was cost-effective but not cost-saving. In a study of the performance of the UK National Health Service compared with other high-income countries (including France, Germany, the Netherlands and the US), the UK demonstrated the lowest per capita healthcare spending.[27] This is certainly reflected in the comparatively lower cost inputs used in our UK model (Table 1). Although reliable and up-to-date inputs from published literature were difficult to obtain, subsequently introducing uncertainty into our model parameters and outcomes, it is obvious that in a health system with overall lower costs, potential gains will be accordingly lower.

Five-year median follow-up data was available from the MINDACT trial. Using a parametric modelling approach, we extrapolated this over a ten-year time horizon. This was important for two reasons: (1) to address regulator decision-making requirements and (2) to predict a more complete recurrence impact of the 70-gene signature-guided strategy. Although the 70-gene signature was designed to predict the chance of breast cancer recurrence within five years after surgery, our modelled population of clinical 'high-risk' patients typically experiences recurrence events within ten years. Furthermore, the risk-reducing effect of adjuvant chemotherapy occurs within the first five years. At 10 years and beyond, the absolute risk of breast cancer mortality for ER+ patients previously treated with endocrine therapy remains low. Therefore, there is low absolute benefit from chemotherapy in this population,

despite the continued response after 10 years. For ER+ patients, this effect cannot easily be parsed out from the patient's receipt of endocrine therapy.[5] It is for these reasons that a ten-year time horizon was used in our analysis, instead of the typical guideline-prescribed lifetime horizon, while also avoiding unnecessary introduction of uncertainty into the model. The standard errors applied in the later cycles of our model are larger, an indication of the uncertainty for the extrapolated period after 5 years.

The strength of this article lies in the use of patient-level data from the MINDACT RCT. This is the best available evidence to model recurrence and survival outcomes and their impact on cost-effectiveness with a high level of accuracy.[6,28] Test utilities directly measured from a sample of MINDACT patients were also integrated into our model.[9] This information reflects differences between subgroups immediately following receipt of personalised recurrence risk information.

Previous cost-effectiveness analyses found the same trend of the cost-effectiveness of the 70-gene signature compared with clinical risk assessments. Despite small differences in recurrence rates between strategies, HRQoL gains and cost savings were apparent.[29,30] This however is contested by other studies which found that the 70-gene signature was not cost-effective or that uncertainty was too high to draw a conclusion.[31-33]

The current analysis confirms findings of cost-effectiveness from earlier analyses but is set apart as the first cost-effectiveness modelling study of the 70-gene signature incorporating data stemming from a prospective RCT. With this data, we demonstrate how minimal expected survival differences in life years are offset by HRQoL gains, with patient outcomes proving to be similar in both strategies. Cost differences vary across countries; some countries see considerable cost savings per patient when using the 70-gene signature in guiding treatment decisions. In a patient-centred simulation, Caruana et al.[34] similarly demonstrate the significant deterioration of HRQoL due to chemotherapy side-effects for MINDACT patients. In both analyses, patients forgoing chemotherapy gain more QALYs. The minimal clinically important difference (MCID), i.e. the smallest benefit of value to patients, was not investigated and defined for the EQ-5D index-derived utilities used this study. It is possible that the observed gain in QALYs (an average of 0.02) may be smaller than a pre-defined MCID.[35,36]

Interpreting the results of the MINDACT trial calls for personalised decisions tailored to the individual patient. For women with ER+/HER2- tumours, the rate of breast cancer–specific survival without distant metastasis remains favourable.

However, observed recurrence differences within MINDACT may mean more to one patient than to another in the real-world, and Cardoso et al.[1] note that risk-benefit considerations must involve shared decision-making between physician and patient. Despite the seemingly small loss in life years which is counterbalanced by gains in quality of life, it is possible that QALYs do not capture the heterogeneity of preferences for patients regarding a chance of a loss in survival time. From a health systems perspective, our model provides information for country-wide policy decision-making. In this perspective, risk-benefit decisions must be weighed against 'average' (sub)groups of patients to decide if the 70-gene signature is suitable to bring into practice.

At the time of writing, the national health authorities of a number of the European countries studied in this analysis have not extended coverage over the 70-gene signature. These authorities have argued that, following the publication of 5-year median results of the MINDACT trial[1], uncertainty remains over the evidence of clinical utility using the 70-gene signature to de-escalate adjuvant chemotherapy.[8,37,38] Furthermore, cost and quality of life inputs required for cost-effectiveness modelling stem from outdated publications. Aside from the recurrence and survival outcomes drawn directly from the MINDACT trial, modelling the full impact of the 70-gene signature on quality of life and country-specific costs will continue to be riddled with uncertainty. These conclusions will likely remain until longer-term follow-up from the MINDACT trial is provided, accompanied by robust utility and cost evidence for this patient population. Future research into uniform cost data collection would be of value, for this study in particular related to chemotherapy costs, treatment-related adverse events and distant metastasis.

Conclusion

With the available evidence, these country-specific models demonstrated that adjuvant chemotherapy strategies guided by the 70-gene signature can save healthcare expenditures over ten years and offer a modest gain in quality-adjusted long-term survival. This information provides clinicians and policy makers with additional evidence of the clinical and economic value of the 70-gene signature for clinical high-risk patients with ER+/HER2- breast cancer in Europe and the US.

Table 1: Model input parameters

Parameters	Mean	SE	Distribution	Source		
Survival probabilities: See Supplementary Table S2 (DMFS), Supplementary Table S3 (OS)						
Probabilities long-term treatment-related adverse events						
Acute myeloid leukemia	0.00025	0.0001	Beta	Wolff et al. 2015[12]		
Congestive heart failure	0.037	0.001	Beta	Boekel et al. 2018[39]		
Utilities						
C-high/G-low/chemotherapy	0.828	0.036	Beta	MINDACT trial, Retel et al. 2013[9]		
C-high/G-low/no chemotherapy	0.838	0.039	Beta	MINDACT trial, Retel et al. 2013[9]		
C-high/G-high	0.832	0.021	Beta	MINDACT trial, Retel et al. 2013[9]		
Distant metastasis free state[1]	0.824	0.002	Beta	Lidgren et al. 2007[10]		
Distant metastasis state	0.685	0.004	Beta	Lidgren et al. 2007[10]		
Disutility chemotherapy (first 6 months)	-0.067	0.004	Beta	Campbell et al. 2011[11]		
Disutility chemotherapy (6 to 18 months)	-0.019	0.004	Beta	Campbell et al. 2011[11]		
Acute myeloid leukemia	0.260	0.040	Beta	Younis ea 2011[13]		
Costs[2]	BE (€)	FR (€)	DE (€)	NL (€)	UK (£)*	US ($)
MammaPrint®[3]	2,675	1,850	2,675	2,675	2,375	4,200
Endocrine therapy total[4]	1,150	550	1,440	1,194	284	459
Tamoxifen, AI, GnRH analogues	337	-	546	381	21	351
PB, calcium, vitamin D, DEXA scan	813	-	893	813	263	108
Chemotherapy total[3]	11,627	9,821	14,314	16,600	5,440	43,307
Chemotherapy	3,064	-	8,579	10,226	4,265	-
Chemotherapy administration	2,367	-	1,039	3,094	-	-
Anti-emetics	459	-	935	108	20	-
Prophylactic G-CSF	2,742	-	3,123	2,535	834	-
Short-term treatment-related AE	2,995	426	637	637	321	-
Monitoring/follow-up first year[5]	151	441	107	151	214	733
Monitoring/follow-up years 2-10[5]	87	441	53	87	120	733
Local/regional recurrence	18,359	18,359	18,359	18,359	15,164	21,659
Distant recurrence[5]	26,992	26,992	26,992	26,992	4,949	125,152
Acute myeloid leukemia[3]	31,259	31,259	31,259	31,259	28,468	35,644
Congestive heart failure[3]	3,710	3,710	3,710	3,710	3,378	7,458

[1]Country-specific population utility norms are applied in a sensitivity analysis (Supplementary Appendix Table S8)

[2]A Gamma distribution was used for costs in the probabilistic sensitivity analysis; details on country specific treatment utilization assumptions and references are listed in Supplementary Methods 2-3 and Supplementary Table 4.

[3]One-off costs.

[4]Per 6-month cycle, for 5 years. Extended tamoxifen up to 7 years use applied for 25% of patients in the DMFS state.

[5]Per 6-month cycle.

*Between 2017-2018, the British pound sterling (£) had an average annual exchange rate of 1.1359 to the Euro (€) according to the European Central Bank.

Abbreviations: AE: adverse events; G-CSF granulocyte-colony stimulating factor; DEXA scan: dual energy X-ray absorptiometry scan; AI: aromatase Inhibitors, PB: Prophylactic bisphosphonates, DMFS: Distant metastasis Free Survival: OS: overall survival

Table 2: Base-case results (Clinical high-risk, ER+/HER2- patients)

Country	Strategy	Costs (CI)	LYs (CI)	QALYs (CI)	Δ Costs	Δ QALYs*	ICER	Budget Impact**	Incremental Net Monetary Benefit
Belgium (€)	70-G	39,571 (30,984-51,049)	8.83 (8.77-8.88)	7.17 (7.11-7.23)	-1,055	0.018	70-G dominates	-4.2 M	2,458
	CA	40,626 (34,350-47,585)	8.85 (8.80-8.90)	7.15 (7.10-7.21)					
France (€)	70-G	36,353 (23,410-45,869)	7.83 (7.78-7.88)	6.36 (6.31-6.41)	-1,234	0.020	70-G dominates	-24.7 M	2,810
	CA	36,611 (31,758-43,245)	7.85 (7.80-7.89)	6.34 (6.29-6.38)					
Germany (€)	70-G	43,483 (34,659-55,191)	8.21 (8.16-8.26)	6.67 (6.61-6.71)	-1,878	0.019	70-G dominates	-45.1 M	3,390
	CA	45,361 (37,989-53,217)	8.23 (8.18-8.27)	6.65 (6.60-6.69)					
Netherlands (€)	70-G	41,582 (33,402-52,276)	8.83 (8.77-8.88)	7.17 (7.11-7.23)	-2,548	0.019	70-G dominates	-12.7 M	3,951
	CA	44,130 (37,269-51,512)	8.85 (8.80-8.90)	7.15 (7.10-7.20)					
UK (£)	70-G	13,711 (11,830-15,981)	8.02 (7.97-8.06)	6.51 (6.46-6.56)	+442	0.019	22,910 70-G more effective, more costly	+8.4 M	1,102
	CA	13,268 (11,441-15,113)	8.03 (7.98-8.08)	6.49 (6.44-6.54)					
US ($)	70-G	104,400 (69,754-154,475)	8.21 (8.16-8.26)	6.67 (6.61-6.71)	-2,869	0.019	70-G dominates	-244 M	4,381
	CA	107,269 (84,515-136,119)	8.23 (8.18-8.27)	6.65 (6.60-6.69)					

CA: clinical assessment strategy (based on the modified Adjuvant! Online); CI: credential intervals, representing the 2.5th and 97.5th percentiles of the 5000 Monte Carlo simulation iterations of the probabilistic analysis; ICER: incremental cost-effectiveness ratio; 70-G: 70-gene signature strategy; LYs: life years; M: million; QALYs: quality-adjusted life years; Δ: difference.

*QALYs rounded to the 3rd decimal point.

**Total costs for the 70-gene signature strategy were multiplied by the current country-specific annual incidence of eligible patients in the target group. For Belgium this amounted to 4,000/year,[21] for France 20,000/year,[22] for Germany 24,000/year,[20] for the NL 5,000/year,[23] for the UK 19,000 patients/year,[18] and for the US 85,000/year.[19]

70-G dominates CA; 70-G is more effective and less costly compared to CA.

Table 3: Patient characteristics

Summary of clinical-pathological characteristics of the Clinical High Risk, ER+, HER2- MINDACT patients (N=2297). Chemotherapy assignment according to randomization.

	Characteristic	Clinical High/ Genomic Low, ER+, HER2-, no chemotherapy (n=693) n(%)	Clinical High/ Genomic Low, ER+, HER2-, chemotherapy (n=709) n(%)	Clinical High/ Genomic High, ER+, HER2- (n=895) n(%)	All Clinical High, ER+, HER2- (N=2297) N(%)
Age (years)	<35 years	10 (1.4)	5 (0.7)	29 (3.2)	44 (1.9)
	35 to <50 years	222 (32.0)	239 (33.7)	310 (34.6)	771 (33.6)
	50 to 70 years	455 (65.6)	455 (64.2)	550 (61.5)	1460 (65.6)
	>70 years	6 (0.9)	10 (1.4)	6 (0.7)	22 (1.0)
Tumor size (cm)	≤1	18 (2.6)	18 (2.5)	13 (1.5)	49 (2.1)
	>1 to 2	265 (38.2)	274 (38.6)	401 (44.8)	940 (40.9)
	>2 to 5	381 (55.0)	390 (55.0)	468 (52.3)	1239 (53.9)
	>5	29 (4.2)	27 (3.8)	13 (1.5)	69 (3.0)
Lymph node status	Negative	352 (50.8)	364 (51.3)	593 (66.3)	1309 (57.0)
	Positive				
	N1	228 (32.9)	239 (33.7)	189 (21.1)	656 (28.6)
	N2	76 (11.0)	73 (10.3)	72 (8.0)	221 (9.6)
	N3	35 (5.0)	30 (4.2)	40 (4.5)	105 (4.6)
	N4+	2 (0.3)	2 (0.3)	1 (0.1)	5 (0.2)
	n.a.	-	1 (0.1)	-	1 (0.04)
Tumor grade	Grade 1	49 (7.0)	41 (5.8)	11 (1.2)	101 (4.3)
	Grade 2	454 (65.5)	461 (65.0)	292 (32.6)	1207 (52.5)
	Grade 3	184 (26.6)	200 (28.2)	589 (65.8)	973 (42.4)
	Undefined	6 (0.9)	7 (1.0)	3 (0.3)	16 (0.7)
Adjuvant treatment received[1]	ET only	597 (86.1)	96 (13.5)	41 (4.6)	734 (32.0)
	ET + ChT	75 (10.8)	576 (81.2)	810 (90.5)	1461 (63.6)
	ChT only	0 (0.0)	7 (1.0)	21 (2.3)	28 (1.2)
	No adjuvant treatment	10 (1.4)	8 (1.1)	4 (0.4)	22 (1.0)
	Missing	1 (0.1)	8 (1.1)	11 (1.2)	19 (0.8)
Country[2]	Belgium	91 (13.1)	91 (12.8)	99 (11.1)	281 (12.2)
	France	241 (34.8)	236 (33.3)	286 (32.0)	763 (33.2)
	Germany	106 (15.3)	109 (15.4)	111 (12.4)	326 (14.2)
	Netherlands	174 (25.1)	163 (23.0)	246 (27.5)	583 (25.4)
	United Kingdom	9 (1.3)	13 (1.8)	19 (2.1)	41 (1.8)
	Other[3]	72 (3.1)	97 (4.2)	134 (5.8)	303 (13.2)

[1]Actual adjuvant treatment received; n=21 patients have missing ChT treatment information; n=50 patients have missing ET treatment information.

[2]It is assumed that the distribution of clinical pathological characteristics are balanced within the country populations as a result of the randomization procedure.

[3]Other countries included Italy, Spain, Switzerland and Slovenia.

References

1. Cardoso F, van't Veer LJ, Bogaerts J, et al. 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. *N Engl J Med* 2016; **375**(8): 717-29.
2. Krop I, Ismaila N, Andre F, et al. Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline Focused Update. *J Clin Oncol* 2017; **35**(4): 2838-47
3. Senkus E, Kyriakides S, Ohno S, et al. Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2015; **26(**Suppl 5):v8-30.
4. Morigi C: Highlights from the 15th St Gallen International Breast Cancer Conference 15–18 March, 2017, Vienna: tailored treatments for patients with early breast cancer. *ecancermedicalscience* 2017; **11**: 732.
5. Peto R, Davies C, Godwin J, et al. Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. *Lancet* 2012; **379**(9814): 432-44.
6. Ramsey SD, Willke RJ, Glick H, et al. Cost-effectiveness analysis alongside clinical trials II-An ISPOR Good Research Practices Task Force report. *Value Health 2012*; **18**(2): 161-72.
7. Kuijer A, Straver M, den Dekker B, et al. Impact of 70-Gene Signature Use on Adjuvant Chemotherapy Decisions in Patients With Estrogen Receptor–Positive Early Breast Cancer: Results of a Prospective Cohort Study. *J Clin Oncol* 2017; **35**(24): 2814-9.
8. National Institute for Health and Care Excellence. Tumour profiling tests to guide adjuvant chemotherapy decisions in early breast cancer. Diagnostics guidance [DG34], 2018. https://www.nice.org.uk/guidance/dg34/ (accessed August 2018).
9. Retèl VP, Groothuis-Oudshoorn CGM, Aaronson NK, et al. Association between genomic recurrence risk and well-being among breast cancer patients. *BMC Cancer* 2013; **13**(1): 1-10.
10. Lidgren M, Wilking N, Jonsson B, et al. Health related quality of life in different states of breast cancer. *Qual Life Res* 2007; **16**(6): 1073-81.
11. Campbell HE, Epstein D, Bloomfield D, et al. The cost-effectiveness of adjuvant chemotherapy for early breast cancer: A comparison of no chemotherapy and first, second, and third generation regimens for patients with differing prognoses. *Eur J Cancer* 2011; **47**(17): 2517-30.
12. Wolff AC, Blackford AL, Visvanathan K, et al. Risk of marrow neoplasms after adjuvant breast cancer therapy: the national comprehensive cancer network experience. *J Clin Oncol* 2015; **33**(4): 340-8.
13. Younis T, Rayson D, Skedgel C: The cost-utility of adjuvant chemotherapy using docetaxel and cyclophosphamide compared with doxorubicin and cyclophosphamide in breast cancer. *Curr Oncol* 2011; **18**(6): 288-96.
14. Fenwick E, Claxton K, Sculpher M: Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Econ* 2001; **10**(8): 779-87.
15. Marseille E, Larson B, Kazi DS, et al. Thresholds for the cost-effectiveness of interventions: alternative approaches. Bulletin of the World Health Organization 2015 **93**:118-24.
16. Janssen MF, Szende A, Cabases J, et al. Population norms for the EQ-5D-3L: a cross-country analysis of population surveys for 20 countries. *Eur J Health Econ* 2019; **20**(2): 205-16.
17. Heins MJ, de Jong JD, Spronk I, et al. Adherence to cancer treatment guidelines: influence of general and cancer-specific guideline characteristics. *Eur J Public Health* 2017; **27**(4): 616-20.
18. Moller H, Henson K, Luchtenborg M, et al. Short-term breast cancer survival in relation to ethnicity, stage, grade and receptor status: national cohort study in England. *Br J Cancer* 2016; **115**(11): 1408-15.

19. Howlader N, Altekruse SF, Li CI, et al. US Incidence of Breast Cancer Subtypes Defined by Joint Hormone Receptor and HER2 Status. *J Natl Cancer Inst* 2014; **106**(5): dju055.
20. Eisemann N, Waldmann A, Katalinic A: Epidemiology of Breast Cancer – Current Figures and Trends. *Geburtshilfe und Frauenheilkunde* 2013; **73**: 130-5.
21. Belgian Cancer Registry. Cancer burden in Belgium 2004-2013. Brussels, 2015. https://kankerregister.org/media/docs/publications/BCR_publicatieCancerBurden2016_web160616.pdf (accessed August 2018).
22. Jéhannin-Ligier K DE, Bossard N, Molinié F, Defossez G, Daubisse-Marliac L, Delafosse P, Remontet L, Uhry Z: Projection de l'incidence et de la mortalité par cancer en France métropolitaine en 2017. Rapport technique. Saint-Maurice, Santé publique France, 2017, pp 80
23. Integraal Kankercentrum Nederland. Nederlandse Kankerregistratie. 2018. https://www.cijfersoverkanker.nl/ (accessed August 2018).
24. Vemer P, Corro Ramos I, van Voorn GA, et al. AdViSHE: A Validation-Assessment Tool of Health-Economic Models for Decision Makers and Model Users. *Pharmacoeconomics* 2016; **34**(4): 349-61.
25. Tsai M, Lo S, Audeh W, et al. Association of 70-Gene Signature Assay Findings With Physicians' Treatment Guidance for Patients With Early Breast Cancer Classified as Intermediate Risk by the 21-Gene Assay. *JAMA Oncol* 2018; **4**(1): e173470.
26. Cusumano PG, Generali D, Ciruelos E, et al. European inter-institutional impact study of MammaPrint. *Breast* 2014; **23**(4): 423-8.
27. Papanicolas I, Mossialos E, Gundersen A, et al. Performance of UK National Health Service compared with other high income countries: observational study. *BMJ* 2019; **367**: l6326.
28. Latimer NR: Survival analysis for economic evaluations alongside clinical trials--extrapolation with patient-level data: inconsistencies, limitations, and a practical guide. *Med Decis Making* 2013; **33**(6): 743-54.
29. Blok EJ, Bastiaannet E, van den Hout WB, et al. Systematic review of the clinical and economic value of gene expression profiles for invasive early breast cancer available in Europe. *Cancer Treat Rev* 2018; **62** :74-90.
30. Retel VP, Joore MA, Knauer M, et al. Cost-effectiveness of the 70-gene signature versus St. Gallen guidelines and Adjuvant Online for early breast cancer. *Eur J Cancer* 2010; **46**(8): 1382-91.
31. Bonastre J, Marguet S, Lueza B, et al. Cost effectiveness of molecular profiling for adjuvant decision making in patients with node-negative breast cancer. *J Clin Oncol* 2014; **32**(31): 3513-9.
32. Oestreicher N, Ramsey SD, Linden HM, et al. Gene expression profiling and breast cancer care: what are the potential benefits and policy implications? *Genet Med* 2005; **7**(6): 380-9.
33. Ward S, Scope A, Rafia R, et al. Gene expression profiling and expanded immunohistochemistry tests to guide the use of adjuvant chemotherapy in breast cancer management: a systematic review and cost-effectiveness analysis. *Health Technol Assess* 2013; **17**(44): 1-302.
34. Caruana E, Foucher Y, Tessier P, et al. Patient-centered simulations to assess the usefulness of the 70-gene signature for adjuvant chemotherapy administration in early-stage breast cancer. *Breast Cancer Res Treat* 2019; **174**(2): 537-42.
35. Coretti S, Ruggeri M, McNamee P: The minimum clinically important difference for EQ-5D index: a critical review. *Expert Rev Pharmacoecon Outcomes Res* 2014; **14**(2): 221-33.
36. McGlothlin AE, Lewis RJ: Minimal clinically important difference: defining what really matters to patients. *JAMA* 2014; **312**(13): 1342-3.

37. San Miguel LD, C.; Gerkens, S.; Harrison, J.; Hulstaert, F.: MammaPrint® test for personalised management of adjuvant chemotherapy decisions in early breast cancer: A rapid assessment, KCE Reports vol. D/2017/10.273/xx. Brussels, Health Technology Assessment (HTA) Brussels: Belgian Health Care Knowledge Centre (KCE), 2017
38. Haute Autorité de santé (HAS) French National Authority for Heath. INHATA Brief Issue 2019. ISBN number 978-2-11-152376-0. https://www.has-sante.fr/upload/docs/application/pdf/2019-03/inahta_brief_genomic_signatures_2019-03-22_14-57-13_608.pdf (accessed June 2020)
39. Boekel NB, Jacobse JN, Schaapveld M, et al. Cardiovascular disease incidence after internal mammary chain irradiation and anthracycline-based chemotherapy for breast cancer. *Br J Cancer* 2018; **119**(4): 408-18.

Supplementary materials

Supplementary Methods 1: Extrapolation of time-to-event endpoints using a Weibull distribution

Weibull hazard function: $h(t)=p\,\lambda\,t^{p-1}$
Weibull survival function: $S(t)=e^{-\lambda t\hat{}p}$

Parameter is the same for all groups but is different for each group.

DMFS outcome:
p=1.438
Group Clin high-Gen low-ACT: λ = 0.00012807
Group Clin high-Gen low-no ACT: λ = 0.00014862
Group Clin high-Gen high: λ = 0.00027045

OS outcome:
p=1.539
Group Clin high-Gen low-ACT: λ = 0.00003603
Group Clin high-Gen low-no ACT: λ = 0.00004692
Group Clin high-Gen high: λ = 0.00008395

Supplementary Methods 2: Treatment Assumptions

Treatment assumptions applied for all country populations are described below. Country-specific deviations from these treatment assumptions are described in detail in the Country Specific Assumptions and Sources section.

Chemotherapy

Chemotherapy treatment assumptions are unique to each country, and are based on the country's most commonly utilized regimens for early breast cancer patients. These are described in the Country Specific Assumptions and Sources section. Beyond chemotherapy drug acquisition, chemotherapy regimens captured administration and monitoring costs. This included port implantation, laboratory tests per cycle of chemotherapy, blood panels, human resources, echocardiogram, and outpatient stay costs.

Endocrine therapy

Based on the MINDACT trial population, it was assumed that 40% of patients were pre-menopausal, and 60% post-menopausal. Costs were calculated as follows: pre-menopausal patients were to receive Tamoxifen (20 mg QD) plus gonadotropin-releasing hormone (GnRH) analogues (3.6 mg implant every 28 days) for ovarian function suppression (OFS) for five years. Post-menopausal patients were to receive 2.5 years of Tamoxifen (20 mg QD), followed by 2.5 years of aromatase inhibitors (1 mg QD). We assumed that 25% of patients in the DMFS state received extended tamoxifen to 7 years. At present, the optimal duration of extended endocrine therapy in adjuvant settings is currently uncertain; we acknowledge that studies are going to determine appropriate treatment schedules up to 10 years of use.

Anti-emetics

Patients treated with an anthracycline combined with cyclophosphamide are assumed to have been offered a four-drug combination of an NK1 receptor antagonist (aprepitant 125 mg oral on day one, 80 mg oral on days 2–3), a 5-HT3 receptor antagonist (ondansetron 8 mg TID day one), dexamethasone (12 mg oral day one, 8 mg oral days 2–4), and olanzapine (10 mg oral day one, 10 mg oral days 2–4).

Prophylactic G-CSF

Updated guidelines have concluded that patients with >20% risk for febrile neutropenia should be offered primary granulocyte-colony stimulating factor (G-CSF). As such pegfilgrastim, a pegylated formulation of G-CSF was assumed to be used prophylactically alongside anthracycline-based chemotherapy regimens. It is now known that continued use of primary G-CSF prophylaxis during

all chemotherapy cycles is warranted.[1] Therefore, it was assumed that patients undergoing anthracycline-based chemotherapy would receive G-CSF prophylaxis for an average of 4 cycles.

Bisphosphonates and screening for fracture risk

Prophylactic bisphosphonates are used for prevention of bone mineral density loss resulting from systemic therapy (e.g. endocrine therapies and GnRH analogues for OFS). A consensus guideline from a European Panel recommends the use of zoledronic acid 4 mg intravenously every six months or clodronate 1600 mg orally daily for three to five years in premenopausal women on adjuvant ovarian suppression and postmenopausal women at intermediate or high risk of recurrence.[2] Additionally, daily intake of calcium (1000 mg) and vitamin D (800 international units) is recommended. Therefore, we assume pre-menopausal patients receive Zolendronic acid (4 mg IV every 6 months for five years) in addition to calcium and vitamin D supplementation. Post-menopausal patients receive clodronate (1600 mg oral daily for five years) plus calcium and vitamin D supplementation. As this population is at high fracture risk, they are assumed to undergo additional screening using physical exams, the fracture risk assessment tool (FRAX), and, as needed, a dual-energy x-ray absorptiometry (DXA) for evaluation of BMD.[2]

Short-term treatment-related adverse events

Probability of experiencing grade III and IV treatment-related adverse events requiring hospitalization were considered. This included neutropenic infection, thrombocytopenia, anemia, nausea/vomiting, emesis, diarrhea, stomatitis, thrombosis, neuropathy, pain, myalgia, and infection.

Long-term treatment-related adverse events

Acute myeloid leukemia (AML) was considered a possible long-term chemotherapy-related adverse event. The cumulative probability of AML provided in the Wolff et al. publication was used to calculate the yearly rate of AML over 10 years.[3] AML costs were applied "one off" and were derived from Wang et al.[4]

Congestive heart failure (CHF) was also considered in the model at the final cycle. A cumulative incidence of 0,037 was applied in the final cycle of the progression free state,[5] as well as one-off costs for treatment of CHF derived from Biermann et al.[6]

Monitoring and Follow-up Care

For the first year of follow-up, patients were assumed to have received two outpatient consultations and one mammography. For subsequent years (years 2 to 10), patients had one annual outpatient consultation and mammography.

Local Regional Recurrence

It was assumed that 3.7% of patients who entered the distant metastasis state had also experienced one loco-regional recurrence (LRR) beforehand, for which they received the same best available treatment, independent of the kind of adjuvant treatment originally received for the primary tumor. This assumption was based upon the proportion of MINDACT patients in the target group of clinical high, ER+/HER2- patients who were known to experience LRR before distant metastasis during the observed 5-year median follow-up period. In support of this assumption, it has also been demonstrated that Luminal A (ER+/HER2-) patients have the lowest rate of LRR compared to other molecular subtypes.[7,8] Costs for LRR were adapted from the de novo economic model reported in the 2013 HTA report by Ward et al.[9] which derives costs associated with LRR from Karnon et al.[10] Among patients who did not experience distant metastasis in the target group, the cumulative incidence of LRR was 1.2%. Costs for this small group of patients were not accounted for in the model.

Distant Recurrence and Palliative care

The treatment pathway for distant recurrence included diagnostic tests to confirm distant recurrence, drug acquisition costs, disease management, and end of life care. The treatment pathway considered adjuvant treatments which are standard of care for metastatic breast cancer in women with ER+/HER2- tumors. Applications of costs and treatment assumptions for the European countries (Belgium, France, Germany, and the Netherlands) were the same, based upon the chart review published by Jerusalem et al. 2015.[11] UK costs for recurrence and end of life care were based upon those reported by Thomas et al,[12] inflated to 2015/16 values using the Hospital and Community Health Services (HCHS) pay and prices inflation index. US costs were based on the publication of Montero et al. 2012[13]

Supplementary Methods 3: Country-Specific Treatment Assumptions and Sources

Belgium:

Healthcare utilization and costs were taken from the Belgian Health Care Knowledge Centre rapid assessment report, "MammaPrint® test for personalized management of adjuvant chemotherapy decisions in early breast cancer".[14] Chemotherapy regimens are those which are most commonly used in Belgium in the MammaPrint target population. The cost of chemotherapy used in this analysis is weighted by the proportions of patients receiving each chemotherapy regimen, as well as other relevant costs including catheter, administrative costs, and clinical biology. The two most frequently used chemotherapy were EC→Paclitaxel in 42.7% of target population patients (four cycles cyclophosphamide and epirubicin, followed by 12 cycles of Paclitaxel) and FEC→docetaxel in 20.1% (three cycles of Cyclophosphamide, epirubicin and flourouracil, followed by a further three of Docetaxel). The KCE report also considered the costs of prophylaxis and management of common chemotherapy-related adverse events, including prophylaxis anti-emetics, neutropenia prophylaxis, management of neutropenia, and hospitalization costs for non-neutropenia AEs. Endocrine therapy costs were not included in the KCE report so were subsequently collected from Belgium-specific cost-effectiveness analyses of Tamoxifen and Aromatase Inhibitors.[15]

France:

The majority of costs for France were provided by the investigators of Optisoins01, a French, multicenter, prospective study, and later checked following their publication in 2019.[16,17] Optisoins01 was an observational study conducted in France in 2014-2016 with early-stage breast cancer patients. The evaluation reflected the main care pathway of early breast cancer patients treated with initial surgery from hospital and health-insurance perspectives. Cost data were collected from patients prior to surgery, after surgery, during and after adjuvant therapy, in- and outpatient care till 1 year follow-up. With regards to G-CSF use, it was assumed that 15% of patients receiving chemotherapy were treated prophylactically with G-CSF as was observed in the Optisoins01 study. Costs were drawn from Tilleul et al. which noted that of the breast cancer patients that receive G-CSF: most receive Pegfilgrastim (66%), 33% Lenograstim and 17% Filgrastim. The total exceeds 100% because patients can receive multiple G-CSF.[18]

Germany:

For Germany, cost and utilization data based on a publication of Lux et al.[19] was used. Costs associated with chemotherapy were based on the most commonly utilized regimen in Germany. This is an anthracycline- and taxane-containing regimen of four cycles of epirubicin (90mg/^{2} BSA, d1, q21d) and cyclophosphamide (600 mg/m^2 BSA, d1, q21d), followed by 12 cycles of paclitaxel (80 mg/m^2 BSA, d1, q7d). Hospital fees and consultation fees for two quarterly visits over a 6 month treatment period were also included. Concomitant and supportive therapies included uromitexan, ondansetron, dexamethasone, aprepitant, clemastine, ranitidine, and pegfilgrastim. Costs for treatment of short-term Grade III/IV treatment-related toxicities were also captured. Adjuvant endocrine therapy for a standard initial duration of five years was considered according to the German Arbeitsgemeinschaft Gynäkologische Onkologie (AGO) 2018 guidelines.[20] All unit costs were sourced from the German national formulary Rote Liste® for endocrine therapy, chemotherapy and concomitant medications.[21] Calculations for unit costs of diagnostic tests, side effects, treatment administration and monitoring were adapted from the the Lux paper, which used a price year of 2016. Costs were inflated to price year 2017/18 for inclusion into the model.

The Netherlands:

For the Netherlands, cost data from national databases such as the handbook for costing studies in the Netherlands from the National Healthcare Insurance board (ZorgInstituut Nederland) and the online pharmacotherapeutic cost database www.medicijnkosten.nl were used for medications.[22,23] Chemotherapy regimens consisted of paclitaxel (80mg/m^2), docetaxel (75mg/m2), doxorubicin (50mg/m^2), cyclophosphamide (500mg/m^2), used in three different regimens. Three cycles of fluorouracil-epirubicin-cyclophosphamide followed by three cycles of docetaxel (T) is the most used regimen (50% according to clinical experts).

Another used regimen is doxorubicin-cyclophosphamide (AC) once every 3 weeks for four cycles, followed by docetaxel once every 3 weeks for four cycles (used in 25% of cases).

And doxorubicin-cyclophosphamide (AC) once every 3 weeks for four cycles, followed by paclitaxel once weekly for twelve cycles (used in 25% of cases).

On the basis of both European Union recommendations[24] and US National Comprehensive Cancer Network guidelines,[25] preferable anthracycline based chemotherapy regimens include subcutaneous Pegfilgrastim for four to six cycles.

United Kingdom:

For the United Kingdom, costs and treatment assumptions were adapted from the de novo economic model reported in the 2013 HTA report by Ward et al[9] and from the value of information analysis based upon the OPTIMA prelim trial reported by Hall et al. 2017.[26]

For endocrine therapy (ET), five regimens were considered: Tamoxifen for five years, Anastrozole for five years, Letrozole for five years, Tamoxifen for two years plus Exemestane for the final three years, Tamoxifen (or other ET regimes) for five years. It is assumed that patients receive one of four adjuvant chemotherapy regimens: (1) FEC 100-T (3+3 cycles, assumed to be given to 25% of patients); (2) TC (4 cycles, assumed to be given to 20% of patients); (3) FEC75 (6 cycles, assumed to be given to 45% patients) and FEC100-Pw (3+3 cycles, assumed to be given to 10% patients). A weighted mean cost of chemotherapy acquisition, delivery and toxicity is derived from this. It was assumed that 25% of patients treated with chemotherapy receive Filgrastim as prophylactic G-CSF (maximum 3 cycles) for the secondary prevention of febrile neutropenia. All costs were inflated to a price level of 2017/18.

United States:

Current National Comprehensive Cancer Network guidelines list 12 acceptable regimens for ER+/HER2- breast cancer and costs of chemotherapy display significant variability.[25] For US chemotherapy regimens, estimated regimen costs were provided through Blue Cross Blue Shield (California).[27] Healthcare Common Procedure Coding System codes were used to define patients' chemotherapy regimens.[28] Docetaxel & cyclophosphamide (TC) and dose-dense doxorubicin & cyclophosphamide + paclitaxel (ddAC+P) represent the most commonly used regimens with 46% and 12% of patients treated as such, respectively. Healthcare Common Procedure Coding System codes were used to capture endocrine therapy costs for Blue Shield Blue Cross. Xie et al.[29] was an additional source for endocrine therapy costs, and were obtained using the wholesale acquisition cost. In case of a drug with multiple package sizes, the price of the largest package was used. The intended dose consumed was estimated based on Food and Drug Administration drug labels.

Questionnaire used for cost parameters (France and United States)

Cost inputs were collected through a questionnaire distributed to Prof.dr. Roman Rouzier, Institut Curie, Paris, France and Blue Cross Blue Shield, California, US. Costs for France were later updated according to the publication of Hequet et al. 2019.[17]

Costs (euros/dollars)			
	Unit costs	**Mean costs**	**Source**
MammaPrint list price			
Endocrine therapy List proportion of patients receiving each regimen			
Tamoxifen (Nolvadex®) 20 mg tablets			
Anastrozole (Arimidex®) 1 mg tablets			
Goserelin (Zoladex®) 3.6 mg implant			
Chemotherapy total (incl:)*			
Chemotherapy *List proportion of patients receiving each regimen* I.e. Combination of: Paclitaxel…(25%) FEC 50/500 (50%) Etc.			
Chemotherapy administration and monitoring costs (e.g. port implantation, blood panels, etc.)			
Anti-emetics/nausea			
Aprepitant (Emend®) 125 mg oral			
Ondansetron 8 mg			
dexamethasone			
olanzapine			
Prophylactic GCSF (Pegfilgrastim, Neulasta) List proportion of patients receiving this.			
Bisphosphonates and monitoring			
zoledronic acid 4 mg (IV)			
clodronate 1600 mg (oral)			
DEXA scan			
Treatment-related grade III and IV adverse events requiring hospitalization (neutropenic infection, thrombocytopenia, anemia, nausea/vomiting, emesis, diarrhea, stomatitis, thrombosis, neuropathy, pain, myalgia, and infection)			
Trastuzumab			
Monitoring/follow-up first year Mammography Outpatient physician visit			

Costs (euros/dollars)			
Monitoring/follow-up subsequent years Mammography Outpatient physician visit			
Recurrence (distant +loc/reg) Treatment local/regional Treatment distant			
End of life care			
Congestive heart failure			

Supplementary Methods 4: Test-Utility measurements alongside the MINDACT trial

Utilities regarding receiving two risk profiles (clinical and genomic) were measured by means of the EQ-5D-3L amongst the first 800 enrolled patients in the MINDACT study, of which a total of n=347 were included in the QoL study.[30] The EQ-5D-3L descriptive system comprises the following five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has three levels: no problems, some problems, and extreme problems. From the following EQ-5D health states, as defined by the EQ-5D descriptive system, a single summary index was obtained by applying a formula that attaches values to each of the levels in the five dimensions. Subsequently, the general population-based value set of the Netherlands was used (because it concerned Dutch patients) to reflect the preferences of local taxpayers and potential receivers of healthcare in that specific country.[31]

These particular utility values capture patient well-being specific to receiving the results of their clinical and gene expression profile. This utility information reflects the real differences between subgroups immediately following the receipt of their personalized recurrence risk information and the accompanying advice surrounding systemic treatment. These test utility values do not reflect the actual experience of chemotherapy and the chemotherapy-related adverse events. We use these "test-utilities" as an alternative baseline utility value at the start of the 6 month cycle. As such, we must also apply the Campbell et al.[32] decrement for chemotherapy-related adverse events for the first 6 months.

We assume that these differences between subgroups are important to capture in the first 6 months, but that patients in the disease free state have returned to the values reported by Lidgren et al.[33] by the start of the 12 month cycle. This is indeed a unique approach to cost-utility analyses, which we believe provides added value and detailed information in an analysis exploring the use of genomic profiles in oncology. Patient quality of life is impacted throughout various stages of their disease course, and we provide an additional step here in capturing this. These "test-utility" values do not and should not be required to capture chemotherapy-related adverse events as well. Therefore we must also apply the Campbell et al. decrement.[32]

	Test Utilities				
	Mean	**95% Confidence interval**	**N**	**SE**	**Source**
C-low/G-low	0.853	0.823-0.884	109	0.015	Retel et al. 2013[30]
C-low/G-high/no chemotherapy	0.873	0.795-0.951	70	0.035	
C-low/G-high/chemotherapy	0.872	0.826-0.918	12	0.022	
C-high/G-low/chemotherapy*	0.828	0.755-0.902	17	0.036	
C-high/G-low/no chemotherapy*	0.838	0.757-0.920	25	0.039	
C-high/G-high*	0.832	0.790-0.874	25	0.021	
C-low/G-no**	0.795	0.732-0.859	33		
C-high/G-no**	0.766	0.711-0.820	56		

*Used in the current model.

**No genomic profile possible or performed.

Supplementary Methods 5: Validation of the model

We validated the model by means of the AdViSHE model, Assessment of the validation status of Health-Economic decision models.[34]

A1/ Face validity testing (conceptual model)
The model has been judged by three experts: one associate professor health economics from the University of Twente, the Netherlands, one Affiliate Associate Professor, Fred Hutchinson Cancer Research Center, University of Washington, and one consultant.
Justification for considering these experts were that they are leading experts in the field, and have ample experience is judging health economic models. All experts agreed with the model.

A2/ Cross validity testing (conceptual model)
The model has been compared to other models found in the literature. The model in fact is based on former models on the cost-effectiveness of the MammaPrint, all based on the advanced modeling course material from the University of York, UK and University of Maastricht, the Netherlands.

B1/ Face validity testing (input data):
Several clinicians and clinical geneticists has been involved in judging the appropriateness of the input parameters. Medical oncologists, surgical oncologists from several high standard, European cancer institutes.

B2/ Model fit testing:
Survival and HRQoL data were based on the MINDACT trial. Extrapolation by means of parametric survival distributions was performed by one of the co-authors, a trained oncology statistician (KJ).

C1/ External review:
The model has been examined by the above mentioned (A1) experts. With one of the experts, all formulas of the Markov model were checked, the adherence and the conditional survival formulas.

C2/ Extreme value testing:
The model has been run for extreme sets of parameter values; e.g. treatment costs, test costs, survival probabilities, utilities.

C3/ Testing of traces:
Due to the close involvement of many clinicians and other experts in the field, the structure of the model has been tested many times. The choice of the structure was based on a true reflection of the real world (we adjusted the randomization structure, as this is not how the MammaPrint will be used in clinical practice)

C4/ Unit testing:
There was a project sheet defined beforehand, this was checked by the experts. The most important submodules of the model were checked by the experts; e.g. adherence, formulas and macros.

D1/ Face validity testing (model outcomes):
The involved clinicians and clinical/molecular geneticists have been asked to judge the outcomes of the model. We have had several rounds of feedback and discussion, using an iterative approach. All outcomes were checked separately, the total costs, (quality adjusted) life years, ICER and budget impact.

D2/ Cross validation testing (model outcomes):
The model outcomes have been compared to the outcomes of other models that address similar problems. The utility data is in general comparable with other publications, however there is limited evidence on the HRQoL with versus without chemotherapy. Especially the cost data and outcomes for the several countries is variable. The ranges are high within countries, however the trends of high or low costs in specific countries are present: the US presents often with the highest costs, followed by the Netherlands and Germany, France, Belgium and the UK. The latter tends to have the lowest costs for the total treatment pathways.

D3/ Validation against outcomes using alternative input data:
The model outcomes have been compared to the outcomes obtained when using alternative input data, e.g. for other utility data and cost data.

D4/ Validation against empirical data:
The model includes already empirical data. It has been validated with former model based data.

E1/ Other validation techniques:
Other validation technique which was applied was naïve benchmarking ("back-of-the-envelope" calculations), to check if the results were reasonable.

Supplementary Methods 6: Scenario analysis using DFS

In a scenario analysis, Disease Free Survival (DFS) was used as an alternative health state to DMFS (data not shown). Within this endpoint, progression captures events such as loco-regional recurrences, contralateral invasive breast cancer, invasive second (non-breast) primary cancer, and ductal carcinoma (in addition to distant metastasis and death).

In our base-case model we chose the time-to-event endpoint of DMFS to create our health states. As the 70-gene signature is used to guide adjuvant chemotherapy treatment decisions, DMFS is the most relevant endpoint because we are solely concerned with the effect of chemotherapy on two major, related events: distant metastases and death. There is currently debate on which trial endpoint is most relevant, with some arguing in favor of DFS.[35,36] We maintain the argument that DFS includes many events that are not relevant to measuring the beneficial effect of adjuvant chemotherapy on overall survival: e.g. local recurrence, DCIS, and secondary cancers.

Using DFS as a health state did not change the conclusion in four out of six countries. For France and the UK, the incremental QALYs remained positive, however the balance between costs (due to more events) and effects was less favorable for the 70-gene signature strategy.

Supplementary Table S1: Country-specific annual discount rates

	Belgium[37]	Germany[38]	France[39]	The Netherlands[22]	United Kingdom[40]	United States
Costs discount rate	0.030	0.030	0.040	0.040	0.035	0.030
Outcomes (Life Years) discount rate*	0.015	0.030	0.040	0.015	0.035	0.030

(Bi-annual rates are applied for each 6-month cycle in the model).

Supplementary Table S2: Conditional probabilities: Distant Metastasis Free Survival (ITT), Weibull distribution

Survival probabilities (DMFS) according to clinical/genomic risk groups and chemotherapy allocation. Intention-to-treat (ITT) population used. Weibull distribution applied.

	Clinical High/Genomic Low, ER+, HER2-, no chemotherapy (n=693)		Clinical High/Genomic Low, ER+, HER2-, chemotherapy (n=709)		Clinical High/Genomic High, ER+, HER2- (n=895)	
Month	**Deterministic P(Surv|year)**	**Standard Error**	**Deterministic P(Surv|year)**	**Standard Error**	**Deterministic P(Surv|year)**	**Standard Error**
0	1.0000	-	1.0000	-	1.0000	-
6	0.998145	0.000932	0.998401	0.000849	0.996625	0.001703
12	0.996753	0.001008	0.9972	0.000951	0.994095	0.001717
18	0.995881	0.00103	0.996449	0.000989	0.992513	0.001591
24	0.995189	0.001089	0.995852	0.001049	0.991258	0.001511
30	0.994599	0.001196	0.995343	0.001139	0.990188	0.001525
36	0.994077	0.001345	0.994892	0.001257	0.989241	0.001647
42	0.993605	0.001526	0.994485	0.001401	0.988385	0.001863
48	0.99317	0.001732	0.99411	0.001564	0.987599	0.002152
54	0.992766	0.001956	0.993761	0.001742	0.986868	0.00249
60	0.992388	0.002193	0.993435	0.001933	0.986182	0.002863
66	0.99203	0.002441	0.993126	0.002134	0.985535	0.00326
72	0.991691	0.002697	0.992833	0.002342	0.984921	0.003675
78	0.991367	0.00296	0.992553	0.002557	0.984336	0.004103
84	0.991057	0.003228	0.992286	0.002777	0.983776	0.004542
90	0.990759	0.0035	0.992029	0.003002	0.983237	0.004989
96	0.990472	0.003777	0.991782	0.003231	0.982719	0.005443
102	0.990195	0.004057	0.991543	0.003464	0.982219	0.005903
108	0.989928	0.00434	0.991311	0.003699	0.981736	0.006368
114	0.989668	0.004625	0.991087	0.003938	0.981267	0.006837
120	0.989416	0.004914	0.99087	0.004179	0.980812	0.00731

Supplementary Table S3: Conditional probabilities: Overall Survival (ITT), Weibull distribution

Survival probabilities (OS) according to clinical/genomic risk groups and chemotherapy allocation. Intention-to-treat (ITT) population used. Weibull distribution applied.

	Clinical High/Genomic Low, ER+, HER2-, no chemotherapy (n=693)		Clinical High/Genomic Low, ER+, HER2-, chemotherapy (n=709)		Clinical High/Genomic High, ER+, HER2- (n=895)	
Month	**Deterministic P(Surv|year)**	**Standard Error**	**Deterministic P(Surv|year)**	**Standard Error**	**Deterministic P(Surv|year)**	**Standard Error**
0	1.0000	-	1.0000	-	1.0000	-
6	0.999482	0.000424	0.999601	0.000378	0.999069691	0.00084519
12	0.99884	0.000558	0.999107	0.000511	0.997917205	0.001039264
18	0.998341	0.00062	0.998723	0.00058	0.997021588	0.001072142
24	0.997902	0.000682	0.998385	0.00064	0.996234367	0.00106972
30	0.997501	0.000772	0.998076	0.000709	0.995514888	0.001080673
36	0.997127	0.000903	0.997788	0.000798	0.994843516	0.001147248
42	0.996773	0.001083	0.997515	0.000914	0.994209044	0.001300636
48	0.996435	0.00131	0.997255	0.001059	0.993604129	0.001551688
54	0.996111	0.001585	0.997006	0.001235	0.993023718	0.001895067
60	0.995799	0.001903	0.996765	0.001441	0.992464082	0.002320316
66	0.995496	0.002263	0.996532	0.001678	0.991922446	0.002818306
72	0.995202	0.002663	0.996306	0.001943	0.991396628	0.003382595
78	0.994916	0.003104	0.996085	0.002237	0.990884724	0.004008963
84	0.994637	0.003583	0.99587	0.002559	0.990385486	0.004694697
90	0.994364	0.004101	0.99566	0.002909	0.989897572	0.005438035
96	0.994097	0.004657	0.995454	0.003285	0.989420044	0.006237805
102	0.993836	0.005252	0.995253	0.003689	0.988951992	0.007093197
108	0.993579	0.005886	0.995055	0.004119	0.988492792	0.008003596
114	0.993326	0.006559	0.99486	0.004576	0.988041645	0.008968496
120	0.993078	0.00727	0.994669	0.005059	0.987598038	0.009987423

Supplementary Table S4: Country-specific costs and sources

Country-specific costs (per 6 month cycle) and sources, inflated to 2017/18 price year (UK values 2015/16)

	Belgium (€)		France (€)		Germany (€)		Netherlands (€)		United Kingdom (£)		United States ($)	
	Mean	Source	Mean	Source	Mean	Source	Mean	Source	Mean	Source	Mean	Source
MammaPrint*	2,675	[41]	1,850	[46]	2,675	[41]	2,675	[41]	2,375	[41]	4,200	[41]
Endocrine therapy total	1,150		550	[16,17]	1,149		1,194		284		459	
Tamoxifen, aromatase inhibitors, GnRH analogues	337	[15]	-		547	[21]	381	[23]	21	[9]	351	[27,29]
Prophylactic bisphosphonates, calcium, vitamin D, DXA scan	813	[23]	-		893	[21]	813	[23]	263	[9]	108	[42]
Chemotherapy total*	11,627		9,821	[16,17]	14,314		16,600		5,440		43,307	[27]
Chemotherapy	3,064	[14]	-		8,579	[19]	10,226	[23]	4,265	[26]	-	
Chemotherapy administration	2,367	[14]	-		1,039	[19]	3,094	[23]	-		-	
Anti-emetics	459	[14]	-		935	[19]	108	[23]	20	[26]	-	
Prophylactic G-CSF	2,742	[14]	-	[18]	3,123	[19]	2,536	[23]	834	[26]	-	
Short-term treatment-related adverse events	2,995	[14]	426		637	[19]	637	[19]	321	[26]	-	
Monitoring/follow-up first year	151	[43]	441	[16,17]	107	[19]	151	[43]	214	[9]	733	[27]
Monitoring/follow-up years 2-10	87	[43]	441	[16,17]	53	[19]	87	[43]	120	[9]	733	[27]
Local/regional recurrence**	18,359	[10]	18,359	[10]	18,359	[10]	18,359	[10]	15,164	[10]	21,659	[10]
Distant recurrence & palliative care	26,992	[11]	26,992	[11]	26,992	[11]	26,992	[11]	4,949	[12]	125,152	[13]
Acute myeloid leukemia**	31,259	[4]	31,259	[4]	31,259	[4]	31,259	[4]	28,468	[4]	35,644	[4]
Congestive heart failure**	3,710	[6]	3,710	[6]	3,710	[6]	3,710	[6]	3,378	[6]	7,458	[44]

G-CSF: granulocyte-colony stimulating factor, GnRH: gonadotropin-releasing hormone.

*Applied one-off.

**Applied only to proportion of patients.

Supplementary Table S5: Two-way sensitivity analysis (based on German model)

Table S5(a): Impact of adherence to MammaPrint "low risk" and clinical assessment on ICER values

		Proportion of Clinical "High Risk" patients receiving chemotherapy										
		0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Proportion of patients adhering to MammaPrint "Low Risk" result	100%	-	€ 314,490	€ 62,183	€ -21,920	€ -63,971	€ -89,202	€ -106,022	€ -118,037	€ -127,048	€ -134,056	€ -139,663
	90%	€ -992,952	-	€ 612,703	€ 211,289	€ 77,485	€ 10,582	€ -29,559	€ -56,320*	€ -75,435	€ -89,771	€ -100,921
	80%	€ -740,645	€ -1,291,166	-	€ 910,917	€ 360,396	€ 176,889	€ 85,136	€ 30,084	€ -6,618	€ -32,833	€ -52,494
	70%	€ -656,543	€ -889,752	€ -1,589,380	-	€ 1,209,131	€ 509,503	€ 276,294	€ 159,689	€ 89,727	€ 43,085	€ 9,769
	60%	€ -614,492	€ -755,948	€ -1,038,859	€ -1,887,594	-	€ 1,507,345	€ 658,610	€ 375,699	€ 234,243	€ 149,369	€ 92,787
	50%	€ -589,261	€ -689,045	€ -855,352	€ -1,187,966	€ -2,185,808	-	€ 1,805,558	€ 807,717	€ 475,103	€ 308,796	€ 209,012
	40%	€ -572,441	€ -648,904	€ -763,599	€ -954,757	€ -1,337,073	€ -2,484,021	-	€ 2,103,772	€ 956,824	€ 574,508	€ 383,350
	30%	€ -560,426	€ -622,143	€ -708,547	€ -838,152	€ -1,054,161	€ -1,486,180	€ -2,782,235	-	€ 2,401,986	€ 1,105,931	€ 673,912
	20%	€ -551,415	€ -603,028	€ -671,845	€ -768,189	€ -912,706	€ -1,153,566	€ -1,635,287	€ -3,080,449	-	€ 2,700,200	€ 1,255,038
	10%	€ -544,407	€ -588,692	€ -645,630	€ -721,548	€ -827,832	€ -987,259	€ -1,252,971	€ -1,784,394	€ -3,378,663	-	€ 2,998,414
	0%	€ -538,800	€ -577,541	€ -625,969	€ -688,232	€ -771,250	€ -887,475	€ -1,061,813	€ -1,352,375	€ -1,933,501	€ -3,676,876	-

White: cost-effective; grey: not cost-effective (with a willingness-to-pay threshold of 30,000 EUR); *base-case value.

Table S5(b): Impact of adherence to MammaPrint "low risk" and clinical assessment on incremental costs

		Proportion of Clinical "High Risk" patients receiving chemotherapy										
		0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Proportion of patients adhering to MammaPrint "Low Risk" result	100%	€ 1,325	€ 826	€ 326	€ -173	€ -672	€ -1,171	€ -1,670	€ -2,169	€ -2,668	€ -3,167	€ -3,666
	90%	€ 2,607	€ 2,108	€ 1,608	€ 1,109	€ 610	€ 111	€ -388	€ -887	€ -1,386	€ -1,885	€ -2,384
	80%	€ 3,889	€ 3,390	€ 2,890	€ 2,391	€ 1,892	€ 1,393	€ 894	€ 395	€ -104	€ -603	€ -1,102
	70%	€ 5,171	€ 4,672	€ 4,172	€ 3,673	€ 3,174	€ 2,675	€ 2,176	€ 1,677	€ 1,178	€ 679	€ 180
	60%	€ 6,453	€ 5,954	€ 5,454	€ 4,955	€ 4,456	€ 3,957	€ 3,458	€ 2,959	€ 2,460	€ 1,961	€ 1,462
	50%	€ 7,735	€ 7,236	€ 6,736	€ 6,237	€ 5,738	€ 5,239	€ 4,740	€ 4,241	€ 3,742	€ 3,243	€ 2,744
	40%	€ 9,017	€ 8,518	€ 8,018	€ 7,519	€ 7,020	€ 6,521	€ 6,022	€ 5,523	€ 5,024	€ 4,525	€ 4,026
	30%	€ 10,299	€ 9,800	€ 9,300	€ 8,801	€ 8,302	€ 7,803	€ 7,304	€ 6,805	€ 6,306	€ 5,807	€ 5,308
	20%	€ 11,581	€ 11,082	€ 10,582	€ 10,083	€ 9,584	€ 9,085	€ 8,586	€ 8,087	€ 7,588	€ 7,089	€ 6,590
	10%	€ 12,863	€ 12,364	€ 11,864	€ 11,365	€ 10,866	€ 10,367	€ 9,868	€ 9,369	€ 8,870	€ 8,371	€ 7,872
	0%	€ 14,145	€ 13,646	€ 13,146	€ 12,647	€ 12,148	€ 11,649	€ 11,150	€ 10,651	€ 10,152	€ 9,653	€ 9,154

Table S5(c): Impact of adherence to MammaPrint "low risk" and clinical assessment on incremental QALYs

		Proportion of Clinical "High Risk" patients receiving chemotherapy										
		0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Proportion of patients adhering to MammaPrint "Low Risk" result	100%	0,000	0,003	0,005	0,008	0,011	0,013	0,016	0,018	0,021	0,024	0,026
	90%	-0,003	0,000	0,003	0,005	0,008	0,011	0,013	0,016	0,018	0,021	0,024
	80%	-0,005	-0,003	0,000	0,003	0,005	0,008	0,011	0,013	0,016	0,018	0,021
	70%	-0,008	-0,005	-0,003	0,000	0,003	0,005	0,008	0,011	0,013	0,016	0,018
	60%	-0,011	-0,008	-0,005	0,003	0,000	0,003	0,005	0,008	0,011	0,013	0,016
	50%	-0,013	-0,011	-0,008	-0,005	-0,003	0,000	0,003	0,005	0,008	0,011	0,013
	40%	-0,016	-0,013	-0,011	-0,008	-0,005	-0,003	0,000	0,003	0,005	0,008	0,011
	30%	-0,018	-0,016	-0,013	-0,011	-0,008	-0,005	-0,003	0,000	0,003	0,005	0,008
	20%	-0,021	-0,018	-0,016	-0,013	-0,011	-0,008	-0,005	-0,003	0,000	0,003	0,005
	10%	-0,024	-0,021	-0,018	-0,016	-0,013	-0,011	-0,008	-0,005	-0,003	0,000	0,003
	0%	-0,026	-0,024	-0,021	-0,018	-0,016	-0,013	-0,011	-0,008	-0,005	-0,003	0,000

NOTE: In the clinical high/genomic low population, when around 99% of clinical high risk cases receive chemotherapy then adherence to treatment strategies guided by MammaPrint® among clinical high/genomic low cases should be at least 70% in order to be cost-effective. In cases where the proportion of chemotherapy given in clinical high risk cases drops to 50%, the adherence towards MammaPrint® low risk should be at least 90%.

Supplementary Table S6: Clinical assessment tools

Reclassification risk results of n=5,402 Luminal A (ER+/HER2-) patients from the MINDACT trial

	Low (Chemotherapy not recommended)	Intermediate (Chemotherapy may be recommended)	High (Chemotherapy recommended)	Missing
mAOL	3105	-	2297	-
MammaPrint	4040	-	1362	-
PREDICT NHS v2.0[1,2]	4655	587	134	26*
PREDICT NHS v2.0[1,3]	3658	1009	709	26*
NCCN/ASCO[25]	35	4190	1177	-

*26 patients were missing information on tumor grade, and were not able to assign an 10y overall survival probability using PREDICT v2.0.

[1]The Cambridge Breast Unit (UK) uses the absolute 10-year survival benefit from chemotherapy to guide decision making for adjuvant chemotherapy as follows: <3% chemotherapy not recommended; 3-5% chemotherapy discussed as a possible option; >5% chemotherapy recommended.

[2]Absolute benefit of 2nd generation chemotherapy, according to Cambridge Breast Unit definition above.

[3]Absolute benefit of 3nd generation chemotherapy, according to Cambridge Breast Unit definition above.

Notes: In the MINDACT trial,[45] all enrolled patients were assigned a binary risk score of "high" or "low" according to the modified Adjuvant! Online clinical assessment tool. A low clinical risk was defined as: "the 10-year probability of breast-cancer–specific survival without systemic therapy of more than 88% among women with estrogen receptor (ER)–positive tumors and more than 92% among women with ER-negative tumors, to account for the 4-percentage-point average absolute benefit of adjuvant endocrine therapy for ER-positive tumors."

All patients in the MINDACT trial with Luminal A (ER+/HER2-) tumors (n=5,402) were assigned a 10 year survival probability according to the PREDICT NHS v2.0 clinical assessment tool. Information on age at diagnosis, mode of detection, tumor size, tumor grade, number of positive nodes, ER status, HER2 status, and Ki67 status was taken from the MINDACT trial, and used within this algorithm to predict the probability of breast cancer overall survival at 10 years (https://www.evidencio.com/models/show/759). The PREDICT v2.0 algorithm calculates individual benefit of adjuvant hormone therapy automatically for all patients with ER+ tumours, and adds this to the 10 year probability of survival.

Important notes regarding data availability:

- The PREDICT v2.0 model defines Ki67 positivity as greater than 10 percent of tumour cells staining positive. However, the MINDACT trial sets this Ki67 positivity threshold at ≥14 %.
- The PREDICT model has the possibility to include whether or not a patient was screen detected or symptomatic or unknown. As the MINDACT dataset does not include this information, all patients have been set to "unknown".
- The mAOL cut off looks at 10y probability of "breast cancer specific survival" according to the Cardoso paper. In their protocol, they simply refer to it as "survival probability". PREDICT v2.0 refers to it as "breast cancer overall survival".
- The NHS recently released a version 2.1 (at the time of publication), in which the original algorithm of the 5 and 10 year survival remained the same as the PREDICT 2.0 version. One new calculation was added in version 2.1 and comprises the additional benefit of using bisphosphonates according to post-menopausal status.

Chemotherapy recommendation according to NCCN/ASCO was also determined for the n=5402 patients. All node-positive patients are recommended adjuvant chemotherapy. All node-negative patients with tumors ≤0.5 cm are not recommended adjuvant chemotherapy. Node-negative patients with tumors >0.5 cm may be recommended adjuvant chemotherapy.

Supplementary Table S7: Markov trace for Years 0, 5, 10 (State occupancies, representing number of simulated patients in each state during given year)

Table S7(a). Genomic strategy

Year	Cycle	State 1: Distant Metastasis Free Survival	State 2: Progressed Disease	State 3: Death
0	**1**	3369.00	0.00	0.00
5	**11**	3117.39	120.52	131.08
10	**21**	2724.39	257.82	368.85

Table S7(b). Clinical strategy

Year	Cycle	State 1: Distant Metastasis Free Survival	State 2: Progressed Disease	State 3: Death
0	**1**	3369.00	0.00	0.00
5	**11**	3119.29	120.19	129.52
10	**21**	2729.11	257.85	382.04

NOTE: These values are based off the simulated population outcome data, and are not representative of outcomes reported in the MINDACT trial. A Weibull parametric survival model using hazard functions derived from MINDACT trial sample data on clinical "high" risk patients with ER+/HER2- tumors was used to impute transition probabilities for the Markov trace.

Supplementary Table S8: Sensitivity analysis of country-specific population utility norms

Country	EQ-5D index value population norms by age group and total population (European VAS value set), Age group 55-64 (Janssen et al. 2019)[47]	Old Incremental QALY (based on utility of distant metastasis-free state: 0.824)	New Incremental QALY (Deterministic analysis)
Belgium	0.881	0.018	0.016
France	0.804	0.020	0.020
Germany	0.881	0.019	0.018
Netherlands	0.869	0.018	0.017
United Kingdom	0.804	0.019	0.020
United States	0.776	0.019	0.020

Supplementary Figure S1: Visual inspection of parametric survival modeling (DMFS and OS)

The following figures provide a graphical representation of all the considered distributions (Weibull, Gompertz, and Exponential), with the accompanying Kaplan-Meier curve for the first 5 years of observed MINDACT data. This has been done to visually inspect the fit of the parametric distributions to the KM curve during the observed period (0-5 years) and the extrapolated period to 10 years, for DMFS (Figures S1[a-c]) and OS (Figures S1[d-f]) separately.

Figure S1(a)

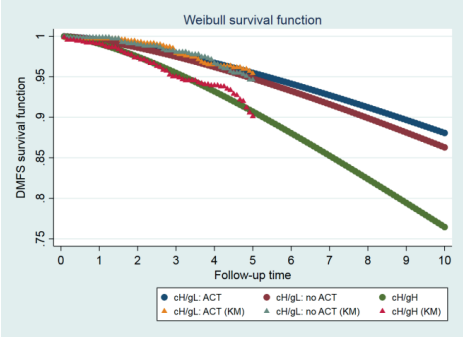


Figure S1(b)

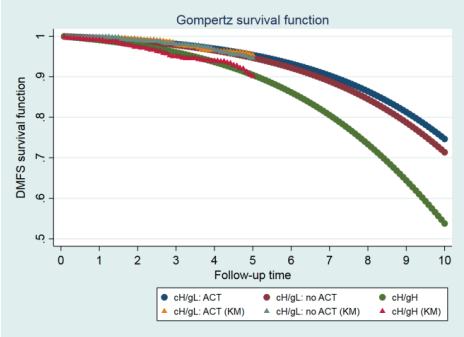


Figure S1(c)

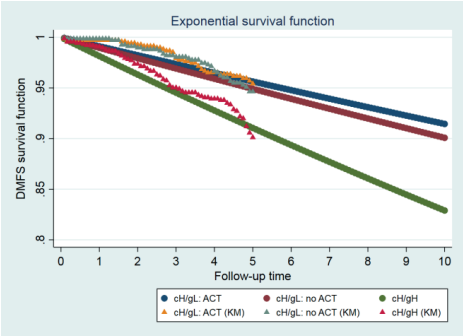


Figure S1(d)

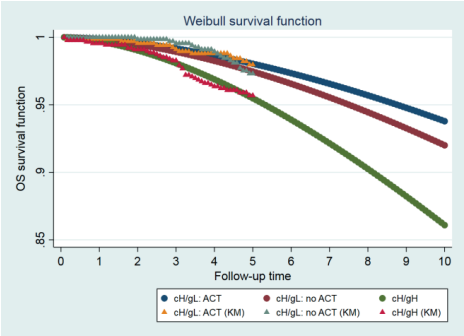


Figure S1(e)

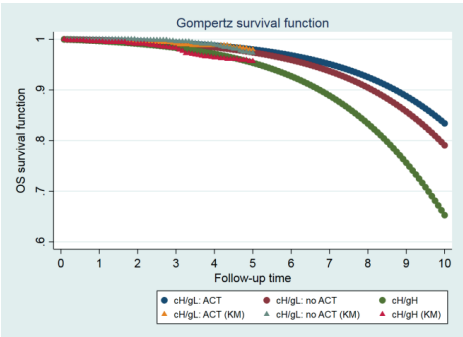
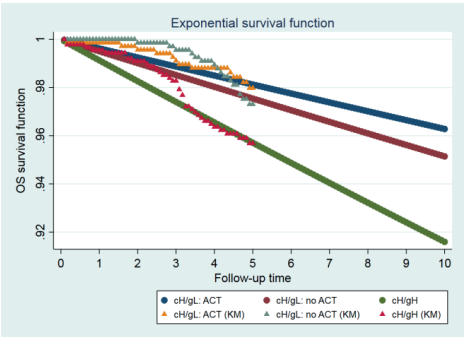


Figure S1(f)



Supplementary Figure S2: Cost-effectiveness planes for clinical high risk, ER+/HER2-

Each quadrant indicates whether a strategy is more or less expensive and more or less effective. The scale for the Y-axes are according to the country-specific currency.

Figure S2(a): Belgium

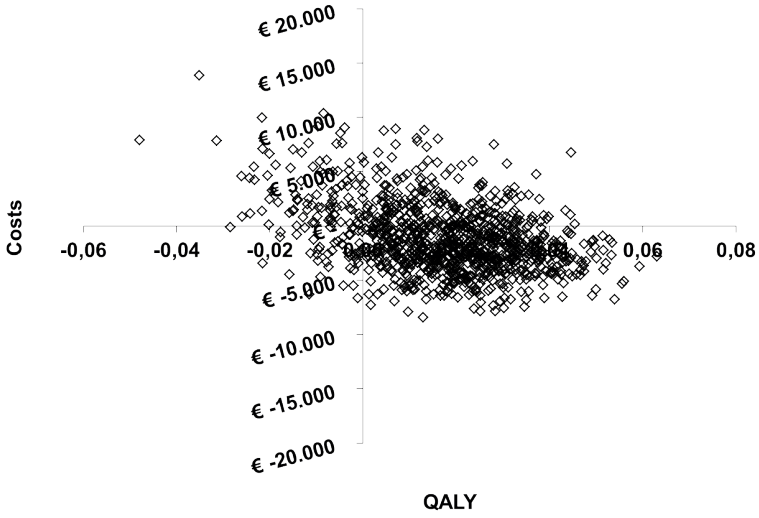


Figure S2(b): France

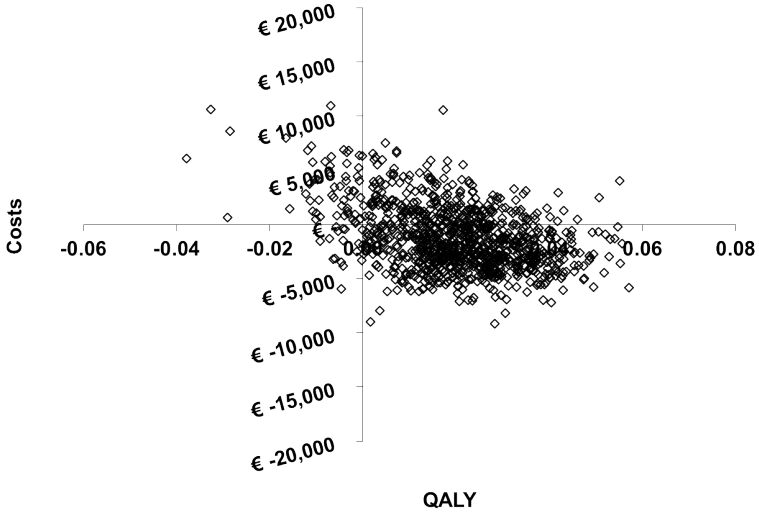


Figure S2(c): Germany

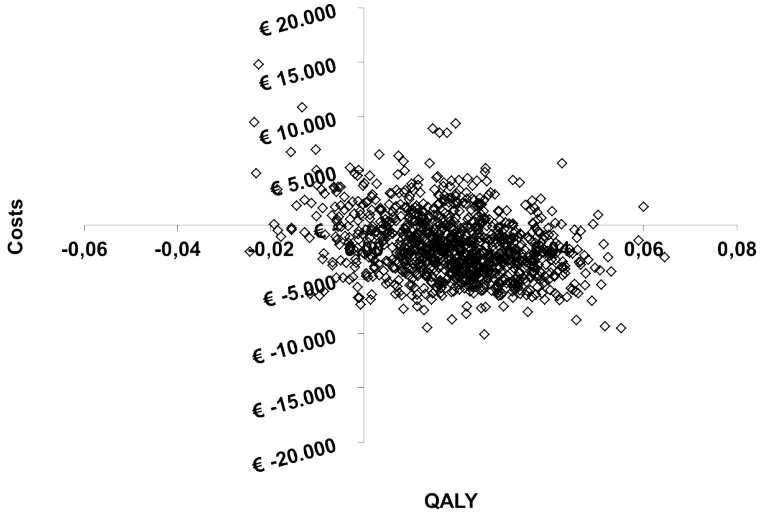


Figure S2(d): The Netherlands

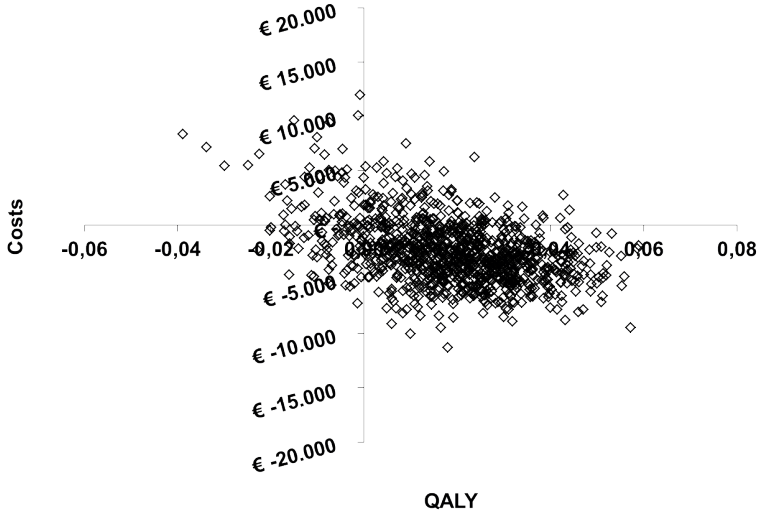


Figure S2(e): United Kingdom

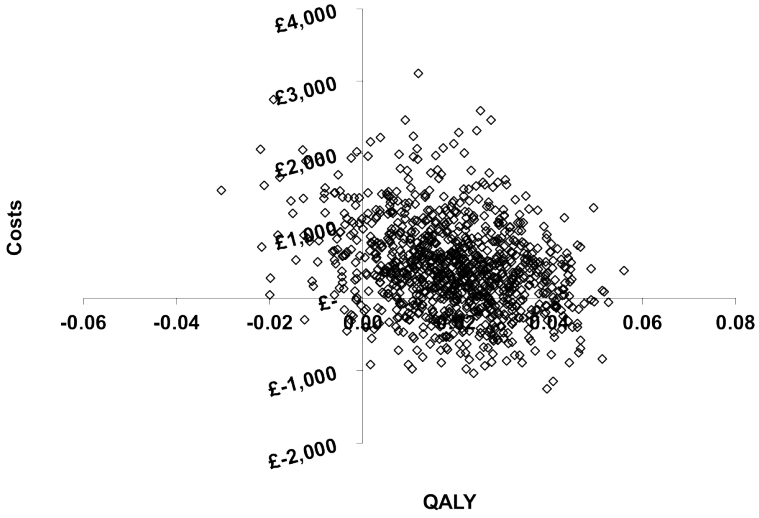
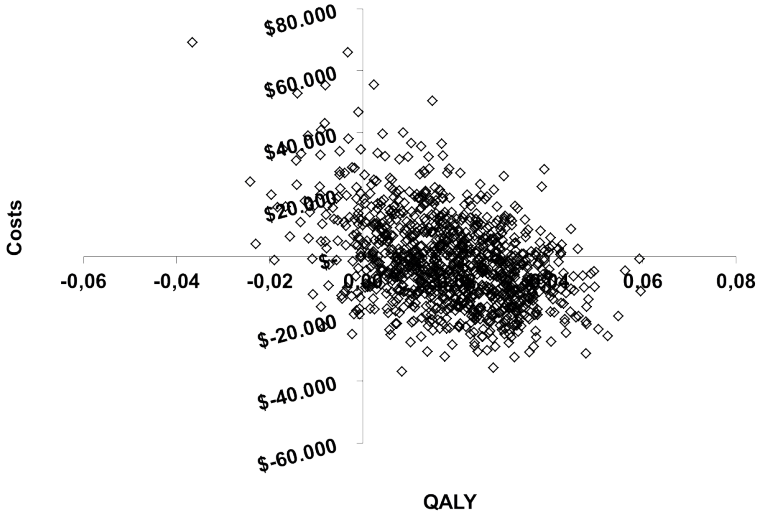
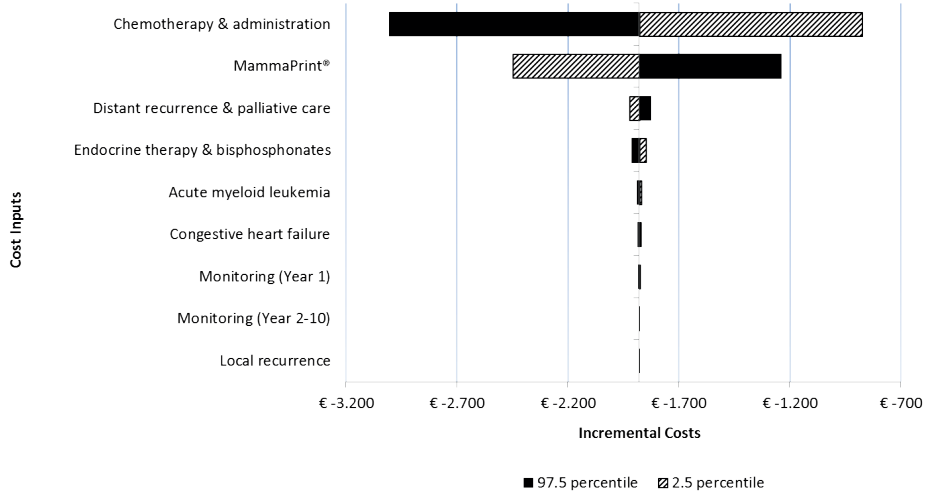


Figure S2(f): United States



Supplementary Figure S3: One-way sensitivity analysis of cost input effect on incremental costs (Germany, €)



References

1. Aarts MJ, Peters FP, Mandigers CM, et al. Primary granulocyte colony-stimulating factor prophylaxis during the first two cycles only or throughout all chemotherapy cycles in patients with breast cancer at risk for febrile neutropenia. *J Clin Oncol* 2013; **31**(34): 4290-6.
2. Hadji P, Coleman RE, Wilson C, et al. Adjuvant bisphosphonates in early breast cancer: consensus guidance for clinical practice from a European Panel. *Ann Oncol* 2016; **27**(3): 379-90.
3. Wolff AC, Blackford AL, Visvanathan K, et al. Risk of marrow neoplasms after adjuvant breast cancer therapy: the national comprehensive cancer network experience. *J Clin Oncol* 2015; **33**(4): 340-8.
4. Wang HI, Aas E, Howell D, et al. Long-term medical costs and life expectancy of acute myeloid leukemia: a probabilistic decision model. *Value Health* 2014; **17**(2): 205-14.
5. Boekel NB, Jacobse JN, Schaapveld M, et al. Cardiovascular disease incidence after internal mammary chain irradiation and anthracycline-based chemotherapy for breast cancer. *Br J Cancer* 2018; **119**(4): 408-18.
6. Biermann J, Neumann T, Angermann CE, et al. Economic burden of patients with various etiologies of chronic systolic heart failure analyzed by resource use and costs. *Int J Cardiol* 2012; **156**(3): 323-5.
7. Lowery AJ, Kell MR, Glynn RW, Kerin MJ, Sweeney KJ. Locoregional recurrence after breast cancer surgery: a systematic review by receptor phenotype. *Breast Cancer Res Treat* 2012; **133**(3): 831-41.
8. McGuire A, Lowery AJ, Kell MR, Kerin MJ, Sweeney KJ. Locoregional Recurrence Following Breast Cancer Surgery in the Trastuzumab Era: A Systematic Review by Subtype. *Ann Surg Oncol* 2017; **24**(11): 3124-32.
9. Ward S, Scope A, Rafia R, et al. Gene expression profiling and expanded immunohistochemistry tests to guide the use of adjuvant chemotherapy in breast cancer management: a systematic review and cost-effectiveness analysis. *Health Technol Assess* 2013; **17**(44): 1-302.
10. Karnon J, Kerr GR, Jack W, Papo NL, Cameron DA. Health care costs for the treatment of breast cancer recurrent events: estimates from a UK-based patient-level analysis. *Br J Cancer* 2007; **97**(4): 479-85.
11. Jerusalem G, Neven P, Marinsek N, et al. Patterns of resource utilization and cost for postmenopausal women with hormone-receptor-positive, human epidermal growth factor receptor-2-negative advanced breast cancer in Europe. *BMC Cancer* 2015; **15**: 787.
12. Thomas RJ, Williams M, Marshall C, Glen J, Callam M. The total hospital and community UK costs of managing patients with relapsed breast cancer. *Br J Cancer* 2009; **100**(4): 598-600.
13. Montero AJ, Eapen S, Gorin B, Adler P. The economic burden of metastatic breast cancer: a U.S. managed care perspective. *Breast Cancer Res Treat* 2012; **134**(2): 815-22.
14. San Miguel LD, Dubois C, Gerkens S, Harrison J, Hulstaert, F. MammaPrint® test for personalised management of adjuvant chemotherapy decisions in early breast cancer: A rapid assessment. Brussels: Health Technology Assessment (HTA) Brussels: Belgian Health Care Knowledge Centre (KCE), 2018. KCE Reports 298. D/2018/10.273/09. https://kce.fgov.be/sites/default/files/atoms/files/KCE_298_Mammaprint_tests_Report__0.pdf (Accessed August 2018)
15. Moeremans K, Annemans L. Cost-effectiveness of anastrozole compared to tamoxifen in hormone receptor-positive early breast cancer. Analysis based on the ATAC trial. *Int J Gynecol Cancer* 2006; **16 Suppl 2**: 576-8.
16. Rouzier & Hequet. Personal Communication. 2018.

17. Héquet D, Huchon C, Soilly AL, et al. Direct medical and non-medical costs of a one-year care pathway for early operable breast cancer: Results of a French multicenter prospective study. *PloS one* 2019; 14(7): e0210917.
18. Tilleul P, Jacot W, Emery C, Lafuma A, Gourmelen J. Management and cost analysis of cancer patients treated with G-CSF: a cohort study based on the French national healthcare insurance database. *J Med Econ* 2017; **20**(12): 1261-7.
19. Lux MP, Nabieva N, Hildebrandt T, et al. Budget impact analysis of gene expression tests to aid therapy decisions for breast cancer patients in Germany. *Breast* 2018; **37**: 89-98.
20. Arbeitsgemeinschaft Gynäkologische Onkologie (AGO) Breast Committee. Diagnosis and Treatment of Patients with Primary and Metastatic Breast Cancer. Recommendations 2018. https://www.ago-online.de/fileadmin/downloads/leitlinien/mamma/2018-03/Gesamt_deutsch/Alle_aktuellen_Empfehlungen_2018.pdf (accessed August 2018).
21. Rote Liste: Arzneimittelinformationen für Deutschland. 2018. https://online.rote-liste.de/ (accessed August 2018).
22. Zorginstituut Nederland. Guideline for economic evaluations in healthcare, 2016. https://english.zorginstituutnederland.nl/publications/reports/2016/06/16/guideline-for-economic-evaluations-in-healthcare. (accessed Ausgust 2018).
23. Zorginstituut Nederland. Medicijnkosten. https://www.medicijnkosten.nl/. (accessed Ausgust 2018).
24. Aapro MS, Bohlius J, Cameron DA, et al. 2010 update of EORTC guidelines for the use of granulocyte-colony stimulating factor to reduce the incidence of chemotherapy-induced febrile neutropenia in adult patients with lymphoproliferative disorders and solid tumours. *Eur J Cancer* 2011; **47**(1): 8-32.
25. Gradishar WJ, Anderson BO, Balassanian R, et al. NCCN Guidelines Insights: Breast Cancer, Version 1.2017. *J Natl Compr Canc Netw* 2017; **15**(4): 433-51.
26. Hall PS, Smith A, Hulme C, et al. Value of Information Analysis of Multiparameter Tests for Chemotherapy in Early Breast Cancer: The OPTIMA Prelim Trial. *Value Health* 2017; **20**(10): 1311-8.
27. Blue Cross Blue Shield. Personal Communication. 2018.
28. Giordano SH, Niu J, Chavez-MacGregor M, et al. Estimating regimen-specific costs of chemotherapy for breast cancer: Observational cohort study. *Cancer* 2016; **122**(22): 3447-55.
29. Xie J, Hao Y, Zhou ZY, Qi CZ, De G, Gluck S. Economic Evaluations of Everolimus Versus Other Hormonal Therapies in the Treatment of HR+/HER2- Advanced Breast Cancer From a US Payer Perspective. *Clin Breast Cancer* 2015; **15**(5): e263-76.
30. Retèl VP, Groothuis-Oudshoorn CGM, Aaronson NK, Brewer NT, Rutgers EJT, van Harten WH. Association between genomic recurrence risk and well-being among breast cancer patients. *BMC Cancer* 2013; **13**: 295-.
31. Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997; **35**(11): 1095-108.
32. Campbell HE, Epstein D, Bloomfield D, et al. The cost-effectiveness of adjuvant chemotherapy for early breast cancer: A comparison of no chemotherapy and first, second, and third generation regimens for patients with differing prognoses. *Eur J Cancer* 2011; **47**(17): 2517-30.
33. Lidgren M, Wilking N, Jonsson B, Rehnberg C. Health related quality of life in different states of breast cancer. *Qual Life Res* 2007; **16**(6): 1073-81.

34. Vemer P, Corro Ramos I, van Voorn GA, Al MJ, Feenstra TL. AdViSHE: A Validation-Assessment Tool of Health-Economic Models for Decision Makers and Model Users. *PharmacoEconomics* 2016; **34**(4): 349-61.
35. Krop I, Ismaila N, Barlow W, Stearns V. Reply to J.L. Blum et al and S. Lange et al. *J Clin Oncol* 2018; **36**(4): 430-1.
36. Lange S, Scheibler F, Fleer D, Windeler J. Interpretation of the Results of the MINDACT Study and Consequent Recommendations in the Updated ASCO Clinical Practice Guideline. *J Clin Oncol* 2018; **36**(4): 429-30.
37. Cleemput I, van Wilder P, Vrijens F, Huybrechts M, Ramaekers D. Guidelines for pharmacoeconomic evaluations in Belgium. Brussels: Health Care Knowledge Centre (KCE), 2008. https://kce.fgov.be/en/guidelines-for-pharmacoeconomic-evaluations-in-belgium (accessed August 2018).
38. Institute for Quality and Efficiency in Health Care. Technical Document: Modelling Cologne, 2008. https://www.iqwig.de/download/TD_CBA_Modellling_v_1_0.pdf (accessed August 2018)
39. Haute Autorité de Santé. Choices in Methods for Economic Evaluation: A methodological guide, 2012. https://www.has-sante.fr/portail/upload/docs/application/pdf/2012-10/choices_in_methods_for_economic_evaluation.pdf (accessed August 2018).
40. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual, 2014. https://www.nice.org.uk/media/default/about/what-we-do/our-programmes/developing-nice-guidelines-the-manual.pdf (accessed August 2018).
41. Agendia NV. Personal Communication. 2018.
42. Shapiro CL, Moriarty JP, Dusetzina S, et al. Cost-Effectiveness Analysis of Monthly Zoledronic Acid, Zoledronic Acid Every 3 Months, and Monthly Denosumab in Women With Breast Cancer and Skeletal Metastases: CALGB 70604 (Alliance). *J Clin Oncol* 2017; **35**(35): 3949-55.
43. Hakkart-van Roijen L, van der Linden N, Bouwmans C, Kanters T, Tan SS. Kostenhandleiding: Methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg. (in Dutch). Rotterdam: Zorginstituut Nederland, 2015. https://www.zorginstituutnederland.nl/binaries/zinl/documenten/publicatie/2016/02/29/richtlijn-voor-het-uitvoeren-van-economische-evaluaties-in-de-gezondheidszorg/Richtlijn+uitvoeren+van+economische+evaluaties+in+de+gezondheidszorg+%28verdiepingsmodules%29.pdf (accessed August 2018).
44. Nolan MT, Plana JC, Thavendiranathan P, Shaw L, Si L, Marwick TH. Cost-effectiveness of strain-targeted cardioprotection for prevention of chemotherapy-induced cardiotoxicity. *Int J Cardiol* 2016; **212**: 336-45.
45. Cardoso F, van't Veer LJ, Bogaerts J, et al. 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. *N Engl J Med* 2016; **375**(8): 717-29.
46. Ministère des Solidarités et de la Santé. Le référentiel des actes innovants hors nomenclature de biologie et d'anatomopathologie (RIHN). Updated 09 April 2020. https://solidarites-sante.gouv.fr/systeme-de-sante-et-medico-social/recherche-et-innovation/rihn (accessed May 2020)
47. Janssen MF, Szende A, Cabases J, et al: Population norms for the EQ-5D-3L: a cross-country analysis of population surveys for 20 countries. *Eur J Health Econ* 2019; 20(2): 205-16.


7

Chapter 7

Outcome without any adjuvant systemic treatment in stage I ER+/ HER2– breast cancer patients included in the MINDACT trial

Josephine M.N. Lopes Cardozo [1,2], Danalyn Byng [3], Caroline A. Drukker [4], Marjanka K. Schmidt [5], Laura J. van 't Veer [6], Fatima Cardoso [7], Martine Piccart [8], Carolien H. Smorenburg [9], Coralie Poncet [2], Emiel J.Th. Rutgers [1]

[1] Department of Surgical Oncology, Netherlands Cancer Institute, Amsterdam, The Netherlands
[2] EORTC Headquarters, Brussels, Belgium
[3] Division of Psychosocial Research and Epidemiology, Netherlands Cancer Institute, Amsterdam, The Netherlands
[4] Department of Surgery, Tergooi Hospital, Hilversum, The Netherlands
[5] Division of Molecular Pathology, Netherlands Cancer Institute, Amsterdam, The Netherlands
[6] Department of Laboratory Medicine, UCSF Helen Diller Family Comprehensive Cancer Center, San Francisco, USA
[7] Breast Unit, Champalimaud Clinical Center/Champalimaud Foundation, Lisbon, Portugal
[8] Institut Jules Bordet, Université Libre de Bruxelles, Brussels, Belgium
[9] Department of Medical Oncology, Netherlands Cancer Institute, Amsterdam

Ann Oncol **33(3),** 310–320 (2022). https://doi.org/10.1016/j.annonc.2021.11.014

Highlights

- Patients who received no AST had a 2.5% lower 8-year DMFI rate than that observed in matched patients who received ET.
- Slightly more locoregional recurrences and contralateral breast cancers were observed in patients who received no AST.
- These effects and side-effects of ET should be discussed with patients even at a very low risk of distant metastasis.

Abstract

Background

Adjuvant systemic treatments (AST) reduce mortality, but have associated short- and long-term toxicities. Careful selection of patients likely to benefit from AST is needed. We evaluated outcome of low-risk breast cancer patients of the EORTC 10041/BIG 3-04 MINDACT trial who received no AST.

Patients and methods

Patients with estrogen receptor-positive, HER2-negative, lymph node-negative tumors ≤2 cm who received no AST were matched 1 : 1 to patients with similar tumor characteristics treated with adjuvant endocrine therapy (ET), using propensity score matching and exact matching on age, genomic risk (70-gene signature) and grade. In a post hoc analysis, distant metastasis-free interval (DMFI) and overall survival (OS) were assessed by Kaplan–Meier analysis and hazard ratios (HR) by Cox regression. Cumulative incidences of locoregional recurrence (LRR) and contralateral breast cancer (CBC) were assessed with competing risk analyses.

Results

At 8 years, DMFI rates were 94.8% [95% confidence interval (CI) 92.7% to 96.9%] in 509 patients receiving no AST, and 97.3% (95% CI 95.8% to 98.8%) in 509 matched patients who received only ET [absolute difference: 2.5%, HR 0.56 (95% CI 0.30-1.03)]. No statistically significant difference was seen in 8-year OS rates, 95.4% (95% CI 93.5% to 97.4%) in patients receiving no AST and 95.6% (95% CI 93.8% to 97.5%) in patients receiving only ET [absolute difference: 0.2%, HR 0.86 (95% CI 0.53-1.41)]. Cumulative incidence rates of LRR and CBC were 4.7% (95% CI 3.0% to 7.0%) and 4.6% (95% CI 2.9% to 6.9%) in patients receiving no AST versus 1.4% (95% CI 0.6% to 2.9%) and 1.5% (95% CI 0.6% to 3.1%) in patients receiving only ET.

Conclusions

In patients with stage I low-risk breast cancer, the effect of ET on DMFI was limited, but overall significantly fewer breast cancer events were observed in patients who received ET, after the relatively short follow-up of 8 years. These benefits and side-effects of ET should be discussed with all patients, even those at a very low risk of distant metastasis.

Introduction

The era of personalized medicine in breast cancer focuses on escalating treatment in patients with a high risk of developing distant metastasis or death and de-escalating treatment in low-risk patients, aiming to avoid under- and overtreatment, respectively.[1,2] Several gene signatures have proven to be successful in identifying a subgroup of patients for whom de-escalation of treatment by omitting adjuvant chemotherapy can be considered.[3,4,5] The use of gene signatures for deciding on administration of adjuvant chemotherapy has been included in international breast cancer guidelines.[2,6,7] For low-risk, early-stage breast cancer patients with estrogen receptor-positive (ER+) disease, who have no indication for adjuvant chemotherapy, adjuvant endocrine therapy (ET) is the standard choice of treatment.[2,7] In all ER+ breast cancers, ET reduces the risk of breast cancer death by ~30%.8 Where adherence rates to ET are relatively high in the setting of clinical trials, the adherence rates in clinical practice are relatively poor; only 50% of breast cancer patients successfully complete 5 years of therapy.[9,10] As expected, side-effects of ET contribute to non-adherence, but low recurrence risk perception is also a factor.[9] Frequently reported side-effects are vasomotor symptoms such as hot flashes and night sweats, musculoskeletal symptoms such as arthralgia and osteoporosis and vulvo-vaginal symptoms such as vaginal dryness or discharge and dyspareunia.[11]

The increasing understanding of the biological heterogeneity of breast cancer and the identification of molecular subtypes and gene-expression signatures results in a continuing refinement of prediction algorithms in early-stage breast cancer.[2] The online PREDICT tool is frequently used by clinicians to determine the 'baseline' risk at diagnosis and the added benefit of AST.[12,13] In the national guideline for breast cancer treatment in the Netherlands, for patients aged 35 years or older with ER+, HER2-negative (HER2–), lymph node-negative (N0) breast cancer with a Bloom and Richardson grade 1 tumor ≤2 cm or a grade 2 or 3 tumor ≤1 cm, the omission of all AST can be considered.14 In international guidelines and national guidelines of other countries, these patients would typically be advised to start ET.[2,7,15-18]

In the randomized, phase III MINDACT trial, treatment allocation and randomization were based on risk stratification. Standard clinical–pathological characteristics were used to determine clinical risk and the 70-gene signature MammaPrint was used to determine genomic risk.[3,4] Patients identified as low risk by both methods did not receive chemotherapy as per study design. These patients were treated according to local guidelines, with most ER+ patients receiving ET.[3,4] However, there is also a group of patients in MINDACT who received no AST. As the Netherlands was the largest recruiter for the MINDACT trial, and their national guideline permits the

omission of all AST in a subgroup of low-risk patients, the majority of the patients in MINDACT who received no AST came from the Netherlands.

The aim of the study was to evaluate the survival of breast cancer patients who participated in the MINDACT trial and did not receive any type of AST and to compare their outcomes to those patients with similar characteristics who did receive ET.

Methods

The MINDACT trial

This study is a downstream project with an exploratory subgroup analysis of patients included in the MINDACT trial (EORCT 10041/BIG 3-04). Women aged 18-70 years with histologically proven operable invasive breast cancer (T1-3N0-1M0) were enrolled in the MINDACT trial between 2007 and 2011.[3,4] Treatment allocation and randomization were based on clinical and genomic risk. A low clinical risk was defined as a 10-year probability of breast cancer-specific survival without AST of >88% for women with ER+ tumors. Clinical risk was determined by Adjuvant! Online (modified from version 8.0 including HER2 status), based on tumor size, hormone and HER2 receptor status, grade and nodal status. In brief, in ER+/HER2– patients, clinical risk in MINDACT was defined as low for N0 patients with grade 1 tumors ≤3 cm, grade 2 tumors ≤2 cm and grade 3 tumors ≤1 cm and for node positive patients only if the tumor was grade 1 ≤2 cm. Genomic risk was based on the 70-gene signature (MammaPrint®).[3,4] Clinical–pathological characteristics, treatment characteristics and outcome data for this study were obtained from the EORTC 10041/BIG 3-04 MINDACT trial database. The cut-off date for the follow-up in this analysis was 26 February 2020; the median follow-up was 8.6 years.[4] The MINDACT trial was approved by the ethics committees of all participating sites, and all patients gave written informed consent. Ethics approval was also obtained for the analysis presented in this study.

Patients and matching

According to the Dutch national guidelines, the omission of AST can be considered in a subgroup of patients with clinical low-risk breast cancer.[14] For the purpose of this analysis, patients with ER+/HER2– N0 tumors, with a tumor size ≤2 cm, were selected from the MINDACT trial database. In this subgroup, 509 patients received no AST (neither chemotherapy nor endocrine therapy nor trastuzumab). These patients could have genomic low- or high-risk tumors.

Using propensity score and exact matching, we aimed to identify a group of patients treated with adjuvant ET with similar tumor characteristics to the patients who received no AST.19 In the subgroup of patients with ER+/HER2– N0 tumors ≤2 cm, patients were matched 1 : 1 based on the propensity score, age (≤50 versus >50 years), genomic risk according to the 70-gene signature (low risk versus high risk) and grade (Supplementary Methods). This resulted in a matched group of 509 patients who received only ET. As we pre-selected patients with tumors ≤2 cm, we did not further match on exact tumor size to increase the number of possible matches. Type of breast and axillary surgery were not included as variables for matching as the impact of differences in the type of surgery on local recurrence and 15-year breast cancer mortality is limited.[20]

Statistical analysis

The primary endpoint for this study was distant metastasis-free interval (DMFI), defined as the time from enrolment until first distant metastasis or breast cancer-related death (including death from unknown cause). Secondary endpoints included overall survival (OS), defined as the time until death from any cause, and breast cancer-specific survival (BCSS), defined as the time until breast cancer-related death. For the primary analysis, all endpoints were assessed in a time-to-event analysis according to the Kaplan–Meier method, and compared for adjuvant systemic therapy (no AST versus only ET) with the log-rank test. Accompanying hazard ratios (HR) were calculated using a univariate Cox proportional hazards model. Data for patients who had no event at the cut-off date were censored at the time of last disease assessment for DMFI and at the last follow-up date for OS and BCSS.

Two secondary analyses were planned; in the first analysis, the estimated benefit of ET according to the PREDICT tool (PREDICT Breast V2.1, The Winton Centre for Risk & Evidence Communication, Centre for Mathematical Sciences, Cambridge) was calculated for all patients who received no AST, based on their clinical–pathological characteristics. In the second analysis, the cumulative incidence functions for locoregional recurrence, contralateral breast cancer (CBC) and distant metastasis were estimated, and sub-distribution HR were estimated in a competing risk model according to the Fine and Gray method.[21,22] For the time to first locoregional recurrence, distant metastasis or any death without locoregional recurrence were considered competing risks. For the time to first CBC, any death without CBC was considered a competing risk. For the time to first distant metastasis, any death without distant metastasis was considered a competing risk. For an additional exploratory analysis, a breast cancer-free interval (BCFI) endpoint was compiled specifically for this study, which was not included in the MINDACT protocol. The STEEP criteria were used for defining this endpoint, BCFI was defined as the time

from enrolment until first distant metastasis, locoregional recurrence, CBC or breast cancer-related death (including death from unknown cause), and analyzed in a time-to-event analysis.[23]

A statistically significant finding was defined as a two-sided P value < 0.05. As this is a post hoc analysis, all analyses are underpowered and attention should go to the 95% confidence intervals (CIs) when interpreting the results. All analyses were carried out using SPSS (version 25.0) and R (version 3.6.3). The R 'MatchIt' package was used for the propensity score and exact matching, and the 'cmprsk' package was used for the competing risk analyses.

PREDICT

The PREDICT tool (PREDICT Breast V2.1) was used to calculate the estimated OS and BCSS for patients who received no AST.[12,13] The estimates for survival and benefit of ET were obtained from 1000 bootstrap samples using the percentile method to account for skewness. Estimations were based on the clinical–pathological characteristics available in the MINDACT database. Ki-67 was available for 87% of patients for whom central pathology review was carried out on their primary tumor sample. For the remaining patients, Ki-67 was coded as unknown. Furthermore, method of detection was not included in the MINDACT database. For this analysis, patients aged >50 years were assigned to have screen-detected tumors, and patients aged ≤50 years to have clinically detected tumors. Sensitivity analyses assigning all patients aged >50 years to have clinically detected tumors did not change the conclusions.

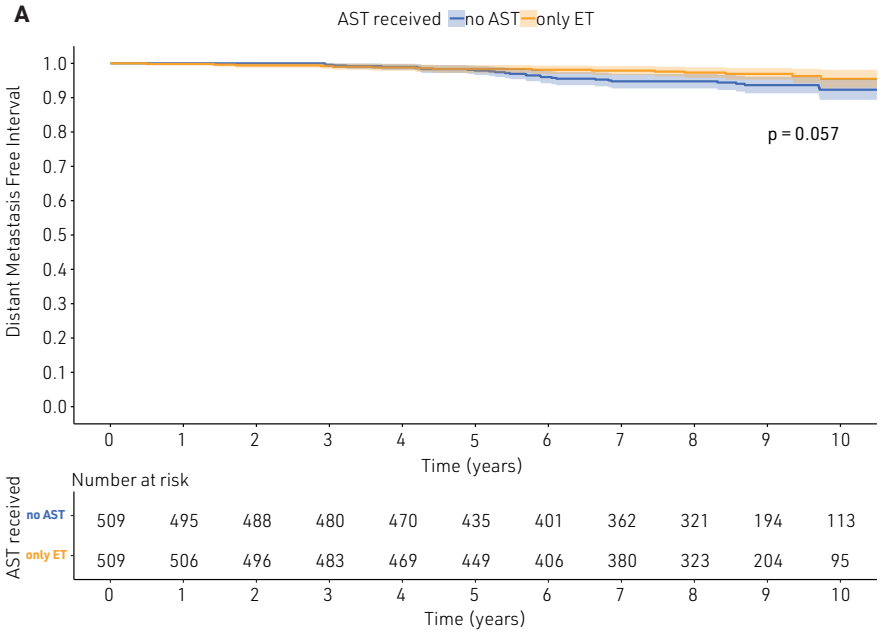
Results

Patient and tumor characteristics

Patients and tumor characteristics of the patients who received no AST and matched group of patients who received only ET are described in Table 1. All patients had ER+/HER2– N0 tumors. In both groups, 95% of patients had a genomic low risk, 76% of patients were older than 50 years, and 65% and 34% of patients had grade 1 and 2 tumors, respectively. Of the 509 patients, 508 patients (99.8%) who received no AST and 504/509 patients (99%) who received only ET were defined as clinical low risk. All patients had tumors ≤2 cm, but a statistically significant difference was seen between tumors ≤1 cm and 1-2 cm. In patients who received no AST, 43% had tumors ≤1 cm, compared to 26% in patients who received only ET (P < 0.001). As expected, a large percentage of patients who received no AST (87%) were from the Netherlands, with the other patients in both groups represented by six other European countries.

Survival outcomes according to adjuvant systemic therapy received

Figure 1 shows the outcomes for DMFI, OS and BCSS for all patients (n = 1018) stratified by adjuvant systemic therapy received (no AST versus only ET). The DMFI rates at 8 years were 94.8% (95% CI 92.7% to 96.9%) for patients who received no AST and 97.3% (95% CI 95.8% to 98.8%) for patients who received only ET [absolute difference at 8 years: 2.5%, HR 0.56 (95% CI 0.30-1.03)].



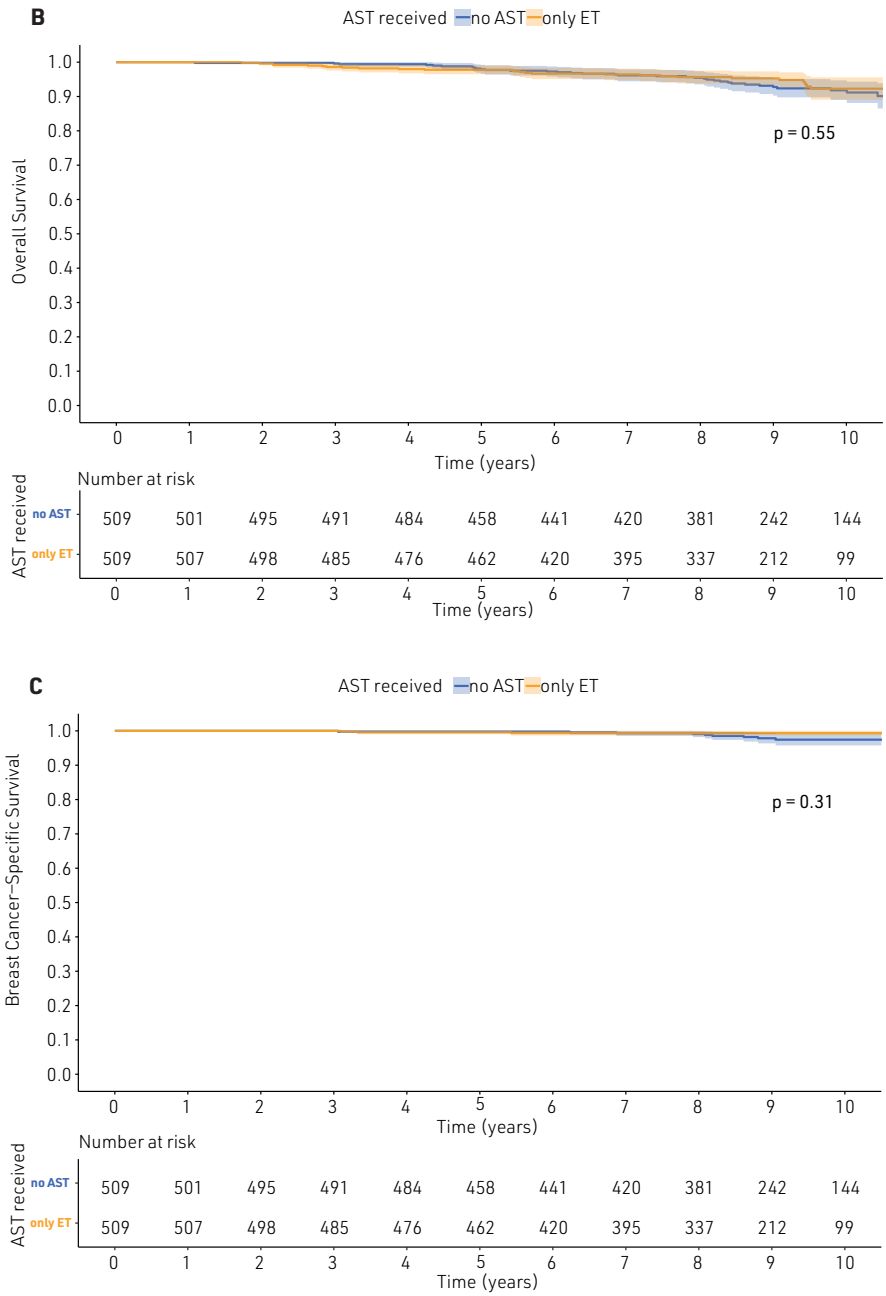


Figure 1. Kaplan–Meier analysis of (A) distant metastasis-free interval (DMFI), (B) overall survival (OS) and (C) breast cancer-specific survival (BCSS) stratified by adjuvant systemic treatment (AST) received. Median follow-up was 8.6 years for this cohort. Estimates are reported at 8 years because at that time point there were still a sufficient number of patients at risk. Kaplan–Meier curves are displayed until 10 years of follow-up due to the high level of censoring beyond this time point. Shading around the curves shows 95% confidence intervals.

OS rates at 8 years were 95.4% (95% CI 93.5% to 97.4%) for patients who received no AST, and 95.6% (95% CI 93.8% to 97.5%) for patients who received only ET [absolute difference at 8 years: 0.2%, HR 0.86 (95% CI 0.53-1.41)].

BCSS rates at 8 years were 99.1% (95% CI 98.2% to 100%) for patients who received no AST, and 99.4% (95% CI 98.6% to 100%) for patients who received only ET [absolute difference at 8 years: 0.3%, HR 0.58 (95% CI 0.20-1.69)].

A summary table with the survival estimates for all endpoints is provided in Supplementary Table S1. Due to the difference in proportions of tumors ≤1 cm and 1-2 cm between groups, DMFI was evaluated according to tumor size, but no statistically significant difference in DMFI was seen in both groups (log-rank P = 0.94 and P = 0.49, respectively) (Supplementary Figure S1).

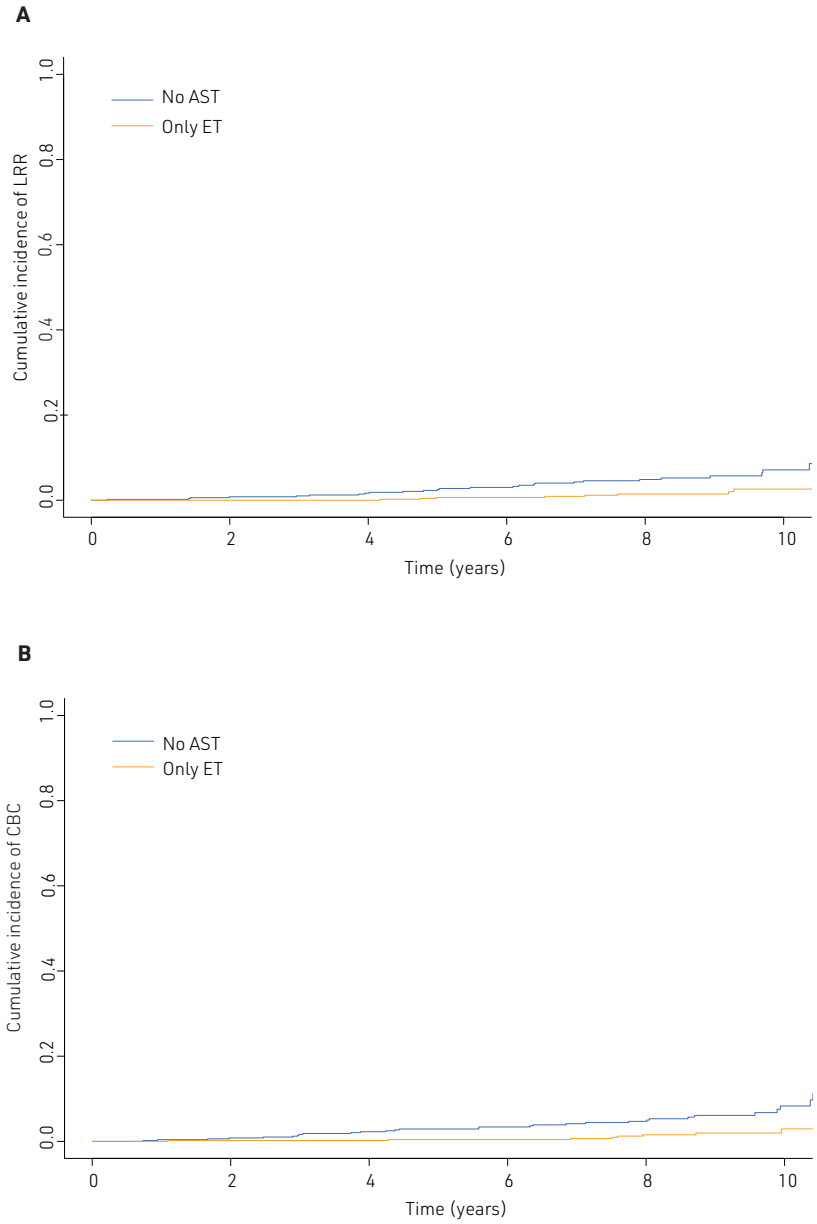
Estimated benefit of ET according to PREDICT in patients receiving no AST

For all patients who received no AST, the estimated OS and BCSS without adjuvant systemic therapy and the added benefit of ET were assessed using the PREDICT tool. The results are summarized in Supplementary Table S2. The estimated OS rate for the 509 patients without AST according to PREDICT was 91.6% at 8 years and 80.0% at 15 years. The estimated BCSS rate without AST according to PREDICT was 97.2% at 8 years and 94.1% at 15 years. The estimated added benefit of ET for this group was 0.86% at 8 years and 1.65% at 15 years according to PREDICT.

Cumulative incidence of locoregional recurrence, contralateral breast cancer and distant metastasis according to adjuvant systemic therapy received

During the entire follow-up period, 27 patients had a locoregional recurrence, 31 patients had a CBC and 23 patients had distant metastasis in those who received no AST, compared to 8, 8 and 11 patients who received only ET, respectively. Figure 2 shows the cumulative incidence of locoregional recurrence, CBC and distant metastasis in both groups. The cumulative incidence of locoregional recurrence at 8 years was 4.7% (95% CI 3.0% to 7.0%) in patients who received no AST and 1.4% (95% CI 0.6% to 2.9%) in patients who received only ET [sub-distribution hazard ratio (SHR): 0.30; 95% CI 0.14-0.66]. The cumulative incidence of CBC at 8 years was 4.6% (95% CI 2.9% to 6.9%) in patients who received no AST and 1.5% (95% CI 0.6% to 3.1%) in patients who received only ET (SHR: 0.26; 95% CI 0.12-0.56). The cumulative incidence of distant metastasis at 8 years was 3.9% (95% CI 2.3% to 6.0%) in patients who received no AST and 2.0% (95% CI 1.0% to 3.6%) in patients who received only ET (SHR: 0.48; 95% CI 0.24-0.99). Figure 3 shows the outcome for BCFI, combining locoregional recurrence,

CBC, distant metastasis and breast cancer-related death into one composite endpoint. At 8 years, BCFI rates were 86.5% (95% CI 83.3% to 89.8%) in patients who received no AST and 94.8% (95% CI 92.7% to 97.0%) in patients who received only ET [absolute difference at 8 years: 8.3%, HR 0.37 (95% CI 0.24-0.57)].



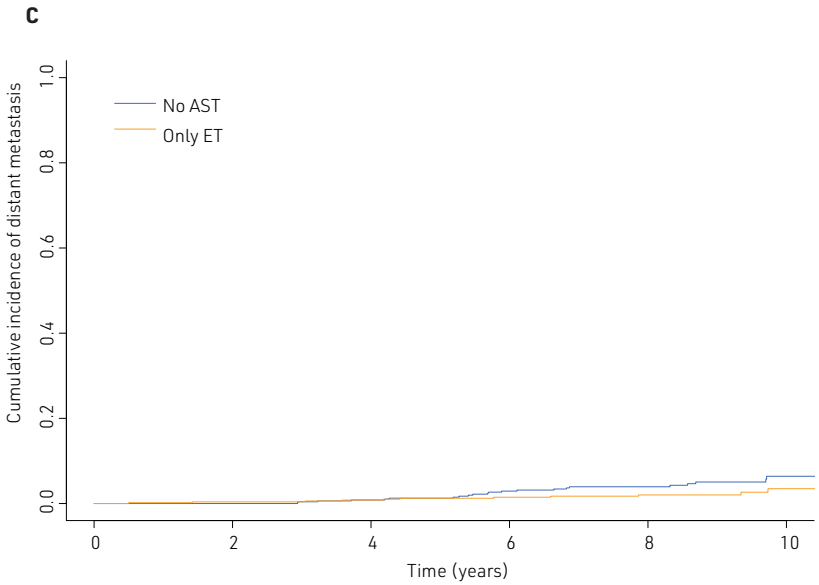


Figure 2. Cumulative incidence of (A) locoregional recurrence, (B) contralateral breast cancer and (C) distant metastasis for the no-AST and only-ET group.

AST, adjuvant systemic treatment; CBC, contralateral breast cancer; ET, endocrine therapy; LRR, locoregional recurrence.

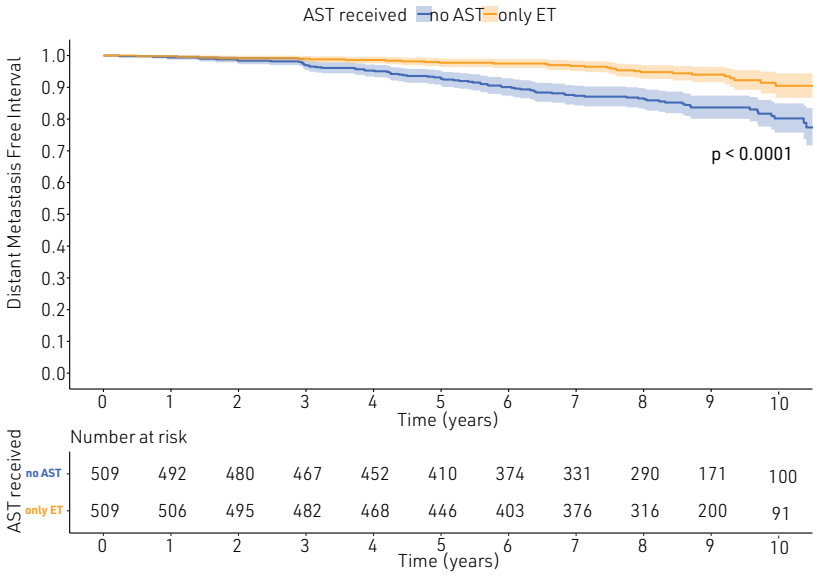


Figure 3. Kaplan–Meier analysis of breast cancer-free interval (BCFI) stratified by adjuvant systemic treatment (AST) received. Median follow-up was 8.6 years for this cohort. Estimates are reported at 8 years because at that time point there were still a sufficient number of patients at risk. Kaplan–Meier curves are displayed until 10 years of follow-up due to the high level of censoring beyond this time point. Shading around the curves shows 95% confidence intervals.

Discussion

In this study, patients with ER+/HER2– N0 ≤2 cm tumors who received no AST had an 8-year DMFI rate of 94.8% which was 2.5% lower than that observed in a matched group of patients who received ET. After a relatively short follow-up of 8 years, no statistically significant difference was observed in OS and BCSS between patients who received no AST or only ET. According to PREDICT,[12,13] the estimated benefit of ET on OS and BCSS for the group of patients who received no AST was <2% up to 15 years. However, the cumulative incidence of locoregional recurrences and contralateral breast cancers at 8 years was ~3% higher in patients who received no AST compared to patients who received ET, contributing to an 8.3% difference in BCFI at 8 years when considering all breast cancer events together.

Breast cancer recurrences are understood to occur at a steady rate up to 20 years after diagnosis in early-stage breast cancer patients with ER+ disease who were treated with 5 years of ET.[24] The Early Breast Cancer Trialists' Collaborative Group (EBCTCG) overview demonstrated that among patients with N0 tumors ≤2 cm, the annual rate of distant metastasis or any breast cancer event was between 1% and 2% up to 20 years post diagnosis.24 The overview emphasizes the clinical problem of late recurrences, specifically in ER+ breast cancer. This phenomenon should be considered alongside any attempt to de-escalate treatment.

The follow-up in our study does not allow for conclusions about long-term consequences of omitting ET altogether. However, if the occurrence of distant metastases is an earlier indicator and surrogate for OS, we can expect a continuation of the trend we see for distant metastasis in our analysis.[25] Several studies in historic cohorts have reported excellent long-term BCSS in subgroups of low-risk patients with ER+/HER2– N0 disease who received no AST. Although these cohorts were limited in size (range 44-124 patients), these studies suggest an excellent long-term survival without AST in a number of breast cancer patients.[26-29]

Breast cancer patients with ER+/HER2– disease have lower rates of locoregional recurrences compared with other subtypes,[30-33] and patients with ER+ disease treated with ET have lower rates of any breast cancer recurrence than patients who receive no ET.34 In studies reporting on relatively similar subgroups of ER+/HER2– patients as in our study, the cumulative incidence of locoregional recurrence ranges from 0.8% to 2.2% at ~5 years[32,33] to 9% at 10 years in an older trial.[31]

Studies from general breast cancer populations show that the cumulative incidence of CBC is around 0.4% per year.[35,36] Furthermore, multiple studies show that systemic

therapies reduce the risk of CBC.[35-38] In our study, we observed a higher cumulative incidence of both locoregional recurrences and CBCs in patients who received no AST. This is within the range mentioned for locoregional recurrence, but higher for CBC. The incidence rates of both locoregional recurrence and CBC in patients receiving ET in our study are somewhat lower than that reported in the literature. This could be due to higher compliance to ET in this well-monitored prospective trial, and the selectivity of patient populations in clinical trials. Nevertheless, this study confirms that in addition to lowering the risk of progression to distant metastases, ET also reduces the incidence of locoregional recurrences and contralateral breast cancers in these low-risk patients. While locoregional recurrences and contralateral breast cancers have a limited impact on survival compared to distant metastases,[39,40] treatment is required and this has a demonstrably negative impact on most dimensions of health-related quality of life.[41]

Studies on adjuvant ET in breast cancer show an effect of similar magnitude across all subgroups with ER+ disease.[8] However, the impact of treatment on the relative risk of progression or recurrence may not be as meaningful to an individual patient, for whom the change in absolute risk is more relevant. This was reflected by the very low estimated added benefit of ET according to PREDICT for the group of no AST patients in our study. While patients derive limited benefit from ET, they are exposed to its side-effects, which are often underestimated. The most frequent side-effects of ET can be categorized into three groups: (i) vasomotor symptoms, including hot flashes and night sweats are reported in 35%-40%; (ii) vulvo-vaginal symptoms, including dyspareunia, vaginal dryness, vaginal bleeding and vaginal discharge are reported in 20%-40%; and (iii) musculoskeletal symptoms, including arthralgia and osteoporosis, leading to a higher incidence of bone fractures, are reported in 30%-60%.[11] Rare but serious side-effects of ET include endometrial cancer, venous thromboembolism and fractures and are reported in 0.5%-6% of patients. Generally, higher incidences of toxicities are reported in patients receiving aromatase inhibitors compared to tamoxifen, and the addition of ovarian function suppression also increases symptoms.[11] Treatment with ET and its side-effects have a considerable impact on the patients' quality of life, when compared with patients who receive no ET, contributing to poor adherence.[9,11,42,43] In patients with a very low risk of distant recurrence and a limited expected benefit of ET, the benefits of the treatment may not outweigh its harms. Ultimately, it is the decision of the patient who, after being informed on the risks and the expected clinical benefit of the considered treatments as well as their harms, should decide whether the benefits outweigh the harms and if the risk without treatment is acceptable.

The follow-up time in this study is not long enough for a complete assessment of DMFI, OS and BCSS, and is a limitation in evaluating this group of women with low-risk breast cancer. However, we circumvent the ethical constraints of long-term trials randomizing patients to ET or no AST. Using propensity score matching to select a control group of patients receiving ET is the best possible approach for such an analysis. There is an imbalance in the country of enrolment between patients in both groups, due to differences in national treatment guidelines, but as expected there was no difference in survival based on country (data not shown). There is also an imbalance in tumor size between both groups, with larger tumors (1-2 cm) in patients receiving ET. However, we observed no differences in DMFI based on tumor size in both groups. DMFI was chosen as the primary endpoint for this analysis as this was the only endpoint prespecified in the MINDACT protocol that included only breast cancer-related events. However, DMFI does not include other breast cancer-related events that are important for the evaluation of ET, which is a limitation. Therefore, the incidence of locoregional recurrences and contralateral breast cancers was evaluated separately, and included in the exploratory BCFI composite endpoint, to assess all breast cancer-related events and the impact of ET. As these were post hoc analyses, we had limited power, but the number of patients in our study is larger than in the historic cohorts reporting on populations who did not receive AST and represents outcomes for a cohort treated according to contemporary guidelines.

In conclusion, with 8.7 years median follow-up, a subgroup of patients with stage I ER+/HER2–, low-risk breast cancer who received no AST had a good 8-year DMFI rate, although a slightly better outcome and lower rates of locoregional recurrence and contralateral breast cancer were observed in patients who received ET. Considering the natural history of ER+/HER2– breast cancer and the projected long-term survival effects of ET, the observed effect of adjuvant ET as well the side-effects should be discussed with patients even at a very low risk of distant metastasis.

Table 1. Patient and tumor characteristic for patients who received no AST and the matched[#] group of patients who received only ET.

	No AST (n=509) N (%)	**Only ET** (n=509) N (%)	**p-value[$]**
Age (years)[#]			
<35	4 (1%)	8 (2%)	0.053
35-50	120 (24%)	116 (23%)	
50-70	385 (76%)	379 (75%)	
≥70	0	6 (1%)	
Genomic risk[#]			
Low risk	484 (95%)	484 (95%)	
High risk	25 (5%)	25 (5%)	
Clinical risk			
Low risk	508 (99.8%)	504 (99%)	0.2
High risk	1 (0.2%)	5 (1%)	
Risk category*			
C-Low/G-Low	483 (95%)	479 (94%)	0.3
C-Low/G-High	25 (5%)	25 (5%)	
C-High/G-Low	1 (0.2%)	5 (1%)	
Type of breast surgery			
Breast conserving surgery	424 (83%)	474 (93%)	<0.0001
Mastectomy	85 (17%)	35 (7%)	
Type of axillary surgery			
ALND	25 (5%)	70 (14%)	<0.0001
SLNB	484 (95%)	439 (86%)	
Tumor size			
≤1 cm	219 (43%)	131 (26%)	<0.0001
1-2 cm	290 (57%)	378 (74%)	
Lymph node status			
Negative	509 (100%)	509 (100%)	
Tumor type			
Ductal	414 (82%)	423 (83%)	0.5
Lobular	46 (9%)	50 (10%)	
Mixed	32 (6%)	21 (4%)	
Other	17 (3%)	15 (3%)	
Tumor grade[#]			
1	329 (65%)	329 (65%)	
2	172 (34%)	172 (34%)	
3	7 (1%)	7 (1%)	
Undefined	1 (0.2%)	1 (0.2%)	

	No AST (n=509) N (%)	**Only ET** (n=509) N (%)	**p-value[$]**
ER status			
Positive	509 (100%)	509 (100%)	
PR status			
Positive	423 (83%)	457 (90%)	0.003
Negative	80 (16%)	49 (10%)	
Unknown	6 (1%)	3 (0.6%)	
HER2 status			
Negative	509 (100%)	509 (100%)	
Country			<0.0001
Netherlands	443 (87%)	62 (12%)	
Other¥	66 (13%)	447 (88%)	
Planned duration of ET			
7 years		264 (52%)	
5 years		245 (48%)	

All patients had tumors that were HR+/HER2-, ≤2cm and were lymph node negative.

[#]Patients were matched based on genomic risk (low risk vs high risk), age (≤50 vs >50) and tumor grade, comparisons of proportions were not applicable.

[$]Chi-square test or Fisher's Exact Test. *Risk category based on clinical and genomic risk. [¥]Other countries included Belgium, France, Germany, Italy, Slovenia and Spain. Merged to maintain patient anonymity.

ALND= axillary lymph node dissection; AST= adjuvant systemic treatment; C-High= clinical high risk; C-Low= clinical low risk; ER= estrogen receptor; ET= endocrine therapy; G-High= genomic high risk; G-Low= genomic low risk; HER2= Human Epidermal growth factor Receptor 2; PR= progesterone receptor; SLNB= sentinel lymph node biopsy

References

1. Curigliano G, Burstein HJ, Winer EP et al. De-escalating and escalating treatments for early-stage breast cancer: The St. Gallen International Expert Consensus Conference on the Primary Therapy of Early Breast Cancer 2017. *Ann Oncol* 2017; **28**(8):1700–12.
2. Burstein HJ, Curigliano G, Loibl S et al. Estimating the benefits of therapy for early-stage breast cancer: The St. Gallen International Consensus Guidelines for the primary therapy of early breast cancer 2019. *Ann Oncol* 2019; **30**(10): 1541–57.
3. Cardoso F, van't Veer LJ, Bogaerts J et al. 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. *N Engl J Med* 2016; **375**(8): 717–29.
4. Piccart M, van 't Veer LJ, Poncet C et al. 70-gene signature as an aid for treatment decisions in early breast cancer: updated results of the phase 3 randomised MINDACT trial with an exploratory analysis by age. *Lancet Oncol* 2021; **22**(4): 476–88.
5. Sparano JA, Gray RJ, Makower DF et al. Adjuvant chemotherapy guided by a 21-gene expression assay in breast cancer. *N Engl J Med* 2018; **379**(2): 111–21.
6. Andre F, Ismaila N, Henry NL et al. Use of biomarkers to guide decisions on adjuvant systemic therapy for women with early-stage invasive breast cancer: ASCO clinical practice guideline update-integration of results from TAILORx. *J Clin Oncol* 2019; **37**(22): 1956–64.
7. Cardoso F, Kyriakides S, Ohno S et al. Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2019; **30**(8):1194–220.
8. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet* 2005; **365**(9472): 1687–717.
9. Chlebowski RT, Kim J, Haque R. Adherence to endocrine therapy in breast cancer adjuvant and prevention settings. *Cancer Prev Res* 2014; **7**(4): 378–87.
10. Pistilli B, Paci A, Ferreira AR et al. Serum detection of nonadherence to adjuvant tamoxifen and breast cancer recurrence risk. *J Clin Oncol* 2020; **38**(24): 2762–72.
11. Condorelli R, Vaz-Luis I. Managing side effects in adjuvant endocrine therapy for breast cancer. *Expert Rev Anticancer Ther* 2018; **18**(11): 1101–12.
12. Wishart GC, Azzato EM, Greenberg DC et al. PREDICT: A new UK prognostic model that predicts survival following surgery for invasive breast cancer. *Breast Cancer Res* 2010; **12**(1): R1.
13. Predict Breast. [https://breast.predict.nhs.uk/].
14. IKNL. Borstkanker Landelijke Richtlijn, versie 2.0. 2020.
15. Guidance for the management of early breast cancer. Recommendations and practice points, 2020.
16. Early and locally advanced breast cancer: diagnosis and management NICE guideline, 2018.
17. Lurie RH, Anderson BO, Abraham J et al. NCCN Guidelines Version 5.2020 Breast Cancer, 2020.
18. Burstein HJ, Prestrud AA, Seidenfeld J et al. American Society of Clinical Oncology clinical practice guideline: Update on adjuvant endocrine therapy for women with hormone receptor-positive breast cancer. *J Clin Oncol* 2010; **28**(23): 3784–96.
19. Austin PC. The use of propensity score methods with survival or time-to-event outcomes: Reporting measures of effect similar to those used in randomized experiments. *Stat Med* 2014; **33**(7): 1242–58.
20. Clarke M, Collins R, Darby S et al. Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials. *Lancet* 2005; **366**(9503): 2087–106.

21. Austin PC, Lee DS, Fine JP. Introduction to the Analysis of Survival Data in the Presence of Competing Risks. *Circulation* 2016; **133**(6): 601–9.
22. Scrucca L, Santucci A, Aversa F. Competing risk analysis using R: An easy guide for clinicians. *Bone Marrow Transplant* 2007; **40**(4): 381–7.
23. Hudis CA, Barlow WE, Costantino JP et al. Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: The STEEP system. *J Clin Oncol* 2007; **25**(15): 2127–32.
24. Pan H, Gray R, Braybrooke J et al. 20-Year Risks of Breast-Cancer Recurrence After Stopping Endocrine Therapy At 5 Years. *N Engl J Med* 2017; **377**(19):1836–46.
25. Tang SC. Reducing the risk of distant metastases: A better end point in adjuvant aromatase inhibitor breast cancer trials? *Cancer Invest* 2008; **26**(5): 481–90.
26. Esserman LJ, Yau C, Thompson CK et al. Use of molecular tools to identify patients with indolent breast cancers with ultralow risk over 2 decades. *JAMA Oncol* 2017; **3**(11): 1503–10.
27. Ohnstad HO, Borgen E, Falk RS et al. Prognostic value of PAM50 and risk of recurrence score in patients with early-stage breast cancer with long-term follow-up. *Breast Cancer Res* 2017; **19**(1):1–12.
28. Pathmanathan N, Balleine RL, Jayasinghe UW et al. The prognostic value of Ki67 in systemically untreated patients with node-negative breast cancer. *J Clin Pathol* 2014; **67**(3): 222–8.
29. Sjöström M, Laura Chang S, Fishbane N et al. Comprehensive transcriptomic profiling identifies breast cancer patients who may be spared adjuvant systemic therapy. *Clin Cancer Res* 2020; **26**(1): 171–82.
30. Adra J, Lundstedt D, Killander F et al. Distribution of Locoregional Breast Cancer Recurrence in Relation to Postoperative Radiation Fields and Biological Subtypes. *Int J Radiat Oncol Biol Phys* 2019; **105**(2): 285–95.
31. Sjöström M, Lundstedt D, Hartman L et al. Response to radiotherapy after breast-conserving surgery in different breast cancer subtypes in the Swedish Breast Cancer Group 91 radiotherapy randomized clinical trial. *J Clin Oncol* 2017; **35**(28): 3222–9.
32. Ignatov A, Eggemann H, Burger E, Ignatov T. Patterns of breast cancer relapse in accordance to biological subtype. *J Cancer Res Clin Oncol* 2018; **144**(7): 1347–55.
33. Arvold ND, Taghian AG, Niemierko A et al. Age, breast cancer subtype approximation, and local recurrence after breast-conserving therapy. *J Clin Oncol* 2011; **29**(29):3885–91.
34. Darby S, McGale P, Correa C et al. Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: Meta-analysis of individual patient data for 10 801 women in 17 randomised trials. *Lancet* 2011; **378**(9804): 1707–16.
35. Giardiello D, Steyerberg EW, Hauptmann M et al. Prediction and clinical utility of a contralateral breast cancer risk model. *Breast Cancer Res* 2019; **21**(1): 144.
36. Kramer I, Schaapveld M, Oldenburg HSA et al. The Influence of Adjuvant Systemic Regimens on Contralateral Breast Cancer Risk and Receptor Subtype. *J Natl Cancer Inst* 2019; **111**(7): 709–18.
37. Davies C, Godwin J, R Gray R et al. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: Patient-level meta-analysis of randomised trials. *Lancet* 2011; **378**(9793): 771–84.
38. Aalders KC, Van Bommel ACM, Van Dalen T et al. Contemporary risks of local and regional recurrence and contralateral breast cancer in patients treated for primary breast cancer. *Eur J Cancer* 2016; **63**: 118–26.

39. Witteveen A, Kwast ABG, Sonke GS et al. Survival after locoregional recurrence or second primary breast cancer: Impact of the disease-free interval. *PLoS One* 2015; **10**(4): 1–12.
40. Neuman HB, Schumacher JR, Francescatti AB et al. Risk of synchronous distant recurrence at time of locoregional recurrence in patients with stage II and III breast cancer (AFT-01). *J Clin Oncol* 2018; **36**(10): 975–80.
41. Thornton AA, Madlensky L, Flatt SW et al. The impact of a second breast cancer diagnosis on health related quality of life. *Breast Cancer Res Treat* 2005; **92**(1): 25–33.
42. Ferreira AR, Di Meglio A, Pistilli B et al. Differential impact of endocrine therapy and chemotherapy on quality of life of breast cancer survivors: a prospective patient-reported outcomes analysis. *Ann Oncol* 2019; **30**(11): 1784–95.
43. Ganz PA, Petersen L, Bower JE, Crespi CM. Impact of adjuvant endocrine therapy on quality of life and symptoms: Observational data over 12 months from the mind-body study. *J Clin Oncol* 2016; **34**(8): 816–24.

Supplementary materials

Supplementary Methods

Additional detail on the definition of endpoints:

The definition of all endpoints are as described in the MINDACT protocol[1,2]:

Distant metastasis free interval (DMFI) was the primary endpoint for this study, defined as the time until first distant metastasis or breast cancer related death (deaths due to progressive disease or treatment toxicity). Patients with unknown cause of death were also considered to have had an DMFI event. Patients with another cause of death (cardiovascular disease, other chronic disease, second primary cancer or other) were censored at their death date.

Distant metastasis free survival (DMFS) was defined as the time until first distant metastatic recurrence or death from any cause. Contralateral breast cancer and secondary cancers were not taken into account as events.

Overall survival (OS) was defined as the time until death from any cause.

Disease-free survival (DFS) was defined as the time until first disease progression (loco-regional and distant recurrences, ipsilateral or contralateral invasive breast cancer or ductal carcinoma in situ and invasive second primary cancer (non-breast)) or death from any cause.

Breast cancer-specific survival (BCSS) was defined as the time until breast cancer related death (deaths due to progressive disease or treatment toxicity).

An exploratory endpoint was compiled specifically for this study, and was not included in the MINDACT protocol. For this endpoint the STEEP criteria were followed[3]:

Breast cancer free interval (BCFI) was defined as the time to first distant metastasis, locoregional recurrence, contralateral breast cancer or breast cancer related death.

Data for patients who had no event at the cutoff date were censored at the time of last disease assessment for distant metastasis free interval, distant metastasis free survival, disease-free survival and breast cancer free interval and at the last follow-up date for overall survival and breast cancer- specific survival. Patients who died more than 2 years after their last disease assessment were censored at the date of last disease assessment.

References:

1. Cardoso F, van't Veer LJ, Bogaerts J, et al. 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. *N Engl J Med*. 2016;375(8):717-729. doi:10.1056/NEJMoa1602253
2. Piccart M, van 't Veer LJ, Poncet C, et al. 70-gene signature as an aid for treatment decisions in early breast cancer: updated results of the phase 3 randomised MINDACT trial with an exploratory analysis by age. *Lancet Oncol*. 2021;22(4):476-488. doi:10.1016/S1470-2045(21)00007-3
3. Hudis CA, Barlow WE, Costantino JP, et al. Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials: The STEEP system. *J Clin Oncol*. 2007;25(15):2127-2132. doi:10.1200/JCO.2006.10.3523

Additional detail on propensity score matching:

We conducted 1:1 nearest neighbor matching without replacement on a propensity score estimated using a logistic regression of the treatment on the covariates, using "exact" matching of specified variables, and with random order of matching. Specifically, the propensity score was estimated using a generalized linear model with a logit link, with binary treatment (no adjuvant systemic treatment vs. endocrine therapy) as the outcome, and age (≤50 vs >50), genomic risk according to the 70-gene signature (low risk vs high risk) and tumor grade as covariates. Exact matching on the same covariates listed above was used in combination with propensity score matching to ensure that doubly robust estimators were derived, which assists in model efficiency. As we pre-selected patients with tumors ≤2cm, we did not further match on exact tumor size to increase the number of possible matches. Patients were not matched on country as this resulted in a 50% reduction of matched patients, and survival was not different between countries. Exact matching and propensity score calculation was performed using the "MatchIt" package (Ho, Imai, King, & Stuart, 2011) in R.

Table S1. Summary of all endpoints for no adjuvant systemic treatment and only endocrine therapy groups

Endpoint	AST received	Patients (N)	Events (O)	% at 5 years (95% CI)	% at 8 years (95% CI)	HR only ET vs no AST (95% CI)
	No AST	509	29	97.9% (96.6-99.2)	94.8% (92.7-96.9)	0.56 (0.30-1.03)
	Only ET	509	16	98.4% (97.2-99.5)	97.3% (95.8-98.8)	
DMFS	No AST	509	46	96.8% (95.3-98.4)	92.5% (90.1-95.0)	0.73 (0.46-1.13)
	Only ET	509	33	97.0% (95.4-98.5)	94.8% (92.7-96.8)	
OS	No AST	509	36	97.9% (96.7-99.2)	95.4% (93.5-97.4)	0.86 (0.53-1.41)
	Only ET	509	28	97.8% (96.5-99.1)	95.6% (93.8-97.5)	
DFS	No AST	509	114	88.4% (85.6-91.3)	80.3% (76.7-84.1)	0.54 (0.40-0.73)
	Only ET	509	65	93.9% (91.8-96.0)	89.3% (86.4-92.2)	
BCSS	No AST	509	10	99.8% (99.4-100)	99.1% (98.2-100)	0.58 (0.20-1.69)
	Only ET	509	5	99.6% (99.0-100)	99.4% (98.6-100)	
BCFI	No AST	509	77	92.7% (90.4-95.0)	86.5% (83.3-89.8)	0.37 (0.24-0.57)
	Only ET	509	30	97.7% (96.4-99.1)	94.8% (92.7-97.0)	

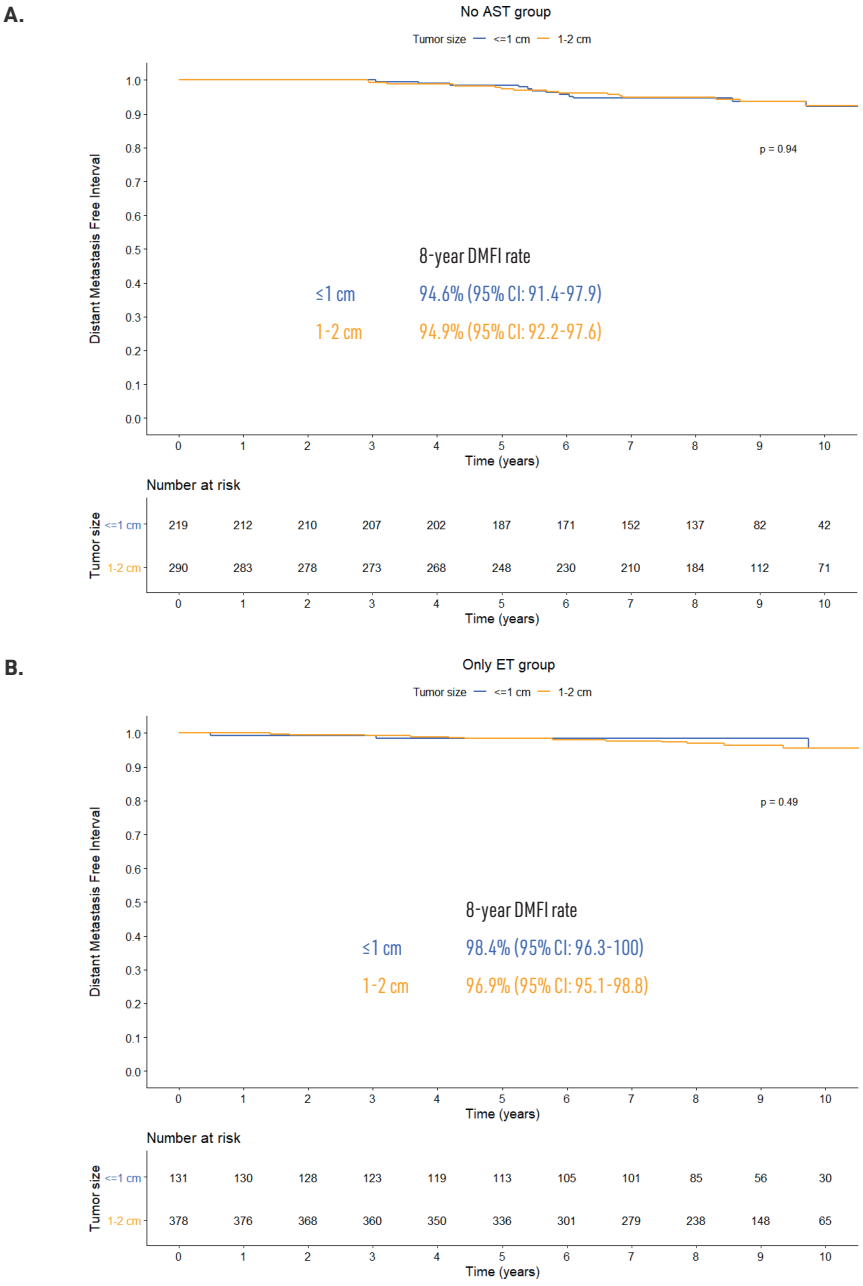
AST: adjuvant systemic treatment; BCFI: breast cancer free interval; BCSS: breast cancer-specific survival; DFS: disease-free survival; DMFI: distant metastasis free interval; DMFS: distant metastasis free survival; ET: endocrine therapy; HR: hazard ratio; OS: overall survival.

Table S2. Estimated OS and BCSS and added benefit of ET according to PREDICT for the no adjuvant systemic treatment group

Endpoint		Survival estimates PREDICT, surgery only (95% CI)	Added benefit ET PREDICT (95% CI)
OS	5 years	95.7% (95.5-95.8)	0.46% (0.44-0.48)
	8 years*	91.6% (91.3-91.9)	0.86% (0.82-0.91)
	10 years	88.5% (88.1-88.9)	1.12% (1.07-1.17)
	15 years	80.0% (79.3-80.8)	1.65% (1.57-1.73)
BCSS	5 years	98.5% (98.5-98.6)	0.46% (0.44-0.48)
	8 years*	97.2% (97.0-97.3)	0.86% (0.82-0.91)
	10 years	96.2% (96.0-96.4)	1.12% (1.07-1.17)
	15 years	94.1% (93.8-94.3)	1.65% (1.57-1.73)

*The estimates are also provided at 8 years as this was the chosen endpoint for survival analyses in this study. BCSS= breast cancer-specific survival; ET= endocrine therapy; OS=overall survival

Figure S1. Kaplan Meier analysis of distant metastasis free interval for **a.** no AST group and **b.** only ET group, stratified by tumor size



Median follow-up was 8.6 years for this cohort. Estimates are shown at 8 years because at that time point there were still a sufficient number of patients at risk. Kaplan-Meier curves are displayed until 10 years of follow-up due to the high level of censoring beyond this time point.


8

Chapter 8

AI-based prevention of interval cancers in a population-based breast cancer program

Danalyn Byng [1], Benjamin Strauch [1], Linda Gnas [2], Christian Leibig [1], Oliver Stephan [2], Stefan Bunk [1], Gerold Hecht [2]

[1]Vara, Berlin, Germany.

[2]Mammography Reference Center North (Referenzzentrum Nord), Oldenburg, Germany.

Eur J Radiology 152**,** 110321 (2022). https://doi.org/10.1016/j.ejrad.2022.110321

Highlights

- Interval cancer rate is an important quality parameter for breast cancer screening.
- AI can accurately detect and localize interval cancers on prior screening exams.
- More breast cancers can be detected at screening without inducing higher recall.
- Some histopathological characteristics may influence likelihood of detection.

Abstract

Purpose

To demonstrate that artificial intelligence (AI) can detect and correctly localise retrospectively visible cancers that were missed and diagnosed as interval cancers (false negative (FN) and minimal signs (MS) interval cancers), and to characterise AI performance on non-visible occult and true interval cancers.

Method

Prior screening mammograms from N = 2,396 women diagnosed with interval breast cancer between March 2006 and May 2018 in north-western Germany were analysed with an AI system, producing a model score for all studies. All included studies previously underwent independent radiological review at a mammography reference centre to confirm interval cancer classification. Model score distributions were visualised with histograms. We computed the proportion and accompanying 95% confidence intervals (CI) of retrospectively visible and true interval cancers detected and correctly localised by AI at different operating points representing recall rates < 3%. Clinicopathological characteristics of retrospectively visible cancers detected by AI and not were compared using the Chi-squared test and binary logistic regression.

Results

Following radiological review, 15.6% of the interval cancer cases were categorised as FN, 19.5% MS, 11.4% occult, and 53.4% true interval cancers. At an operating point of 99.0% specificity, AI could detect and correctly localise 27.5% (95% CI: 23.3–32.3%), and 12.2% (95% CI: 9.5–15.5%) of the FN and MS cases on the prior mammogram, respectively. 228 of these retrospectively visible cases were advanced/metastatic at diagnosis; 21.1% (95% CI: 16.3–26.8%) were found by AI on the screening mammogram. Increased likelihood of detection of retrospectively

visible cancers with AI was observed for lower-grade carcinomas and those with involved lymph nodes at diagnosis. Among true interval cancers, AI could detect and correctly localise in the screening mammogram where subsequent malignancies would appear in 2.8% (95% CI: 2.0–3.9%) of cases.

Conclusions

AI can support radiologists by detecting a greater number of carcinomas, subsequently decreasing the interval cancer rate and the number of advanced and metastatic cancers.

Introduction

Background

Among women participating in breast cancer screening, a proportion of breast cancers will be clinically detected and diagnosed in the interval between screening rounds due to the onset of clinical symptoms following a prior normal screening round.[1] Some of these cancers are retrospectively visible on the prior screening images. Approximately 20–25% of interval cancers are classified as minimal signs, and a further 20–25% are considered missed as false negatives.[1-5] The European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis (EUREF) have objective criteria for defining minimal signs and false negative interval cancers.[1] The latter are clearly visible abnormalities warranting assessment, whereas the former are a possible subtle abnormality on screening mammography, which may or may not warrant further assessment. Other interval cancers are occult, not visible on mammography at screening or diagnosis but are symptomatic and possibly visible with other imaging modalities (e.g. ultrasound) at diagnosis. Finally, true interval cancers are not visible at all on the prior normal mammogram. These make up the majority of cases clinically diagnosed between screening intervals.[1,2]

Interval cancers are generally understood to have poorer prognostic characteristics and are related to poorer breast cancer specific mortality outcomes compared to screen-detected cancers.[6-8] As such, the interval cancer rate is an important surrogate parameter for evaluating the quality and effectiveness of organised population-based screening programs, as outlined by EUREF.[1] Using information on interval cancers, program sensitivity can be calculated as the proportion of screen-detected breast cancers, among all cancers occurring in screening participants within a defined period of time.[1] The denominator effectively includes all interval cancers in addition to screen-detected cancers.

The introduction of artificial intelligence (AI) can possibly improve program sensitivity, but this requires an evaluation of its performance on interval cancers to determine if AI can detect cancers that would otherwise be overlooked or missed by radiologists during screening. To date, performance of different AI systems on datasets composed exclusively of interval cancers has been evaluated using small retrospective series in Germany (N = 29) and Sweden (N = 429), and more recently a larger cohort from the Netherlands (N = 666).[9-11] In some cases, performance was evaluated using thresholds corresponding to recall rates from 5% to 10%, exceeding what is acceptable in clinical practice under European guidelines.[1]

Given the poor prognosis of many interval cancers at diagnosis, understanding which breast cancer histopathological characteristics may lend themselves to increased likelihood of detection with AI at screening remains an important undertaking. Therefore, evaluation of AI performance on a larger cohort of interval cancer cases with detailed clinicopathological information, using thresholds relevant to clinical practice is warranted.

Objectives

Using a representative dataset of interval cancers diagnosed between 2006 and 2018, the primary objective of the study is to understand the performance of AI on the screening studies prior to the interval cancer diagnosis. We assessed whether AI can detect retrospectively visible cancers that are clinically diagnosed in the interval between screening rounds (false negatives and minimal signs), showing a potential of AI to prevent interval cancers before they occur. The analysis included characterising and comparing the cancers that are detected by AI and those not, as well as verifying whether the AI localised the correct region within the false negative and minimal signs cases. Furthermore, we characterised AI performance on occult and true interval cancers, to determine if an increased model score (i.e. indication of suspiciousness) can be produced for some studies, and to determine if AI could accurately indicate where on the screening image a true interval breast cancer would later develop.

Materials and methods

Data sources

An anonymised interval cancer dataset was provided by the Mammography Reference Centre North (Referenzzentrum Mammographie Nord) in Oldenburg, Germany for this study. The dataset consists of mammography studies from women attending biennial breast cancer screening at 7 screening units in the German federal state of Lower Saxony between January 2006 to February 2017, who had an interval cancer diagnosis between March 2006 and May 2018. The dataset contains images from the screening examination prior to the clinical diagnosis of the interval cancer, corresponding diagnostic images, and meta-data on clinicopathological characteristics of the diagnosed cancer. Data on interval cancers from screening units and the local cancer registry were transferred to the reference centre for the determination of interval cancer rates, in accordance with currently valid legal provisions and the procedure specified in the official Cancer Registration Act (Landeskrebsregistergesetz) in Lower

Saxony. Ethics approval for this study was granted by the ethics committee of the Medical Association of Lower Saxony under Bo/60/2021.

Interval cancer radiological review

The Mammography Reference Centre North is one of 5 regional reference centres in Germany providing radiological review as part of the quality assurance requirements for the national mammography screening program.[12] Under the German cancer screening guidelines, evaluation of the mammography screening program requires determination of the interval cancer rate. Interval cancers are defined as cancers clinically diagnosed in the 24 month period following a normal breast screening, before the next scheduled screening round.[12] All screening units are linked to the local cancer registry to collect information about interval cancer cases. These are further classified according to EUREF guidelines into the following categories: false negative, minimal signs, occult, and true interval.[1] Additional criteria are applied to true interval cancers. While some are completely non-visible at screening, others may be considered stable breast lesions, benign on previous biopsy, without observable changes over a prolonged period. All interval cancer cases without available mammograms at the point of diagnosis are classified as "unclassifiable".

All interval cancer cases, including accompanying screening and diagnostic mammograms and documentation required for classification, are sent from the screening unit to a regional reference centre, where a random sample of 10 cases per year are selected for radiological review. The review is carried out as an individual evaluation by at least three doctors, including the head of reference centre, the deputy head of the reference centre, and 1–2 qualified radiologists. Review is a two-stage process: (1) blinded diagnosis, a preliminary assessment based on screening mammograms and any prior images; followed by a (2) findings-oriented diagnosis, whereby the final assessment and classification is made using the diagnostic mammograms and documentation, in addition to screening documentation. The results of the review, including any deviations from the original classification, are assessed annually in a professional discussion between the screening unit and the reference centre.

Participants

We included studies from the Mammography Reference Centre North's radiological review of women with an interval cancer diagnosis who were participating in biennial breast cancer screening at one of the screening units in the catchment area, operating under the German national breast cancer screening program. Women were ages 50–69 at the time of screening, with an interval cancer diagnosis confirmed by biopsy and reported in the cancer registry. Interval cancer review and classification must have been confirmed by the reference centre, with a diagnosis occurring within

24 months after a normal screening mammogram and before the next scheduled mammogram. Cases were excluded if they had not been assessed according to EUREF and did not receive an accompanying classification of false negative, minimal signs, occult, or true interval cancer. We further excluded cases with diagnosis dates within 3 months of screening, or those with incomplete screening records, missing prior screening images, or without an interval cancer diagnosis date. Only cases for which all imaging studies from screening are available, comprising the four standard views (bilateral craniocaudal and mediolateral oblique) were included.

Variables

To characterise the interval cancers according to clinicopathological characteristics, data on age at screening, classification of breast density based on the mammographic appearance of the tissue according to the American College of Radiology Breast Imaging Reporting and Data System (ACR BI-RADS) Atlas,[13] hormone receptor (HR) status (including oestrogen and progesterone receptor status), HER2neu status, tumour size, lymph node (LN) status, distant metastasis at diagnosis, grade and Ki-67 proliferation rate were extracted from the cancer registry datafiles by the reference centre and provided for this study. Information on laterality and subsite descriptors of the position of the lesion in the breast were also included.

Image analysis

All available full-field digital mammography images obtained from the prior screening round were analysed with a commercially available AI system (Vara 1.0.7). For each study, the model outputs a floating point exam-level model score ranging from 0.0 to 1.0, with higher scores corresponding to increased suspiciousness for cancer. For AI to make a distinct decision about whether a case is positive (should be considered for recall) or negative (no follow-up), a threshold to split positive and negative classes at a given model score is necessary. Accordingly, a cancer is considered detected when its accompanying prediction score is above the threshold. To determine a threshold, we used a separate threshold setting dataset of 20,000 follow-up-proven negative cases from the German screening system. The threshold was chosen to deliver a specificity of 99.0% of the AI system on this separate dataset. A specificity of 99% was chosen deliberately to configure the algorithm to have a much smaller false positive rate (1%) than a typical European double-reader screening system (~4%),[12,14] and to illustrate the potential of AI to already prevent a substantial proportion of interval cancers at this operating specificity. Further details of the threshold setting procedure are described in the Supplementary Appendix.

The AI also produces a single corresponding localization marker per exam (Fig. 1). To determine if the AI system detected a retrospectively visible interval cancer and

correctly localised it within the image, all false negative and minimal signs cases detected by the AI at 99.0% specificity underwent localization review by a radiologist with 10 years of experience reading breast images. Though true interval cancers may have no lesion or tissue change visible to the human eye at screening, the same exercise was undertaken for the true interval cancer cases which yielded scores above the threshold to assess the potential of AI to identify the area of the breast in which a true interval cancer later develops. The radiologist was provided with the screening study, the localization markers by the AI in the screening study, the diagnostic mammograms from diagnosis and the localization information (laterality and subsite descriptors of the position of the lesion in the breast) provided by the reference centre (Fig. 1), and was asked to assess whether the AI localised the correct region.

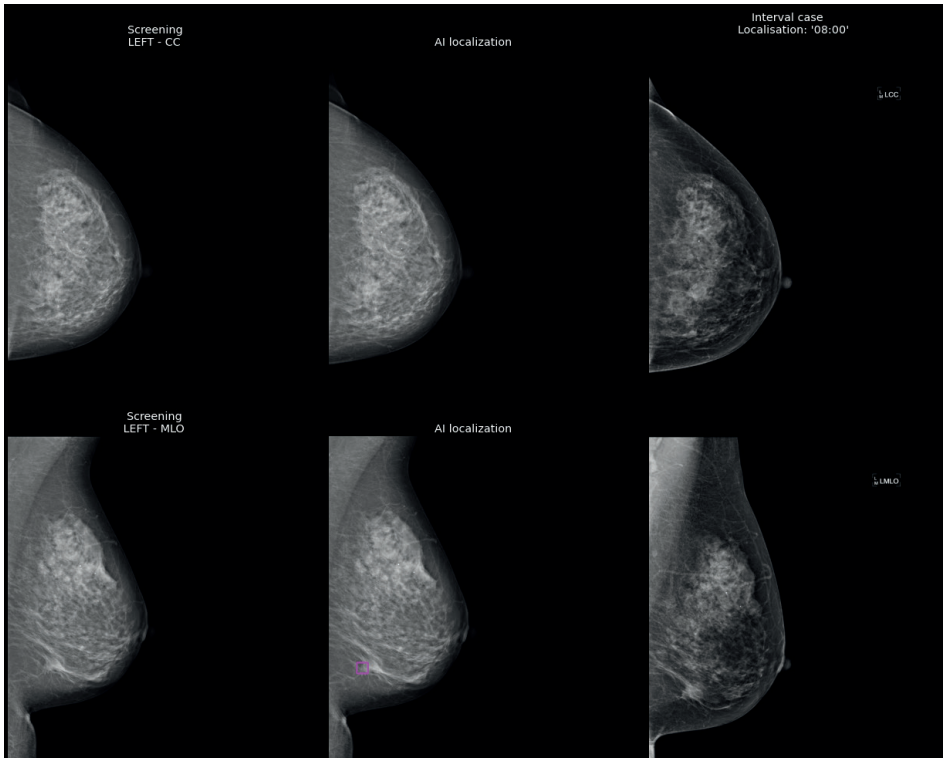


Figure 1. Localisation procedure used to verify correct localisations for cases above the threshold. Images presented to the radiologist for assessing the correct localisation. The leftmost panel shows the original screening images, only showing the laterality where the cancer was detected later. The middle panel shows the screening images overlaid with the AI's localisation in the form of one rectangle marker per case. The rightmost panel shows the diagnostic images from the point of diagnosis. The localisation of the cancer (written above the diagnostic images) was determined blinded beforehand, during the blinded independent radiological review in the reference centre. This localisation happened before and independently from the assessment by AI. During our localisation check, the radiologist's task was to assess whether the AI's localisation and the localisation from the reference centre correspond to the same lesion.

Statistical methods

To illustrate the scope of different algorithm operating points, we assess the algorithm's sensitivity over a set of high specificities (>90%). Performance on sensitivity is stratified according to interval cancer classification (false negative, minimal signs, occult, true interval).

We report the proportions of retrospectively visible (false negative and minimal signs) interval cancer studies identified as suspicious by the algorithm at the different exemplary specificity values (97.0–99.5%). The proportion of false negative, minimal signs, and true interval cases correctly localised at the exemplary operating point (99.0%) is also reported. As advanced breast cancer is an important endpoint for evaluating breast cancer screening effectiveness, we separately computed the proportion of retrospectively visible cancers detected by AI that were locally advanced (T2N1M0; any T3 and T4; any N2N3M0) and metastatic (M1) at diagnosis. Wilson score intervals are used to compute accompanying 95% confidence intervals (CI) for the reported proportions.

We compared the clinicopathological characteristics of retrospectively visible cancers detected by AI and those not using the Chi-square test, and univariate and multivariable binary logistic regression. The models included covariates for age at screening, ACR breast density classification, HR status, HER2neu status, tumour size, LN status, distant metastasis at diagnosis, grade, and Ki-67.

To assess and compare the model score distributions for false negatives, minimal signs, occult and true interval cancers, we discretised scores into ten equal-width bins and visualised their distributions with relative frequency histograms. Model score distributions for cancers not visible on the prior mammogram (occult and true interval) were also compared to score distributions derived from the 20,000 negative cases used for threshold setting.

Results

Cohort summary

Data from N = 6,261 women diagnosed with interval breast cancers were provided (Fig. 2). After exclusion based on the prespecified diagnosis interval and information missing at random, the final cohort resulted in N = 2,396 women with interval cancers diagnosed 3–24 months after their biennial screening appointment. Interval cancer diagnoses occurred between October 2006 and March 2018. Radiological review conducted at the Mammography Reference Centre North classified the diagnoses as

follows: 374 (15.6%) were false negative, 468 (19.5%) minimal signs, 274 (11.4%) were occult, and 1,280 (53.4%) were true interval cancers. Patient and tumour-specific characteristics are listed in Table 1. Median age at screening was 60 years old (range 50–69). Based on the available information from the cancer registry, 4.4% of the cases were in situ lesions, 26.9% were locally advanced invasive carcinomas (≥stage IIB) or metastasized at diagnosis. Cancers were predominantly intermediate and high grade (II/III).

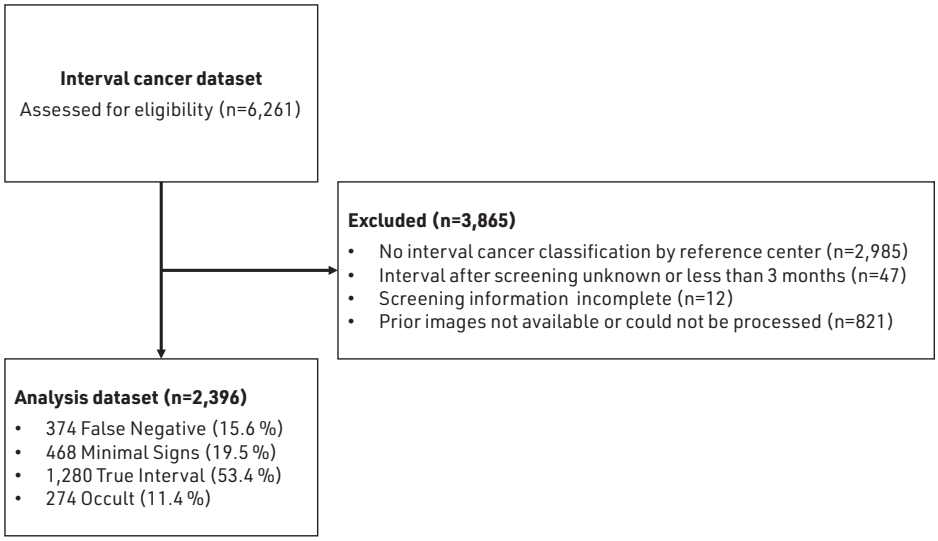


Figure 2. Flow diagram of initial number of interval cancer cases and those excluded for any reason.

Detection of interval cancers by AI

Performance of AI on detecting interval cancers across different possible specificity values are shown in Fig. 3. The algorithm performed best in detecting retrospectively visible cancers, i.e. false negative and minimal signs cases on the prior screening mammogram. A small proportion of occult and true interval cancers were flagged as suspicious by the system.

Detection and localisation of retrospectively visible cancers by AI

At an operating point of 99.0% specificity, AI could detect 30.2% (113/374; 95% CI: 25.8–35.0%) of the false negative cases, and 13.7% (64/468; 95% CI: 10.9–17.1%) of the minimal signs cases (Table 2). With decreasing specificity and corresponding increasing sensitivity, AI could detect larger proportions of retrospectively visible cancers.

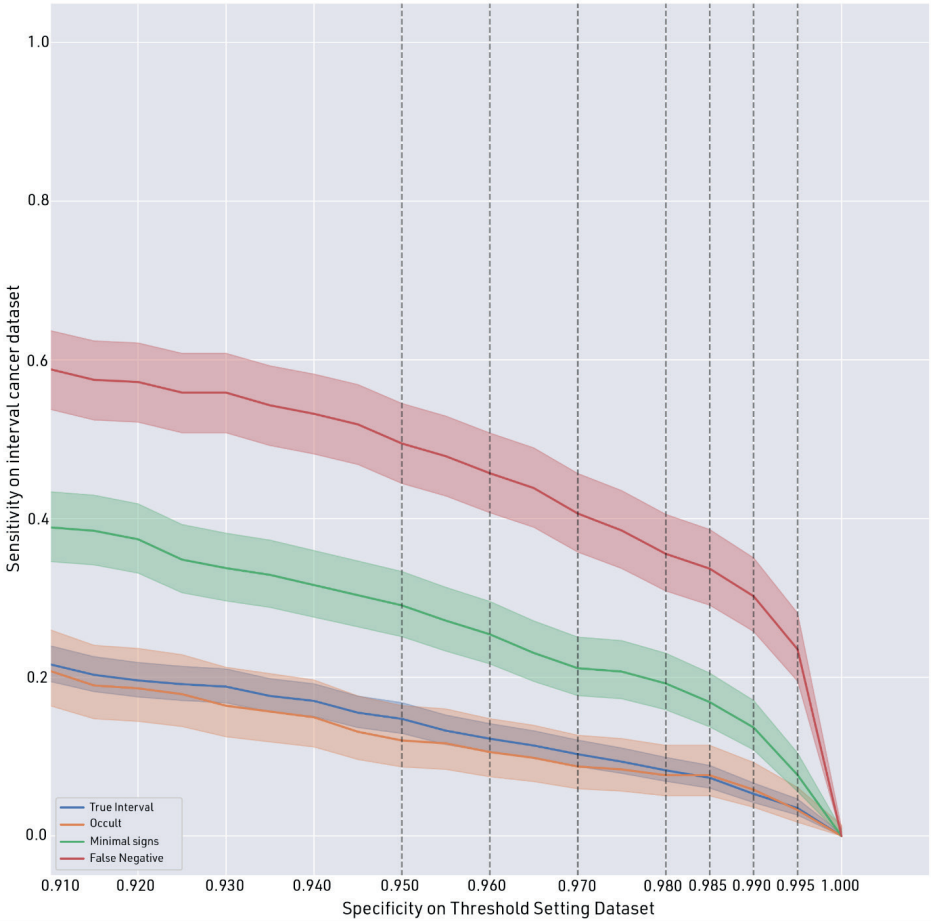


Figure 3. Performance of AI on interval cancers (N = 2,396). Performance of AI on detecting interval cancers by type. Different possible specificity values based on the threshold setting dataset are represented by vertical lines. Methods for threshold setting are described in the Supplementary Appendix.

AI demonstrated strong localization performance on most detected false negative and minimal signs cases, with 91.2% and 89.1% accuracy, respectively. This corresponds to a potential reduction of false negative and minimal signs interval cancers by 27.5% (103/374; 95% CI 23.3–32.3%) and 12.2% (57/468; 95% CI 9.5–15.5%), respectively. Among these cancers, 228/842 were locally advanced or metastatic at diagnosis. AI could detect and correctly localise 21.1% (48/228; 95% CI 16.3–26.6%) of these cases. Examples of false negative interval cancer cases with correct and incorrect AI localisations are provided in Fig. 4, Fig. 5.

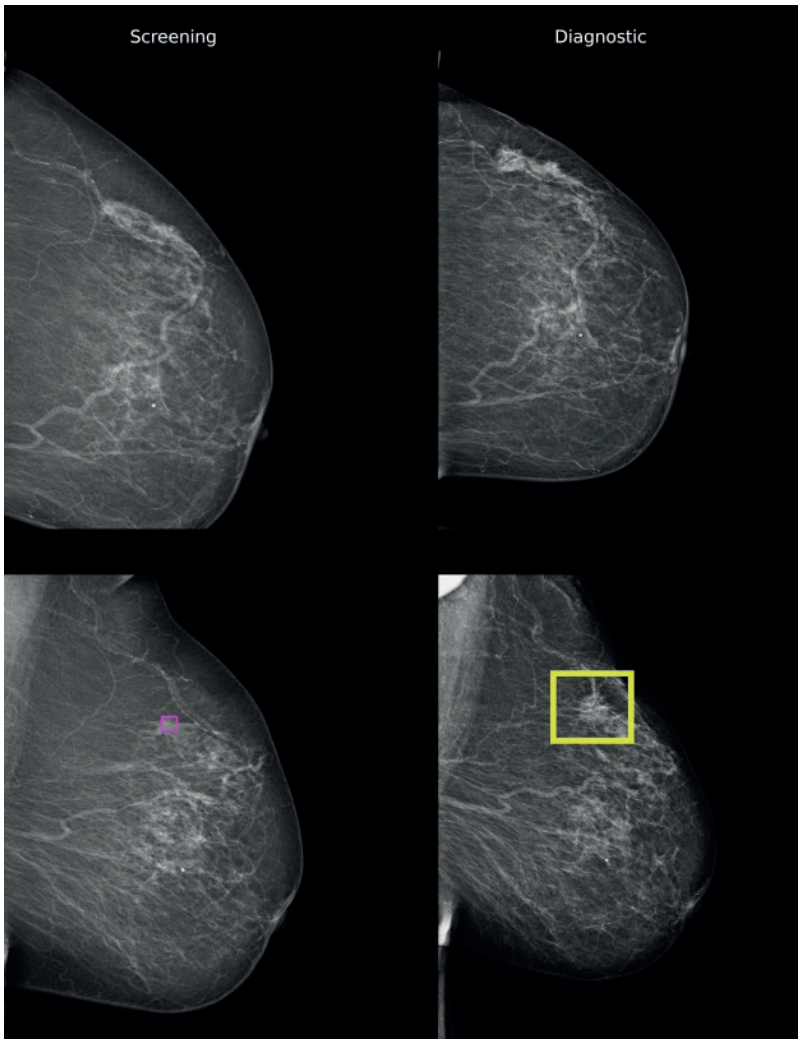


Figure 4. Example of a correctly localised false negative interval cancer on the screening mammography. False negative interval cancer detected by AI and correctly localised with AI marker in the left MLO view of the screening mammography. 68-years old at screening, diagnosed with an interval cancer 20-months post-screening. Grade 3, T2N2aM0 invasive ductal carcinoma.

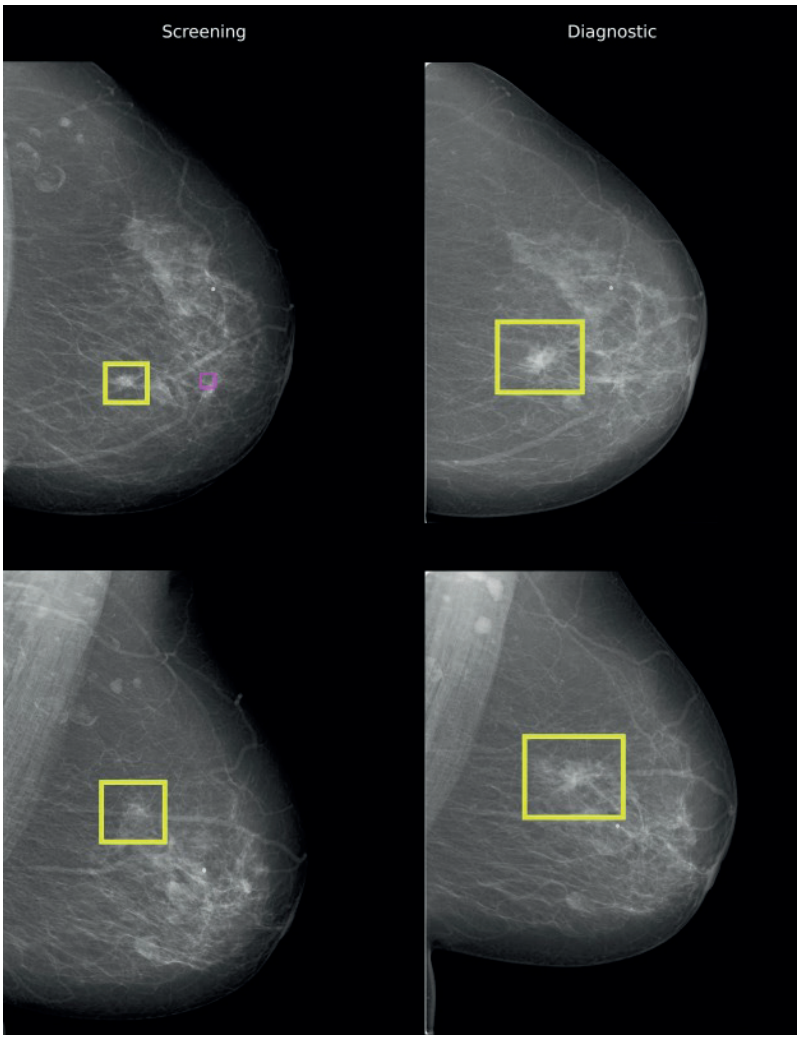


Figure 5. Example of an incorrectly localised false negative interval cancer on the screening mammography. False negative interval cancer detected by AI but incorrectly localised with AI marker on left CC view. 61-years old at screening, diagnosed with interval cancer 22-months post-screening. Grade 3, T1cN2aM0 invasive ductal carcinoma.

Correct localisation of true interval breast cancer on prior image

Upon the localisation check, the radiologist determined that AI could detect and correctly localise where the subsequent interval cancer would be diagnosed on the screening image with an accuracy of 63% (36 of 57 studies with heightened suspiciousness scores). This corresponds to 2.8% (36/1,280; 95% CI 2.0–3.9%) of all true interval studies in the cohort.

Factors associated with detection by AI

Across all clinicopathological characteristics of the retrospectively visible cancers, no statistically significant differences were observed between the cancers detected by AI and those not, with the exception of high grade tumours: compared to low grade, there was a lower likelihood of detecting high grade tumours with AI (OR 0.42, 95% CI: 0.21–0.85) (Table 3). When controlling for all tumour characteristics simultaneously in the multivariable analyses (Table 4), this was particularly pronounced among false negative cancers (OR 0.33, 95% CI: 0.12–0.88), and cancers diagnosed in the 12–24 month period after screening (OR 0.31, 95% CI: 0.12–0.78). There was a trend towards better detection with AI in older age groups ≥ 55, compared to ages 50–54. Increased likelihood of detection with AI was also observed for carcinomas with involved lymph nodes at diagnosis compared to lymph node negative carcinomas (LN1 vs. LN0, OR 1.67 [95% CI: 1.10–2.54]; LN2 + vs. LN0, OR 1.16 [95% CI: 0.63–2.06]).

Distribution of model scores by interval cancer type

Fig. 6 shows the distribution of model scores by interval cancer classification. False negative cases show a clear left-skewed distribution, an indication of a higher percentage of cases with high model scores. Occult and true interval cancers show a right-skewed distribution, indicating a higher percentage of lower model scores. Scores of minimal signs cases tend more towards a uniform distribution. Fig. 7 shows that while both true intervals as well as occult cases have no visual signs of malignancy (visible to the human eye on mammography) at the point of screening, they still tend to yield an increased model score compared to a background population of screening negatives.

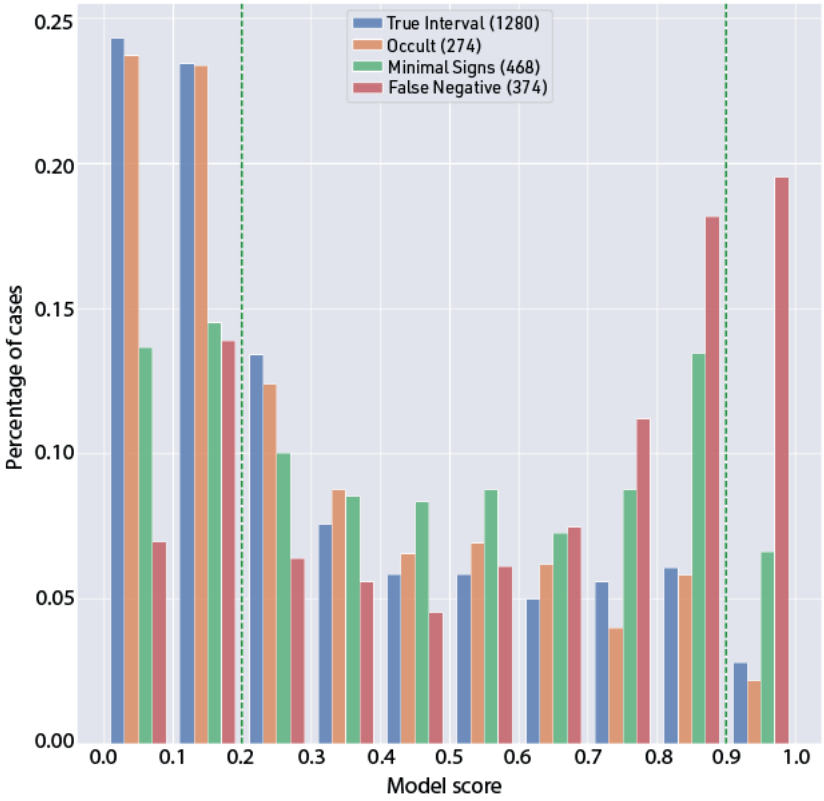


Figure 6. Distribution of model scores by interval cancer classification. Distribution of AI model scores derived from screening mammograms for true interval, occult, minimal signs, and false negative interval cancer cases. Higher scores correspond to increased suspiciousness for cancer. The vertical line at 0.2 represents a possible threshold to automatically classify a subset of non-visible cancers during radiologic review, potentially reducing manual classification workload for screening program quality assurance. The vertical line at 0.9 represents a possible threshold to provide an estimation of the underlying false negative rate among interval cancers during radiologic review, as one-fifth of false negatives score above this threshold.

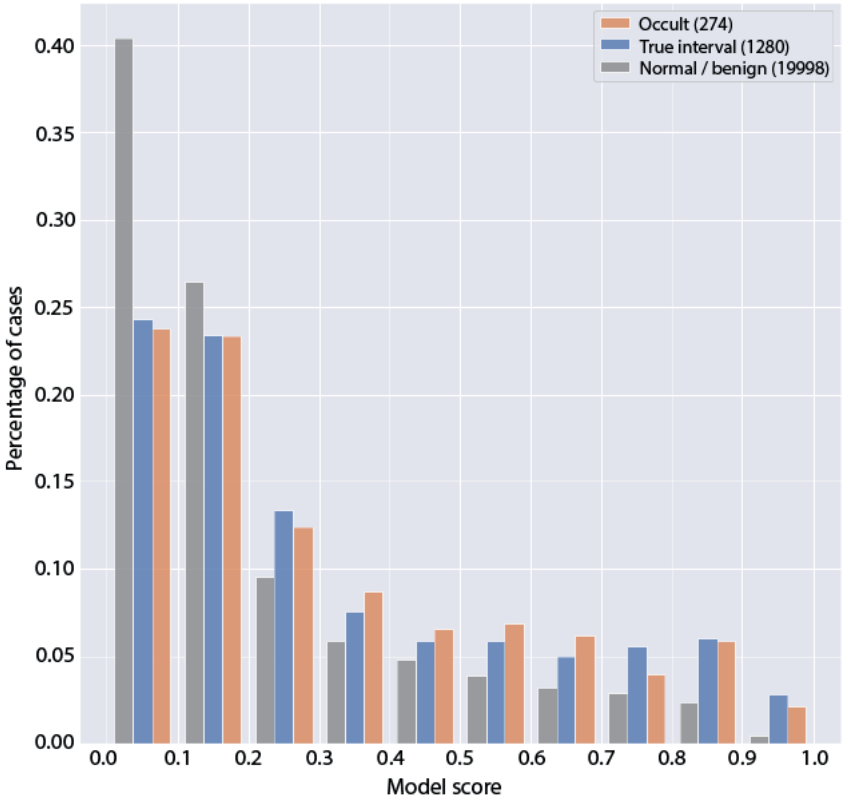


Figure 7. Comparison of distribution of AI model scores for occult and true interval cancers vs. normal cases with proven follow-up. The distribution of AI model scores derived from screening mammograms for true intervals and occult cases are compared to scores for follow-up proven normal (screening negatives) studies. Despite having no visible signs of cancer at the point of screening, true interval and occult cases yield an increased model score compared to normal studies.

Discussion

With the use of AI in screening, there is potential to improve upon program sensitivity by effectively increasing the number of screen-detected cancers that would have otherwise been missed by the radiologist. Using an AI system trained entirely on screen-detected cancers, we evaluated performance on a large cohort of N = 2,396 interval breast cancers. AI could potentially reduce the number of false negatives subsequently diagnosed as interval cancers by up to 27.5% based on an AI specificity of 99.0%, corresponding to a recall rate of approximately 1%. Among all retrospectively visible cancers, including minimal signs, the algorithm could detect 21.0% (177/842), and correctly localise 160/177 of cases detected at this threshold.

These findings outperform other AI systems evaluated on interval cancer datasets of composed of false negative and minimal signs cases.[9-11] For example, at a similar threshold (corresponding recall rate of ~ 1%), the AI algorithm reported by Lång et al. could detect and correctly localise 11.8% of retrospectively visible cancers.[10] In all studies, decreasing the AI's operating specificity would result in more cancers detected, but authors also concluded that this was only possible at the cost of unacceptably high recall of women for further investigation. Using the same AI algorithm reported by Lång et al. but set at a lower specificity of 97.5%, Wanders et al. evaluated performance on a different cohort of false negative cases. The algorithm could produce a heightened model score for 129 of 666 cases (19.4%), compared to 35.6–40.6% FNs detected by this AI at similar thresholds.[11] The Wanders et al. study did not include a comprehensive check for AI localisation accuracy, but noted cases could undergo direct additional examination or imaging.

Our findings of performance across all interval cancers, including true interval and occult cancers, demonstrates that AI is also able to produce high model scores for studies where no suspicious malignancy is discernible at screening. Previous CAD studies have emphasised the importance of analysing prior mammograms, hypothesising that screening mammograms obtained prior to interval cancer diagnosis could contain subtle signs of abnormality.[15,16] For example, an early sign of breast cancer is architectural distortion, characterised by subtle contraction of the breast tissue.[17,18] Among the screening images which AI flagged as suspicious and later had a true interval breast cancer diagnosis, AI could accurately indicate where on the image a malignant lesion would later develop in 2.8% of the cases. The present findings of heightened suspiciousness scores for non-retrospectively visible tumours might have important implications for understanding radiological features and underlying biological mechanisms of true interval and occult cancers, and should be explored further.

While AI findings on non-visible cancers may not yet be clinically actionable at the point of screening, they do provide an indication for increased surveillance. Dembrower et al. have suggested that studies with the highest AI scores deemed normal by both readers in a double-reader setting should undergo an enhanced assessment with supplementary imaging.[19] Another proposed approach is an AI model based on short-term risk prediction to enhance early detection for women deemed to be at high risk of a cancer diagnosis within 5 years.[20] Our findings could similarly be used to necessitate shorter intervals between screening, or use of supplemental imaging for selected women. This would be subject to further validation, as we cannot at present directly suggest that all screening exams with high suspiciousness as indicated by AI be recalled immediately. While we could not directly assess the number of additional false positives that could be induced by introducing AI, we limited our assessment to capture the range of acceptable recall rates as outlined in EUREF.[1] The choice of 99% specificity was intended to illustrate that the algorithm's specificity should operate above the screening program's specificity in order to limit a possibly detrimental impact on false positive rates. An assessment of the best specificity for prospective clinical use would require a simulation also taking into account sensitivity on screen-detected cancers, in order to assess the sensitivity–specificity trade-off on a representative screening population given an actual point of integration into the screening workflow. This is outside the scope and data available in this study, but is the subject of future work.

We evaluated the prognostic tumour characteristics of the interval cancers detected by AI, with special emphasis on features used to predict subsequent mortality from breast cancer.[21] Among the retrospectively visible cancers, 27.1% were at an advanced or metastatic stage at diagnosis, 21.1% of which could be detected by AI and localised at the exemplary operating point. Earlier detection of these cancers represents an important opportunity to improve long-term disease-specific outcomes.[8] With the exception of grade III tumours, there were no statistically significant differences observed between the cancers detected by AI and those not. This is confirmation that AI does not perform better or worse on certain breast cancer histopathological characteristics. However, long-term follow-up information was not available for the women in this cohort. This information would allow for a broader conclusion to be made about the use of AI to improve breast cancer screening sensitivity, as advanced breast cancer detection and breast cancer-specific survival are important endpoints for evaluating breast cancer screening effectiveness.

There is also potential for AI to be used as a tool during radiological review to automate classification of a subset of interval cancers. Until AI is able to redefine categorization of interval breast cancers in practice, this could help radiologists

with mandatory quality assurance work in organised breast cancer screening programmes. Using findings in Fig. 6 as an example, a low threshold of 0.2 can be used to automatically classify non-visible cancers. 47.08% of occult and 47.73% of true interval cancers could be classified correctly at the cost of classifying 20.86% of false negatives incorrectly. Alternatively, a high threshold of 0.9 could be used to find false negatives automatically and provide an estimation of the underlying false negative rate among interval cancers. In that case, 19.52% of false negatives could be detected correctly, while misclassifying 2.81%/2.19% of true intervals/occult cases, respectively. Given the large volume of interval cancers to assess, subjective nature of mammography interpretation and inter-observer variability, such a tool can provide up-to-date and accurate metrics evaluating radiologist performance within the screening program.

The evidence from this study suggests that AI can serve an important role in breast cancer screening. AI can accurately detect a proportion of cancers that would otherwise be missed by screening radiologists and subsequently diagnosed in the interval between screening rounds. We have demonstrated that this use of AI can decrease the interval cancer rate and thus improve overall program sensitivity. Our findings were based on an AI algorithm trained exclusively on screen-detected cancers and negatives with proven follow-up. With further exposure to data on interval cancers, future work will focus on enhancing the algorithm's sensitivity and discriminative ability.

Table 1. Cohort and tumour characteristics (N=2,396)

	Frequency N (%)
Interval cancer classification	
False negative	374 (15.6 %)
Minimal signs	468 (19.5 %)
Mammographically occult	274 (11.4 %)
True interval	1,280 (53.4 %)
Median age in years, IQR	60 (50-69)
Age group at screening	
50-54	642 (26.8 %)
55-59	543 (22.7 %)
60-64	569 (23.7 %)
65-69	640 (26.7 %)
Unknown	2 (0.1 %)
Time interval between screening and diagnosis	
3-12 months	787 (32.8 %)
13-24 months	1,609 (67.2 %)
ACR breast density	
I	222 (9.3 %)
II	1,085 (45.3 %)
III	751 (31.3 %)
IV	166 (6.9 %)
Unknown	172 (7.2 %)
Hormone receptor status (oestrogen receptor and/or progesterone receptor)	
Positive	1,506 (62.9 %)
Negative	383 (16.0 %)
Not determined/unknown	507 (21.1 %)
HER2/neu status	
Positive	370 (15.4 %)
Negative	1,456 (60.8 %)
Not determined/unknown	570 (23.8 %)
Tumour size	
Tis (in situ or Paget's disease)	107 (4.5 %)
T0 (no evidence of the primary tumour)	113 (4.7 %)
T1 (< 2 cm)	1,066 (44.5 %)
T2 (2-5 cm)	826 (34.5 %)
T3 (> 5 cm)	128 (5.3 %)
T4 (tumour grown into chest wall or skin; inflammatory breast cancer)	37 (1.5 %)
Unknown	119 (5.0 %)

	Frequency N (%)
Nodal status	
Positive	791 (33.0 %)
Negative	1,389 (58.0 %)
Unknown	216 (9.0 %)
Distant metastasis at diagnosis	
Yes	61 (2.5 %)
No	1,428 (59.6 %)
Unknown	907 (37.8 %)
Grade	
I	182 (7.6 %)
II	1,121 (46.8 %)
III	839 (35.0 %)
Unknown	254 (10.6 %)
Ki67	
< 20 %	669 (27.9 %)
≥ 20 %	732 (30.6 %)
Unknown	995 (41.5 %)

Table 2. Retrospectively visible cancers detected by AI at given operating points

Operating point specificity	False negative n/N (%, 95% CI)	Minimal signs n/N (%, 95% CI)	Locally advanced or metastatic at diagnosis n/N (%, 95% CI)	Total retrospectively visible cancers n/N (%, 95% CI)
99.5	88/374 (23.5 %, 19.5-28.1 %)	36/468 (7.7 %, 5.6-10.5 %)	40/228 (17.5 %, 13.2-23.0 %)	124/842 (14.7 %, 12.5-17.3 %)
99.0*	113/374 (30.2 %, 25.8-35.0 %)	64/468 (13.7 %, 10.9-17.1 %)	54/228 (23.7 %, 18.6-29.6 %)	177/842 (21.0 %, 18.4-23.9 %)
98.5	126/374 (33.7%, 29.1-38.6 %)	79/468 (16.9 %, 13.8-20.5 %)	63/228 (27.6 %, 22.2-33.8 %)	205/842 (24.3 %, 21.6-27.4 %)
98.0	133/374 (35.6 %, 30.9-40.5 %)	90/468 (19.2 % (15.9-23.0 %)	68/228 (29.8 %, 24.3-36.1 %)	223/842 (26.5 %, 23.6-29.6 %)
97.0	152/374 (40.6 %, 35.8-45.7 %)	99/468 (21.2 %, 17.7-25.1 %)	74/228 (32.5 %, 26.7-38.8 %)	251/842 (29.8 %, 26.5-32.6 %)

*An operating point of 99.0% specificity was selected as the exemplary operating point to emulate a recall rate of approximately 1%. At this operating point, 103 of 113 false negatives, 57 of 64 minimal signs, and 48 of 54 retrospectively visible locally advanced/metastatic cases detected by AI were correctly localised.

Table 3. Retrospectively visible cancers (false negative and minimal signs) detected by AI (specificity 99.0%)

		Not detected by AI N (%)	Detected by AI N (%)	Univariable OR 95% CI, *P*-value	Overall *P* value*
Age at diagnosis	**50-54**	147 (22.1)	30 (16.9)	-	0.33
	55-59	132 (19.8)	44 (24.9)	1.63 (0.97-2.77, p=0.07)	
	60-64	178 (26.8)	48 (27.1)	1.32 (0.80-2.21, p=0.28)	
	65-70	208 (31.3)	55 (31.1)	1.30 (0.80-2.14, p=0.30)	
ACR breast density	**I**	73 (11.0)	20 (11.3)	-	0.80
	II	325 (48.9)	87 (49.2)	0.98 (0.57-1.73, p=0.93)	
	III	194 (29.2)	56 (31.6)	1.05 (0.60-1.91, p=0.86)	
	IV	23 (3.5)	5 (2.8)	0.79 (0.24-2.22, p=0.68)	
	Missing	50 (7.5)	9 (5.1)	0.66 (0.27-1.52, p=0.34)	
Hormone receptor status	**Negative**	83 (12.5)	15 (8.5)	-	0.30
	Unknown	158 (23.8)	41 (23.2)	1.44 (0.76-2.82, p=0.27)	
	Positive	424 (63.8)	121 (68.4)	1.58 (0.90-2.94, p=0.13)	
HER2 status	**Negative**	402 (60.5)	108 (61.0)	-	0.96
	Unknown	176 (26.5)	45 (25.4)	0.95 (0.64-1.40, p=0.80)	
	Positive	87 (13.1)	24 (13.6)	1.03 (0.61-1.67, p=0.92)	
Tumour size	**T0**	27 (4.1)	3 (1.7)	-	0.40
	T1	306 (46.0)	83 (46.9)	2.44 (0.84-10.40, p=0.15)	
	T2	227 (34.1)	70 (39.5)	2.78 (0.94-11.86, p=0.10)	
	T3	32 (4.8)	9 (5.1)	2.53 (0.68-12.27, p=0.20)	
	T4	10 (1.5)	1 (0.6)	0.90 (0.04-8.01, p=0.93)	
	Tis	29 (4.4)	6 (3.4)	1.86 (0.44-9.52, p=0.41)	
	Unknown	34 (5.1)	5 (2.8)	1.32 (0.30-6.92, p=0.72)	
Lymph node status	**N0**	392 (58.9)	95 (53.7)	-	0.18
	N1	121 (18.2)	48 (27.1)	1.64 (1.09-2.44, p=0.02)	
	N1mi	11 (1.7)	3 (1.7)	1.13 (0.25-3.69, p=0.86)	
	N2	47 (7.1)	11 (6.2)	0.97 (0.46-1.87, p=0.92)	
	N3	30 (4.5)	8 (4.5)	1.10 (0.46-2.37, p=0.82)	
	Unknown	64 (9.6)	12 (6.8)	0.77 (0.38-1.44, p=0.44)	
Metastatic breast cancer	**M0**	395 (59.4)	113 (63.8)	-	0.27
	M1	20 (3.0)	2 (1.1)	0.35 (0.06-1.22, p=0.16)	
	Unknown	250 (37.6)	62 (35.0)	0.87 (0.61-1.22, p=0.42)	
Grade	**1**	44 (6.6)	22 (12.4)	-	0.003
	2	342 (51.4)	104 (58.8)	0.61 (0.35-1.08, p=0.08)	
	3	202 (30.4)	39 (22.0)	0.39 (0.21-0.72, p=0.002)	
	Unknown	77 (11.6)	12 (6.8)	0.31 (0.14-0.68, p=0.004)	
Ki67 proliferation	**High Ki67**	187 (28.1)	35 (19.8)	-	0.08
	Low Ki67	177 (26.6)	55 (31.1)	1.66 (1.04-2.68, p=0.04)	
	Unknown	301 (45.3)	87 (49.2)	1.54 (1.01-2.40, p=0.049)	

**P*-value based on the Pearson Chi-squared test.

Table 4. Likelihood of detection with AI based on tumour characteristics

		Logistic regression by interval cancer type	
	Full cohort: Detected by AI (n=177) vs. not detected (n=665) Multivariable OR (95% CI, *P*-value)	False negative interval cancers: Detected by AI (n=113) vs. not detected (n=261) Multivariable OR (95% CI, *P*-value)	Minimal signs interval cancers: Detected by AI (n=64) vs. not detected (n=404) Multivariable OR (95% CI, *P*-value)
Age at diagnosis			
50-54	-	-	-
55-59	1.62 (0.95-2.80, p=0.08)	2.23 (1.03-4.98, p=0.045)	1.22 (0.55-2.75, p=0.63)
60-64	1.35 (0.80-2.29, p=0.26)	1.73 (0.85-3.63, p=0.13)	0.72 (0.29-1.73, p=0.46)
65-70	1.34 (0.80-2.26, p=0.27)	1.59 (0.77-3.39, p=0.22)	1.14 (0.52-2.53, p=0.75)
ACR breast density			
I	-	-	-
II	0.94 (0.54-1.68, p=0.83)	0.61 (0.29-1.28, p=0.18)	2.27 (0.73-10.00, p=0.21)
III	1.07 (0.60-1.98, p=0.82)	0.82 (0.37-1.84, p=0.63)	3.02 (0.94-13.58, p=0.09)
IV	0.70 (0.21-2.02, p=0.53)	0.30 (0.04-1.54, p=0.18)	3.49 (0.55-22.40, p=0.17)
Unknown	0.61 (0.24-1.45, p=0.27)	0.47 (0.11-1.71, p=0.27)	2.01 (0.43-10.95, p=0.38)
Hormone receptor status			
Negative	-	-	-
Positive	1.10 (0.59-2.16, p=0.78)	2.50 (0.86-9.15, p=0.12)	0.48 (0.20-1.21, p=0.11)
Unknown	1.00 (0.35-2.87, p>0.99)	1.56 (0.32-8.31, p=0.59)	0.94 (0.17-5.10, p=0.95)
HER2 status			
Negative	-	-	-
Positive	1.08 (0.63-1.82, p=0.77)	0.91 (0.40-1.97, p=0.81)	1.20 (0.52-2.57, p=0.66)
Unknown	1.09 (0.45-2.58, p=0.84)	1.98 (0.60-6.66, p=0.26)	0.38 (0.07-1.75, p=0.23)
Tumour size			
<2 cm	-	-	-
2-5 cm	1.11 (0.76-1.63, p=0.59)	1.02 (0.59-1.75, p=0.94)	1.09 (0.58-2.03, p=0.78)
>5 cm	0.82 (0.36-1.73, p=0.62)	0.86 (0.29-2.29, p=0.76)	0.67 (0.14-2.34, p=0.56)
In situ	0.80 (0.23-2.43, p=0.70)	0.52 (0.11-2.07, p=0.38)	1.28 (0.04-17.74, p=0.86)
Unknown	0.80 (0.18-3.43, p=0.76)	0.56 (0.08-3.59, p=0.55)	2.93 (0.18-55.11, p=0.45)
Lymph node status			
N0	-	-	-
N1	1.67 (1.10-2.54, p=0.02)	1.48 (0.81-2.69, p=0.20)	2.57 (1.32-5.00, p=0.005)
N2+	1.16 (0.63-2.06, p=0.63)	1.10 (0.49-2.40, p=0.81)	1.37 (0.49-3.46, p=0.53)
Unknown	1.18 (0.41-3.16, p=0.75)	1.44 (0.40-4.99, p=0.57)	0.64 (0.05-4.99, p=0.71)
Metastatic breast cancer			
M0	-	-	-
M1	0.37 (0.06-1.37, p=0.20)	0.36 (0.05-1.50, p=0.21)	NA*
Unknown	0.92 (0.61-1.38, p=0.71)	0.80 (0.45-1.40, p=0.44)	1.21 (0.63-2.28, p=0.56)
Grade			
1	-	-	-
2	0.59 (0.33-1.08, p=0.08)	0.58 (0.24-1.37, p=0.21)	0.68 (0.28-1.80, p=0.42)
3	0.42 (0.21-0.85, p=0.02)	0.33 (0.12-0.88, p=0.03)	0.51 (0.17-1.63, p=0.25)
Unknown	0.37 (0.14-0.95, p=0.04)	0.35 (0.09-1.25, p=0.11)	0.23 (0.03-1.22, p=0.11)
Ki67 proliferation			
Low Ki67	-	-	-
High Ki67	0.78 (0.46-1.31, p=0.35)	1.03 (0.49-2.16, p=0.93)	0.62 (0.27-1.39, p=0.25)
Unknown	1.06 (0.68-1.66, p=0.80)	1.30 (0.69-2.47, p=0.42)	0.67 (0.32-1.40, p=0.29)

*Point estimates indicated with NA do not have any observations

Logistic regression by interval year	
Interval cancers diagnosed within 12 months post-screening: Detected by AI (n=67) vs. not detected (n=246) Multivariable OR (95% CI, *P*-value)	**Interval cancers diagnosed 13-24 months post-screening: Detected by AI (n=110) vs. not detected (n=419) Multivariable OR (95% CI, *P*-value)**
-	-
2.35 (1.01-5.68, p=0.05)	1.27 (0.61-2.67, p=0.53)
0.92 (0.37-2.28, p=0.86)	1.87 (0.96-3.74, p=0.07)
0.87 (0.35-2.19, p=0.77)	1.66 (0.86-3.29, p=0.14)
-	-
0.77 (0.26-2.65, p=0.66)	0.88 (0.45-1.76, p=0.70)
0.68 (0.21-2.44, p=0.53)	1.20 (0.60-2.52, p=0.61)
0.92 (0.13-5.77, p=0.93)	0.41 (0.06-1.76, p=0.28)
0.41 (0.07-2.12, p=0.30)	0.68 (0.22-1.94, p=0.49)
-	-
2.66 (0.89-10.02, p=0.11)	0.61 (0.27-1.45, p=0.25)
4.31 (0.54-39.37, p=0.17)	0.48 (0.14-1.68, p=0.24)
-	-
1.52 (0.62-3.56, p=0.35)	0.96 (0.46-1.90, p=0.90)
0.59 (0.08-3.51, p=0.57)	1.41 (0.51-3.77, p=0.49)
-	-
0.98 (0.50-1.93, p=0.95)	1.11 (0.67-1.80, p=0.69)
0.78 (0.21-2.52, p=0.69)	0.75 (0.23-2.09, p=0.61)
NA	1.25 (0.32-4.30, p=0.74)
0.68 (0.03-23.97, p=0.81)	0.62 (0.10-3.36, p=0.59)
-	-
1.39 (0.67-2.82, p=0.37)	1.94 (1.13-3.31, p=0.02)
0.44 (0.13-1.28, p=0.16)	2.00 (0.94-4.10, p=0.07)
2.94 (0.13-29.30, p=0.39)	1.18 (0.36-3.59, p=0.77)
-	-
NA*	0.64 (0.09-2.68, p=0.58)
0.99 (0.49-2.00, p=0.98)	0.80 (0.47-1.35, p=0.42)
-	-
0.79 (0.27-2.48, p=0.67)	0.47 (0.23-1.01, p=0.047)
0.52 (0.15-1.79, p=0.29)	0.31 (0.12-0.78, p=0.01)
0.22 (0.03-1.36, p=0.13)	0.40 (0.12-1.28, p=0.13)
-	-
0.49 (0.19-1.23, p=0.14)	0.92 (0.47-1.79, p=0.82)
0.90 (0.41-2.04, p=0.81)	1.05 (0.60-1.86, p=0.86)

References

1. Perry N, Broeders M, de Wolf C, Törnberg S, Holland R, von Karsa L. European guidelines for quality assurance in breast cancer screening and diagnosis. Fourth edition-summary document. *Ann Oncol* 2008; **19**(4): 614-22.
2. Houssami N, Hunter K. The epidemiology, radiology and biological characteristics of interval breast cancers in population mammography screening. *NPJ Breast Cancer* 2017; **3**(1): 1-13.
3. Houssami N, Irwig L, Ciatto S. Radiological surveillance of interval breast cancers in screening programmes. *Lancet Oncol* 2006; **7**(3): 259-65.
4. Hovda T, Hoff SR, Larsen M, Romundstad L, Sahlberg KK, Hofvind S. True and Missed Interval Cancer in Organized Mammographic Screening: A Retrospective Review Study of Diagnostic and Prior Screening Mammograms. *Acad Radiol* 2021.
5. Maes RM, Dronkers DJ, Hendriks JH, Thijssen MA, Nab HW. Do non-specific minimal signs in a biennial mammographic breast cancer screening programme need further diagnostic assessment? *Br J Radiol* 1997; **70**: 34-8.
6. Tsuruda KM, Hovda T, Bhargava S, Veierød MB, Hofvind S. Survival among women diagnosed with screen-detected or interval breast cancer classified as true, minimal signs, or missed through an informed radiological review. *Eur Radiol* 2021; **31**(5): 2677-86.
7. Niraula S, Biswanger N, Hu P, Lambert P, Decker K. Incidence, characteristics, and outcomes of interval breast cancers compared with screening-detected breast cancers. *JAMA Netw Open* 2020; **3**(9): e2018179-e.
8. Mook S, Van't Veer LJ, Rutgers EJ, et al. Independent prognostic value of screen detection in invasive breast cancer. *J Natl Cancer Inst* 2011; **103**(7): 585-97.
9. Graewingholt A, Rossi PG. Retrospective analysis of the effect on interval cancer rate of adding an artificial intelligence algorithm to the reading process for two-dimensional full-field digital mammography. *J Med Screen* 2021; **28**(3): 369-71.
10. Lång K, Hofvind S, Rodríguez-Ruiz A, Andersson I. Can artificial intelligence reduce the interval cancer rate in mammography screening? *Eur Radiol* 2021; **31**(8): 5940-7.
11. Wanders AJ, Mees W, Bun PA, et al. Interval Cancer Detection Using a Neural Network and Breast Density in Women with Negative Screening Mammograms. *Radiology* 2022: 210832.
12. Kääb-Sanyal VH, Elisabeth. Jahresbericht Evaluation 2018: Deutsches Mammographie-Screening-Programm. Berlin, 2020.
13. D'Orsi C, Bassett L, Feig S. Breast imaging reporting and data system (BI-RADS). *Breast imaging atlas* 2018.
14. Domingo L, Hofvind S, Hubbard RA, et al. Cross-national comparison of screening mammography accuracy measures in U.S., Norway, and Spain. *Eur Radiol* 2016; **26**(8): 2520-8.
15. Banik S, Rangayyan RM, Desautels JL. Detection of architectural distortion in prior mammograms. *IEEE Trans Med Imaging* 2010; **30**(2): 279-94.
16. Rangayyan RM, Banik S, Desautels JE. Detection of architectural distortion in prior mammograms via analysis of oriented patterns. *J Vis Exp* 2013; (78).
17. Gaur S, Dialani V, Slanetz PJ, Eisenberg RL. Architectural distortion of the breast. *AJR Am J Roentgenol* 2013; **201**(5): W662-70.
18. Bahl M, Baker JA, Kinsey EN, Ghate SV. Architectural Distortion on Mammography: Correlation With Pathologic Outcomes and Predictors of Malignancy. *AJR Am J Roentgenol* 2015; **205**(6): 1339-45.

19. Dembrower K, Wåhlin E, Liu Y, et al. Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologist workload: a retrospective simulation study. *Lancet Digit Health* 2020; **2**(9): e468-e74.
20. Yala A, Mikhael PG, Strand F, et al. Multi-institutional validation of a mammography-based breast cancer risk model. *J Clin Oncol* 2021: JCO. 21.01337.
21. Giuliano AE, Connolly JL, Edge SB, et al. Breast cancer—major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA Cancer J Clin* 2017; **67**(4): 290-303.

Supplementary materials

Supplementary Methods A. 1. Threshold setting

A threshold setting dataset of negative cases was used to estimate the models' specificity on a German screening population. The dataset comprises 20,000 screening examinations from 20,000 women and was collected from 8 screening units from the German screening program. All cases were negative and also had a negative follow-up screening study after a minimum of 24 months. No cases from either the interval cancer dataset nor the threshold setting dataset were used for training the AI model.

Specificity is determined by running the model on the threshold setting dataset and setting a model decision threshold such that a certain specificity is reached. For example, a specificity of 98% means we choose the detection threshold for the model such that 98% of the 20,000 negative cases are below the threshold (classified correctly as negative) and then observe how many interval cancers would be detected, i.e. are above the threshold (classified correctly as positive). Sample weights were applied to the threshold setting dataset to reflect the actual distribution of study types in the German breast screening population according to screening stage (Supplementary Table A. 1.).

Supplementary Table A. 1. Derivation of sample weights for threshold setting dataset

Clinical subgroup	Percentage in threshold setting dataset	Actual percentage in German screening population	Weight
Benign biopsies	1.19%	0.51%	0.51% / 1.19% = 0.43
Recalled but no biopsy	18.02%	1.80%	1.80% / 18.02% = 0.10
Consensus conference but no recall	15.28%	9.30%	9.30% / 15.28% = 0.61
No consensus conference	65.52%	87.80%	87.80% / 65.52% = 1.30


9

Chapter 9

Discussion, future perspectives, and conclusion

General discussion

The last decade has seen significant reforms in how early-stage breast cancer and DCIS are understood and treated. These changes would not have been made possible without an interdisciplinary approach to understanding these diseases and the preferences that the women affected by them have. In this discussion, we revisit the three complementary themes in the management of early-stage breast cancer. We uncover a few factors that affect the de-escalation of low-value treatment using approaches grounded in health technology assessment.

In the first theme, we focus on screen-detected primary DCIS using the research performed within the PRECISION (PREvent ductal Carcinoma In Situ Invasive Overtreatment Now) Consortium. As part of the consortium, we undertook an HTA-based modeling exercise to describe the potential and the consequences of biomarker-based strategies to select low-risk women for an active surveillance strategy. However, because the depth of understanding for DCIS remains limited compared to invasive breast cancer, it was essential first to illustrate the disease etiology of DCIS, and to characterize real-world healthcare utilization and treatment preferences. We show how it was necessary to leverage real-world data to uncover real-world patient experience and foreshadow future opportunities for treatment optimization. The second theme focuses on treatment de-escalation for early-stage breast cancer, based on the first results of the EORTC 10041/BIG 3-04 MINDACT (Microarray in Node-Negative and 1 to 3 Positive Lymph Node Disease May Avoid Chemotherapy) phase 3 randomized control trial of the 70-gene signature. We discuss the history of adoption and policy implications of the 70-gene signature as a part of the wider discussion of biomarker technologies that have transformed personalized treatment decision-making. Finally, a complementary final theme and chapter highlights a promising new technology: artificial intelligence for to improve cancer detection at breast cancer screening to decrease the interval cancer rate. We finish with an exploration of the careful balance between overdiagnosis, overtreatment, and underdiagnosis.

Using statistical models to reflect reality

In **Chapter 2,** we use a multi-state model to illustrate the disease etiology of untreated DCIS and calculate the life years associated with different treatment strategies. Modeling the life course of different women with DCIS was an important starting point for future cost-effectiveness modeling. It served to highlight the diverse pathways that women could experience. Using an extensive population-based cancer registry like the *Surveillance, Epidemiology, and End Results Program*

(SEER) also made it possible to identify the largest cohort of women reported to date who did not receive surgery or radiotherapy for their DCIS. The SEER dataset is also rich in clinicopathological and sociodemographic information, which helps us understand who is more likely to receive specific treatment modalities and how this impacts their health outcomes. Among untreated women, a subgroup with low-risk features (white women aged 50 to 69 at diagnosis, with estrogen receptor-positive (ER+), grade I/II DCIS less than 2 cm) could be identified and used as inputs for a multi-state model. Despite the limitations of incomplete data drawn from cancer registries, this data builds a foundation for future health economic modeling which incorporates the costs and health-related quality of life associated with being in each health state.

An essential feature of multi-state modeling, making it particularly relevant to cost-effectiveness modeling, is that it considers competing risks. A competing risk is an event whose occurrence precludes the primary event of interest.[1] This is useful when modeling the entire life course of women with DCIS – because DCIS itself is not deadly.[2,3] It is the possible progression to invasive breast cancer and the subsequent possibility of that breast cancer metastasizing that directly impacts a woman's mortality. Yet the risk of progression occurs alongside other risks present among screening-age women. For example, statistically speaking, the presence of one or more comorbidities may put these women at a higher likelihood of dying from non-oncological causes.[3] Furthermore, and most relevant to the PRECISION consortium, is the understanding that each woman has a different lifetime risk of experiencing invasive breast cancer after DCIS.

> ***Further research, practice and policy implications:*** To build a multi-state model for DCIS and the sub-populations of women, vast swathes of readily available individual-level data on real women with DCIS were necessary. Fortunately, this study was published at a time characterized by a growing appreciation for the use of real-world data (RWD) to uncover novel insights on treatment outcomes and to inform reimbursement decisions. Real-world data offers tremendous potential to provide a holistic picture of an individual's health status and inform healthcare decision-making, and can be generalized to populations beyond clinical trials. Thus healthcare decision-makers should be prepared to leverage RWD to fill evidence gaps.

Leveraging real-world data to uncover real-world patient experiences

Despite being considered the "gold standard" for clinical evidence,[4] there is a limit to what data from randomized controlled trials (RCTs) can tell us. While RCT data can bring us closest to understanding treatment effects in controlled experimental settings, the benefit may not be so apparent once treatments are administered in non-trial settings.[5] Furthermore, RCTs cannot tell us anything about the patient preferences that impact why some women choose specific treatment strategies over others. RCTs also don't tell us about treatment accessibility in the real world or disparities in outcomes among certain groups of women. By design, trial numbers are too small to uncover these important facets, and follow-up is too short to understand long-term consequences.

Through the PRECISION consortium, the LORD, LORIS, and COMET trials set out to explore the safety of an active surveillance strategy. An active surveillance strategy may only be successful if we correctly identify women with low-risk DCIS and they adhere to annual mammography surveillance. Screening with digital mammography is already the foundation of every surveillance regime in women treated with breast conserving surgery; breast ultrasound or magnetic resonance imaging (MRI) as an adjunct to mammography can also be considered for women with additional risk factors.[6] For women with a history of breast cancer and DCIS, regular follow-up screening is intended to ensure that any invasive recurrences are caught early and have the best possible chance of successful treatment. Tailored screening, capturing patient preference and individual risk factors would help clinicians consider which women are at increased risk of recurrence.[7]

In **Chapter 3**, we set out to characterize the real-world uptake of surveillance breast imaging in a contemporary cohort of women with primary screen-detected DCIS using extensive data from the National Cancer Database DCIS Special Study. We wanted to understand whether certain characteristics, including the type of treatment a woman received and any sociodemographic characteristics, had any relationship to adherence to clinical guidelines that warranted annual mammography following therapy for all women.

This registry-based surveillance cohort included 12,559 women with primary DCIS and had detailed information on imaging for up to 10 years post-diagnosis. Given the already good prognosis of adequately treated DCIS, it was essential to understand how women were followed-up after treatment and whether there would be health-related consequences of under-utilization of screening. We found that a large proportion of women treated with breast conserving surgery do not adhere

to guideline-recommended annual surveillance imaging. This phenomenon is especially the case among women with DCIS- and treatment-specific characteristics related to a higher risk of subsequent ipsilateral invasive breast cancer. Women who did not undergo adjuvant radiotherapy or endocrine therapy following surgery were more likely to be non-screeners or inconsistent mammography screeners than those who did. A similar phenomenon of uptake of surveillance imaging driven by receipt of radiotherapy was observed among two separate cohorts of Dutch women with a history of invasive breast cancer.[8,9] In our study we also found that compared to women with screen-detected DCIS, women who had clinically diagnosed DCIS detected through palpation or breast symptoms also had a higher probability of not being consistent screeners. This finding is indicative of a group of women who are healthcare "under-users," who generally may not have attended regular screening mammography before their diagnosis. Indeed, uninsured women were less likely to be consistent screeners than government-insured women.

Racial and ethnic disparities in uptake were also evident for Black women compared to white women, and Hispanic women compared to non-Hispanic. These disparities in uptake occur despite what is already known, for example, about Black women being at significantly higher risk of subsequent invasive breast cancer,[10] which we also identified in Chapter 2. This risk is coupled with the higher likelihood of a diagnosis of poor prognosis triple-negative subtype and higher breast cancer mortality among Black women compared to white women.[10-12] This is further complicated by racial disparities in access to breast cancer diagnosis and treatment.[13] It has been argued that existing breast cancer screening recommendations put Black women at a disadvantage.[14]

> ***Further research, policy and practice implications:*** The consequences of this situation for women who did not adhere to annual surveillance are clear: lower rates of surveillance uptake can lead to advanced disease at presentation of recurrence, and this is directly related to poorer breast cancer-specific survival.[15] We found the rate of ipsilateral invasive breast cancer diagnosis to be higher in patients who received early surveillance imaging (within 12 months of definitive surgery) compared to those who did not. Early screening was a good proxy for adherence to annual screening (69% of early screeners were consistent screeners thereafter), suggesting that the time to detection of invasive recurrences may be shorter among DCIS patients who adhere to the screening guidelines. Establishing a regular pattern of surveillance soon after diagnosis may promote timely detection of ipsilateral recurrence in the long term.

In **Chapter 4,** we direct our attention to the Netherlands, where the LORD study is underway. Following difficulties recruiting and randomizing low-risk women to standard surgical treatment or active surveillance, study coordinators changed the LORD study design from a randomized controlled trial to a preference-based study. This change created a unique opportunity to explicitly measure the preferences of treatment strategies among these women in a setting where sociodemographic factors may not contribute to the uptake of treatment or adherence to surveillance imaging. We compared the learnings from these women to responses drawn from Dutch oncologists involved in the treatment of women with DCIS who completed the same questionnaire.

We used a discrete choice experiment (DCE) as a "stated preference" method, asking respondents to choose between alternative treatment strategies from a set of hypothetical scenarios generated from an experimental design.[16] When deciding upon treatment strategies for low-risk DCIS, the extensiveness of the locoregional treatment was consistently shown to be an important factor for both patients and care providers. Yet, the most apparent discordance in preference between the two groups was related to the risk of ipsilateral invasive breast cancer: we found this to be most important to oncologists and least important to patients. These women had very strong preferences for an active surveillance strategy with no surgery, irrespective of the 10-year risk of ipsilateral invasive breast cancer this carried. Meanwhile, physicians nearly exclusively chose strategies to minimize breast cancer risk.

Despite promoting shared decision-making,[17] discordant preferences between patients and health care providers are common.[18] A systematic literature review of 28 DCEs eliciting patient and healthcare provider preferences for healthcare interventions found that the most significant discordance between patients and healthcare providers was for disease progression and mortality outcomes. Healthcare providers believed this to be more important than patients. On the other hand, patients ranked factors related to treatment safety and processes more highly than healthcare providers. These processes included treatment-related adverse events and the delivery and timing of treatment. These were the same patterns of discordance we uncovered in **Chapter 4**.

> ***Further research, policy and practice implications:*** The differences in treatment preferences and how individuals relate this back to their understanding of the risk of disease progression could foreshadow the willingness of healthcare providers and patients to adopt new treatment strategies. Shared decision-making processes should alleviate this problem.

Shared decision-making is a form of communication between healthcare providers and patients where both parties have a mutual understanding.[15] It is understood to be the key to good healthcare, oriented to the needs of the individual and thus improves satisfaction or quality of life despite illness.[17] While the healthcare provider is and remains the expert in all medical matters, patients should have all information relevant to their decision-making. Their healthcare providers should fully inform them of the treatment options and the trade-offs between risks and benefits. Then patients can be empowered to openly share their thoughts, concerns, questions, and expectations. The final aim is to jointly decide on appropriate medical treatment, with both parties taking responsibility for this decision. Doing so can also increase compliance with and adherence to treatment, further increasing the chances of successful treatment outcomes.

How risky is it?

The concept of risk was central to all of the chapters presented in this dissertation. We started with the basic understanding that diagnosis with primary DCIS is related to an elevated risk of subsequent iIBC. It is now recognized that this risk is not evenly distributed across the population of women with primary DCIS, with evidence pointing to certain prognostic features associated with a higher and lower risk of invasive recurrence. However, one's risk of experiencing iIBC can be precluded by their risk of another health issue. Discussions around risk and screen-detected DCIS thus must be interwoven with an understanding of lead-time bias, whereby cases detected by screening appear to have prolonged survival because the disease was detected at an earlier time point, but not because mortality was delayed. In fact, contemporary research into breast cancer overdiagnosis uses a lead-time approach which accounts for competing mortality risk and a mixture of progressive and nonprogressive cancer. Ryser et al. defined the rate of overdiagnosis as the proportion of screen-detected breast cancer cases (including invasive breast cancer and DCIS) that were either non-progressive or progressive but would not have progressed to clinical (symptomatic) disease before the woman died of causes unrelated to breast cancer.[20] The researchers concluded that in a European-style screening setting (biennial screening), among screen-detected cases, overdiagnosis rates are likely to be driven by DCIS given its limited propensity to progress.

In **Chapter 4** we found that the immediacy of treatment-related adverse effects seemed more relevant to women than the future risk of breast cancer. The study sample did however contain an overrepresentation of patients who selected an active surveillance strategy. A post-hoc effect-modifier analysis found that

compared to women who underwent surgery, women choosing active surveillance would be more inclined towards scenarios with shorter follow-up intervals while accepting scenarios with higher risk of invasive breast cancer. In the discussion of **Chapter 4** we highlighted an important consideration that likely factors into a patient's treatment choice: the understanding of one's risk of upstaging to breast cancer. This was not measured in the DCE, but is relevant given the challenge of identifying patients with a core needle biopsy showing DCIS with "low-risk" clinicopathological characteristics who have concurrent invasive carcinoma in the breast. Two retrospective studies conducted on Dutch pathology registry data have shown between 13 to 18% of screen-detected DCIS were "underestimated" and upstaged to invasive DCIS following excision.[21,22] Notably, upstaging rates were lower among women with non-palpable and low-to-intermediate grade DCIS. An important takeaway from these studies is that even with possible upstaging, overall survival should not be significantly compromised. Access to high-quality annual mammography in the Netherlands is readily available, and invasive carcinomas can be treated on time. Nevertheless, the prediction of upstaging of DCIS to invasive disease remains an important area of ongoing research and will serve to identify the lowest achievable upstaging rate among women eligible for clinical trials of active surveillance. This may in turn address some of the challenges with trial accrual, and better inform the understanding of the risk of upstaging.

Balancing the trade-offs between individual risks and benefits of treatment strategies for DCIS was the focus of **Chapter 5.** We showed the results of an early economic evaluation of an active surveillance strategy based upon selecting women with low-risk features who can forgo surgery. Women with low-to-intermediate grade, estrogen receptor-positive (ER+) DCIS make up approximately 50% of screen-detected primary DCIS. These 'low-risk' women are the focus of the LORD, LORIS, and COMET trials studying the safety of an active surveillance strategy. Cost-effectiveness analyses can provide unique insights into the downstream costs associated with selecting different treatment strategies for this group of women. Only breast conserving surgery ± radiotherapy was used as a comparator, as mastectomy may be considered overtreatment for many women with small, localized DCIS. While initial treatment costs may be lower (i.e. 0€) for women undergoing active surveillance, this may not be the case over time as these women are expected to experience slightly higher rate of ipsilateral invasive breast cancer, thus possibly incurring higher downstream costs. Nevertheless, costs are not the only endpoint central to cost-effectiveness analyses. In this analysis, forgoing surgery among these women resulted in significant gains in quality of life, despite an expected elevated rate of iIBC and somewhat reduced life years. We provided results using incremental life years and quality-adjusted life years to offer a critical contrast when

applying utilities. Whereas incremental life years are reflected in the differences in overall survival between groups, utilities adjust this metric by incorporating the impact of patient preference and quality of life. In the base-case model, introducing an active surveillance strategy would result in life years lost (-0.06, 95% confidence interval (CI) -0.26 to 0.16) across the cohort. The application of utilities however positively shifted the health effects towards an average QALY gain of 0.4. Therefore this early economic evaluation demonstrated that introducing an active surveillance option to select women with low-risk features can be a cost-effective alternative to immediate surgery and adjuvant radiotherapy.

The question that remains to be answered in full is: what is the best biomarker to identify women who can safely forgo surgical treatment? In addition to low-to-intermediate grade/ER+ DCIS, we considered a different biomarker that can select a smaller, more defined group of women who gain no benefit from surgical intervention. Selecting women based on COX-2 expression and adipocyte size, we modelled a 'hypothetically perfect biomarker' scenario where iIBC rates in the low-risk group would match a healthy population without a history of DCIS.[23] The scenario analysis showed a higher QALY gain among this group: 1.02 incremental QALYs compared to 0.81 QALYS among the low-risk group defined in the base case analysis.

Further research, policy and practice implications: The potential to decrease low-value treatment for low-risk DCIS is clear. However, it will still be approximately 5 to 10 years before results from ongoing prospective studies on biomarker-based treatment de-escalation strategies become available. Modelling the variation in cost-effectiveness results in chapter 5 showed us that QALY gains were inconsistent across the population, reflecting inherent limitations to identifying and using prognostic biomarkers of progression to invasive disease.[24,25] Lips and colleagues recently published findings that one in five ipsilateral invasive breast cancers following DCIS are unrelated to the initial DCIS lesion and are actually new primary tumours.[26] This shifts our understanding of DCIS as a precursor lesion to invasive breast cancer, rendering the notion of prognostic biomarkers to predict invasion irrelevant for a significant proportion of women with DCIS. While the PRECISION consortium continues to develop evidence around the biological underpinnings of DCIS, decision makers should be prepared to accept that minimal life-years will be lost on average for active surveillance, regardless of the strategy to select low-risk women. However, this will always be balanced with significant cost-savings and gains in quality-adjusted life years.

Using predictive vs. prognostic biomarkers to measure individual risk and benefit of treatment

It may be helpful to take a step back to explore the concept of biomarkers within the context of oncology. Biomarkers for personalized oncologic care present two opportunities for understanding disease outcomes. First, prognostic value of a biomarker provides information about an individual's future risk of progression. Second, predictive value signifies whether an individual will likely respond beneficially to a specific type of ((neo-)adjuvant) therapy compared to no or an alternative treatment.

Many studies have been undertaken to identify biomarkers for clinical decision-making in breast cancer. Yet, for several reasons, few meaningful biomarkers have been identified and are used in practice.[25,27] Firstly, studies may not have utilized the correct statistical approach. For example, establishing treatment benefit does not follow a purely prognostic analysis but could require interaction analysis in which a statistically significant difference is demonstrated between treatment groups. This is further complicated by the fact that biologically, potential biomarkers may underestimate the complexity of a drug's mechanism of action, only explaining a fraction of the factors and mechanisms which affect treatment-related tumour behaviour and patient progression outcomes. Finally, as we have seen with the 70-gene signature, economic, regulatory, access, and preference-related aspects affect uptake significantly.

In early-stage breast cancer, gene expression profiles (GEPs) like the 70-gene signature are among a suite of biomarker technologies that have transformed personalized treatment decision-making.[28,29] Numerous clinical guidelines support the use of GEPs to aid in adjuvant treatment decisions for a subset of early breast cancer patients identified through clinicopathological characteristics.[30,31] Despite the early enthusiasm shown by clinicians in using GEPs to de-escalate adjuvant chemotherapy,[32] questions persisted about the consistency of chemotherapy benefits and the impact on the risk of progression within specific patient subgroups. Most of these patients are already effectively managed with optimal local treatment and adjuvant endocrine therapy. Patients with a genomic risk profile contradictory to clinical assessments of risk based on prognostic factors such as grade and nodal status have been of particular interest,[28,33] because the actual small effect of adjuvant chemotherapy was less evident in these populations.

Chapter 6 took patient-level data from the phase III EORTC 10041/BIG 3-04 Microarray in Node-Negative and 1 to 3 positive Lymph Node Disease May Avoid Chemotherapy (MINDACT) trial to model treatment strategies guided by the 70-

gene signature. The analysis focused on patients with discordant risk results: they were at high clinical risk of distant metastases as defined by the modified Adjuvant! Online clinicopathological assessment, but low genomic risk as defined by the 70-gene signature. In the trial, these patients were randomized to receive adjuvant chemotherapy on either the genomic or the clinical risk.[28] Using this subgroup, we conducted cost-effectiveness analyses for six countries: Belgium, France, Germany, the Netherlands, the United Kingdom, and the United States. Each model considered country-specific costs, available treatment regimens, and utility values unique to their populations. Our models showed that treatment strategies guided by the 70-gene signature saved costs in five of six countries, gained QALYs, and were cost-effective in all six countries given country-specific willingness-to-pay thresholds.

Following a theme similar to the cost-effectiveness of active surveillance for primary low-risk DCIS presented **in Chapter 5**, **Chapter 6** describes how minimal survival differences are balanced against quality of life gains per patient. Considerable cost savings are realized when using the 70-gene signature compared to clinical assessment in guiding treatment decisions.

HTA as an iterative process

The study presented in **Chapter 6** was certainly not the first to establish the cost-effectiveness of the 70-gene signature. However, it was the first to use data from a prospective, randomized controlled trial. In 2009, an HTA study providing an overview of how to enable decisions on coverage and reimbursement of the 70-gene signature in the Netherlands was published.[34] This was quickly followed in 2010 and 2013 by two analyses of the cost-effectiveness of the 70-gene signature. The former published analysis was based on a series of validation studies and compared the 70-gene signature to the St. Gallen guidelines and Adjuvant! Online.[35] The latter analysis utilized data from a prospective cohort of patients from the RASTER study.[36,37] With the final cost-effectiveness analysis based on the randomized controlled trial, we demonstrated that the conclusion that the 70-gene signature would be cost-effective remained similar throughout the iterative process.

> ***Further research, policy and practice implications:*** It is now widely accepted that iteratively-conducted economic evaluations should be part of the health technology assessment process, starting at the earliest stages of developing new technologies.[38-40] This facilitates the incorporation of new evidence as it becomes available at different points in time. It has also been suggested that economic evaluations in early stages, e.g. alongside phase I and II clinical research, can be advantageous for the uptake of new medical technologies.[41]

Evidence-based practice and policy-making: the 70-gene signature case

The HTA analyses conducted for the 70-gene signature provide a clear example of the use of iterative processes to anticipate future developments and barriers to uptake of the promising GEPs. **Chapter 6** continued the relevant conversation surrounding the economic impact and clinical utility of using GEPs in practice. Clinical guidelines from the American Society for Clinical Oncology, the European Society for Medical Oncology, the National Comprehensive Cancer Center, and others have long since included the 70-gene signature in their respective clinical guidelines for early-stage breast cancer. They acknowledge that it, and other GEPs, are a valuable tool to determine if adjuvant chemotherapy is warranted after the surgical removal of a tumour.[42-45] Yet, not all insurance funds have heeded this advice by extending insurance coverage of this or other breast cancer GEPs.

In 2018, the Netherlands' national healthcare institute (Het Zorginstituut Nederland (ZIN)) made the decision to no longer reimburse the 70-gene signature in the country's primary health insurance package, because ZIN found insufficient evidence that chemotherapy could be safely waived for some women based on the 70-gene signature test result, based on the publication in the New England Journal of Medicine.[28]

Our cost-effectiveness analysis in **Chapter 6** showed that using the 70-gene signature could be acceptable even if it meant a small increased risk of distant metastasis, but quality-adjusted life years gained. An important aspect of this modeling approach is that it carefully considers the impact of patient quality of life. Curiously, ZIN made its decision to stop reimbursement despite more than a decade of research to support the clinical utility of the 70-gene signature and the long and close involvement of Dutch physicians and researchers in its development.

> ***Further research, policy and practice implications:*** Indeed, policymakers at the ZIN and elsewhere face complex challenges when deciding to implement tools to de-escalate oncological treatment. Policymakers must make these decisions on behalf of the population, given the available information, with the primary goal of ensuring equal access to (innovations in) health care while maintaining affordable health insurance premiums. Policymakers are tasked with supporting effective treatment delivery while avoiding adverse effects to patients and the healthcare system through soaring costs. In situations where the evidence surrounding the clinical benefit of a new healthcare intervention is not yet clear – or are not convincing enough to warrant reimbursement, it could be worthwhile to additionally consider the

result of cost-effectiveness analyses in these decisions. Cost-effectiveness analyses bring additional important insights for reimbursement decisions as they also directly model societal costs and patient quality of life.

Other critics of the initial results of the MINDACT trial argued that given the persistent long-term risks of recurrence, the results were likely to shift in favour of the use of chemotherapy. However, upon release of the long-term follow-up results of the MINDACT trial, there was no clear chemotherapy benefit in the women with high clinical risk and low genomic risk at 8.7 years median follow-up.[46] Today, we have seen how GEPs have dramatically lowered the use of adjuvant chemotherapy in this subset of breast cancers, without adversely affecting clinical outcomes.[42] Testing with GEPs is now recommended for most ER+, HER2-negative cancers, irrespective of grade or menopausal status, in cases with up to one positive lymph node.[42-45] Given these clear benefits and wide-spread use, there is still hope that the ZIN will reverse their decision on coverage of the 70-gene signature.

The more you know, the less you need

The MINDACT trial also created the opportunity for a posthoc study of breast cancer outcomes in a subgroup of women with the lowest risk of recurrence. In **Chapter 7,** we evaluated a suite of different statistical endpoints to better understand endocrine therapy's added value for women with ER+, HER2-negative, lymph node-negative tumours ≤ 2 cm. Across all ER+ breast cancers, endocrine therapy can reduce the risk of breast cancer death by approximately 30 %. However, this is related first and foremost to lowering the risk of distant metastasis, which is the major cause of death in breast cancer patients. Given their breast cancer features, the women in this study already have a very low risk of distant metastasis, yet most clinical guidelines still recommend endocrine therapy.[44,45,47,48]

The side-effects of endocrine therapy are often underestimated but are essentially the reason for relatively poor adherence—only 50% of breast cancer patients complete five years of treatment.[49,50] The most frequent therapy-induced side effects include vasomotor symptoms, including hot flashes and night sweats, vulvo-vaginal symptoms, and musculoskeletal symptoms leading to a higher incidence of bone fractures.[51] If these women must endure overall poor quality of life and heightened risk of treatment-related adverse events, then it becomes essential to measure the absolute benefit of endocrine therapy for them accurately.

We first focused on endpoints relating to the risk of distant metastasis and overall survival, finding an absolute 2.5 % lower distant metastasis free interval among women treated with endocrine therapy at 8 years. There was no statistically significant difference in overall survival or breast cancer-specific survival between women who received endocrine therapy and those who did not during the observation period. A greater magnitude of difference only emerged when considering the cumulative incidence of locoregional recurrences and contralateral breast cancers, contributing to an 8.3 % difference between groups when considering all possible breast cancer events together. The decision to forgo treatment is more challenging than decisions surrounding escalating treatment or using alternative treatments, but patients generally are more willing to accept de-escalation for treatments with worse side effects.[52]

The best protection is in early detection

For many women with breast cancer, particularly those with high-risk features associated with poor prognoses, the best chances at progression-free survival are rooted in early detection. In a study of patients with high-risk tumours according to the 70-gene signature, a significant worse difference in the distant metastasis-free interval was observed for women whose cancer was detected outside of screening as an interval cancer, as compared to screen-detected cancers.[53] Method of detection remains a significant independent prognostic factor. In **Chapter 8** we introduce a new artificial intelligence (AI)-based technology to support image interpretation during breast cancer screening with mammography. This technology has been demonstrated in retrospective simulation studies to increase the number of screen-detected cancers, in turn potentially decreasing the interval cancer rate in population-based screening.[54]

The study on interval cancers follows the publication from Leibig et al[49] showing the results of a retrospective simulation study of this AI using mammography examinations from screen-detected cancers and follow-up proven normal examinations. A combined approach is used where two AI systems of normal triage and cancer detection work together to achieve joint improvement of screening sensitivity and specificity of a radiologist. This combined approach works within a decision-referral pathway using the two AI systems with complementary algorithmic thresholds: one for pre-screen triage of normal mammography examinations and the other for post-screen of examinations for cancers potentially missed by the radiologist. The remaining screening examinations with scores falling between the algorithmic thresholds are referred to the radiologist for interpretation. The intention behind such an approach is to facilitate safe clinical adoption of an AI-based system for breast cancer screening without replacing the essential role of the human reader.

Chapter 8 and Leibig et al.[54] have illustrated the potential to improve cancer detection and reduce workload at screening. However, the integration of this system and its performance on live prospective data has not yet been formally studied, and is the focus of an ongoing nationwide prospective observational study in Germany (DRKS00027322). To date, no prospective evidence (either RCT, test accuracy study or cohort study) of any AI system for breast cancer screening has been published.

Evaluating the prospective performance of AI and the interaction with users is particularly important given the history of poor performance of computer-aided detection (CAD) solutions. Early reports of efficacy were based on studies in controlled, experimental settings, but this did not translate into higher cancer detection in clinical settings. Large retrospective registry-based studies disproved claims by CAD, concluding that CAD does not improve diagnostic accuracy of mammography and may result in missed cancers.[55] Like CAD, AI for breast cancer screening with mammography needs to be trialled extensively in a variety of clinical settings to understand the downstream effects.

The consequence of overdiagnosis is overtreatment

Interventions relating to breast cancer screening present a unique challenge. A delicate balance must be struck between not missing cancers at screening, decreasing the number of interval breast cancers, and preventing further overdiagnosis among screen-detected cancers. Early HTA could uncover whether improvements in cancer detection and workload reduction translate to improved health outcomes at the population-level, and cost savings for the healthcare system.

The latest natural history models of breast cancer have estimated that among a population of screening-age women undergoing biennial screening, 15.4% of screen-detected cancers were estimated to be overdiagnosed.[20] More than one-third of these were due to detecting indolent preclinical cancer, and the remaining were due to detecting progressive preclinical cancer in women who would have died of an unrelated cause before clinical diagnosis. The consequence of overdiagnosis is overtreatment, and this has a profound negative impact on affected women.[56] Studies have shown that the introduction of population-based screening programs has led to an increased detection in biologically low-risk, and ultralow-risk cancers.[57,58] These low-risk and ultra-low risk are overrepresented within screen-detected cancer, but also occur among interval cancers.[53]

> ***Further research, policy and practice implications:*** Risk-based breast cancer screening could be a promising solution to prevent overdiagnosis and overtreatment for some, and underdiagnosis and undertreatment for others.[59]

It incorporates risk factors like family history of breast cancer, breast density, personal history of breast biopsies, polygenic risk score representing the cumulative effects of genetic variants, and sequencing for moderate- and high-penetrance germline mutations.[60] This allows women to undergo different screening strategies, where the interval between screening rounds and imaging modalities can differ. Another proposed approach to risk-based screening is an AI model based on short-term risk prediction to enhance early detection for women deemed to be at high risk of a cancer diagnosis within 5 years.[61]

In the Netherlands, there are many in the medical community seriously considering whether population-based breast cancer screening should be continued 'as-is' – even despite very high uptake and good outcomes in the population. A large prospective cohort study PRISMA (Dutch Personalised RISk-based MAmmography screening) has already begun, with the aim of updating and validating an existing breast cancer risk prediction model to guide Dutch screening policy. This follows the growing trend elsewhere in Europe,[62] the UK,[63] Canada,[64] and the United States[59] to similarly study risk-based screening approaches in their populations. While results from these trials will not be available for several years, women have already expressed high interest in receiving information on breast cancer risk estimates alongside tailored screening recommendations.[65,66]

Future perspectives and conclusion

Much of what has been achieved regarding breast cancer survival outcomes are due to early detection and surgical, chemoradiation, and targeted treatments. Yet the standard "one-size-fits-all" screening and treatment pathway has resulted in very different levels of benefit and harm amongst women with breast cancer and DCIS. This PhD delved into some of these harms, the multi-faceted consequences, and potential solutions. The following two overlapping opportunities emerged, the first related treatment optimization, and the second to optimizing screening and surveillance:

Opportunity Number 1: Apply risk assessment at the point of diagnosis to establish the best individualized treatment and surveillance pathway. Treatment should be de-escalated for low-risk women to maintain their quality of life, while freeing up resources to find meaningful treatment for high-risk women.

> **Opportunity Number 2:** Explore alternatives to population-based breast cancer screening which employ risk-based approaches to address overdiagnosis, and AI-based technologies to catch more aggressive cancers earlier. Optimize post-diagnosis surveillance to improve adherence and ensure that any recurrences are caught as early as possible.

Biomarkers and prognostic factors have a powerful potential to predict the trajectory of a woman's experience with breast cancer. On their own, however, they can only explain a fraction of the mechanisms which affect treatment-related tumour behaviour and patient progression outcomes.[63] Yet, if we continue to pursue safe de-escalation of treatment as a worthwhile goal, we must be able to accurately stratify patients according to their disease risk and chances of benefiting from a given therapy.

For invasive breast cancer, selecting patients for chemotherapy based on a combination of clinicopathological characteristics and their genomic risk score remains encouraging. As precision medicine sits at the forefront of clinical oncology, gene expression profiles like the 70-gene signature will remain an important, cost-effective tool for identifying therapeutic strategies tailored to the individual. Prognostic factors for DCIS surgery de-escalation are still in the exploration phase. Early HTA has revealed promise in terms of cost-effectiveness and in the willingness of women to undergo an active surveillance strategy.

It is essential to remember that biomarkers that allow us to de-escalate treatment in certain patients also have a crucial role to play for other patients. They should help us escalate treatment for other higher-risk individuals to prevent recurrence and advanced disease, which would ultimately require costlier therapies. Regular surveillance imaging can also play a complementary role in finding recurrences earlier. Meanwhile, AI could be a promising technology that prevents missing cancers at screening, subsequently decreasing the number of advanced interval breast cancers.

This PhD acknowledges the challenges in finding and translating predictive biomarkers into clinical practice to inform decisions for women with breast cancer and DCIS and using AI technology without inducing further overdiagnosis among screen-detected cancer. Researchers must continue verifying the cost-effectiveness of approaches based on their use before and after prospective validation. Early HTA has earned itself an essential role in facilitating the adoption of technologies that improve health outcomes at the population level and bring cost savings to the healthcare system.

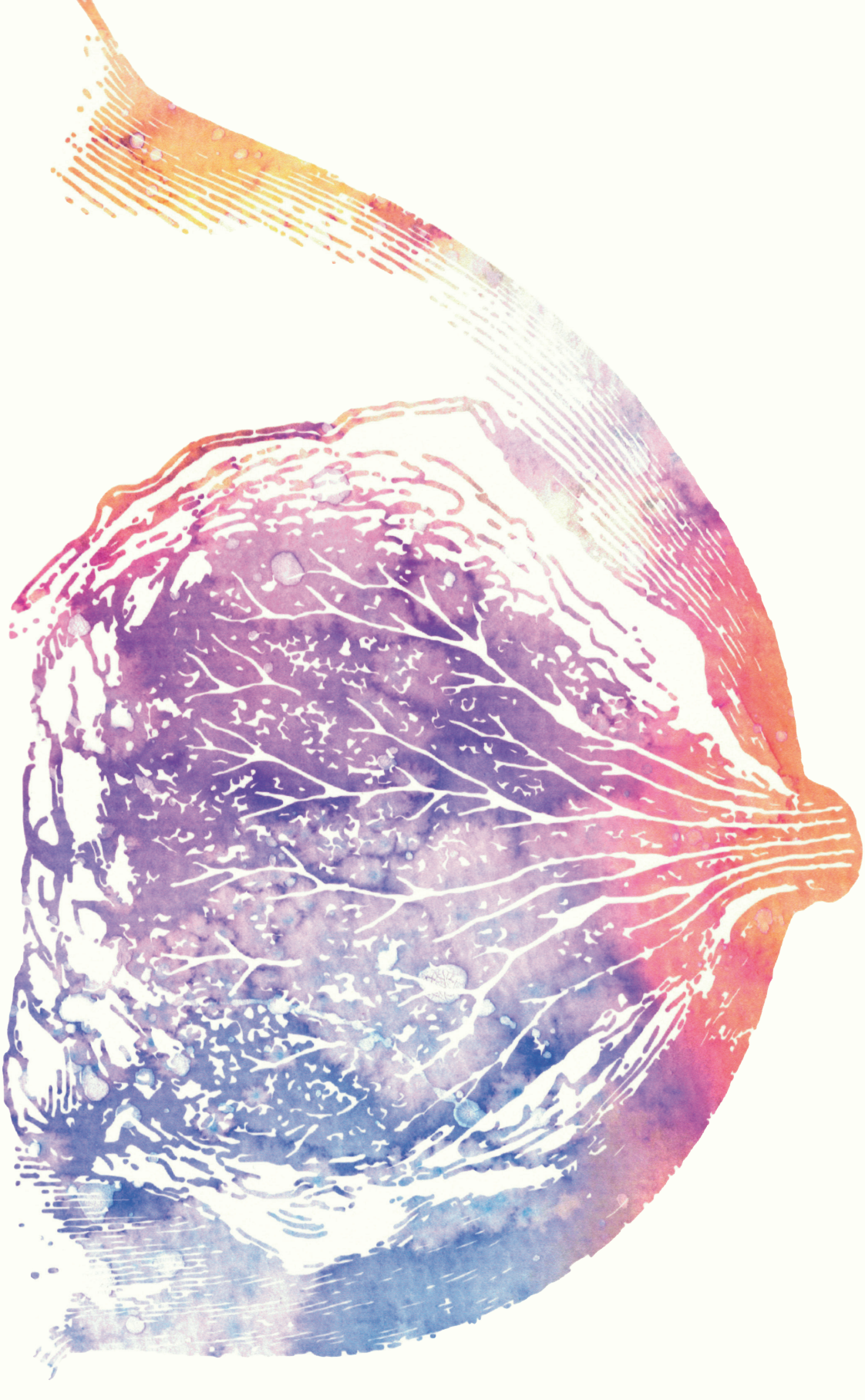
References

1. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc* 1999; **94**(446): 496-509.
2. Worni M, Akushevich I, Greenup R, et al. Trends in Treatment Patterns and Outcomes for Ductal Carcinoma In Situ. *J Natl Cancer Inst* 2015; **107**(12): djv263.
3. Elshof LE, Schmidt MK, Rutgers EJT, van Leeuwen FE, Wesseling J, Schaapveld M. Cause-specific Mortality in a Population-based Cohort of 9799 Women Treated for Ductal Carcinoma In Situ. *Ann Surg* 2018; **267**(5): 952-8.
4. Evans D. Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. *J Clin Nurs* 2003; **12**(1): 77-84.
5. Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. *Soc Sci Med* 2018; **210**: 2-21.
6. Monticciolo DL, Newell MS, Moy L, Niell B, Monsees B, Sickles EA. Breast cancer screening in women at higher-than-average risk: recommendations from the ACR. *J Am Coll Radiol* 2018; **15**(3): 408-14.
7. Kwan J, Croke J, Panzarella T, et al. Personalizing post-treatment cancer care: a cross-sectional survey of the needs and preferences of well survivors of breast cancer. *Curr Oncol* 2019; **26**(2): 138-46.
8. Grandjean I, Kwast AB, de Vries H, Klaase J, Schoevers WJ, Siesling S. Evaluation of the adherence to follow-up care guidelines for women with breast cancer. E*ur J Oncol Nurs* 2012; **16**(3): 281-5.
9. Draeger T, Voelkel V, Schreuder K, et al. Adherence to the Dutch Breast Cancer Guidelines for Surveillance in Breast Cancer Survivors: Real-World Data from a Pooled Multicenter Analysis. *Oncologist* 2022.
10. Visser LL, Groen EJ, van Leeuwen FE, Lips EH, Schmidt MK, Wesseling J. Predictors of an Invasive Breast Cancer Recurrence after DCIS: A Systematic Review and Meta-analyses. *Cancer Epidemiol Biomarkers Prev* 2019; **28**(5): 835-45.
11. Hill DA, Prossnitz ER, Royce M, Nibbe A. Temporal trends in breast cancer survival by race and ethnicity: A population-based cohort study. *PLoS One* 2019; **14**(10): e0224064.
12. Clarke CA, Keegan TH, Yang J, et al. Age-specific incidence of breast cancer subtypes: understanding the black-white crossover. *J Natl Cancer Inst* 2012; **104**(14): 1094-101.
13. Mootz A, Arjmandi F, Dogan BE, Evans WP. Health Care Disparities in Breast Cancer: The Economics of Access to Screening, Diagnosis, and Treatment. *J Breast Imaging* 2020; **2**(6): 524-9.
14. Rebner M, Pai VR. Breast cancer screening recommendations: African American women are at a disadvantage. *J Breast Imaging* 2020; **2**(5): 416-21.
15. Lu WL, Jansen L, Post WJ, Bonnema J, Van de Velde JC, De Bock GH. Impact on survival of early detection of isolated breast recurrences after the primary treatment for breast cancer: a meta-analysis. *Breast Cancer Res Treat* 2008; **114**(3): 403.
16. Bridges JF, Hauber AB, Marshall D, et al. Conjoint analysis applications in health--a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value Health* 2011; **14**(4): 403-13.
17. Elwyn G, Frosch D, Thomson R, et al. Shared decision making: a model for clinical practice. *J Gen Intern Med* 2012; **27**(10): 1361-7.

18. Harrison M, Milbers K, Hudson M, Bansback N. Do patients and health care providers have discordant preferences about which aspects of treatments matter most? Evidence from a systematic review of discrete choice experiments. *BMJ Open* 2017; **7**(5): e014719.
19. Katz SJ, Belkora J, Elwyn G. Shared decision making for treatment of cancer: challenges and opportunities. *J Oncol Pract* 2014; **10**(3): 206-8.
20. Ryser MD, Lange J, Inoue LYT, et al. Estimation of Breast Cancer Overdiagnosis in a U.S. Breast Screening Cohort. *Ann Intern Med* 2022; **175**(4): 471-8.
21. Meurs CJC, van Rosmalen J, Menke-Pluijmers MBE, et al. A prediction model for underestimation of invasive breast cancer after a biopsy diagnosis of ductal carcinoma in situ: based on 2892 biopsies and 589 invasive cancers. *Br J Cancer* 2018; **119**(9): 1155-62.
22. Mannu GS, Groen EJ, Wang Z, et al. Reliability of preoperative breast biopsies showing ductal carcinoma in situ and implications for non-operative treatment: a cohort study. *Breast Cancer Res Treat* 2019; **178**(2): 409-18.
23. Almekinders MMM, Schaapveld M, Thijssen B, et al. Breast adipocyte size associates with ipsilateral invasive breast cancer risk after ductal carcinoma in situ. *NPJ Breast Cancer* 2021; **7**(1): 31.
24. Shee K, Muller KE, Marotti J, Miller TW, Wells WA, Tsongalis GJ. Ductal carcinoma in situ biomarkers in a precision medicine era: current and future molecular-based testing. *Am J Pathol* 2019; **189**(5): 956-65.
25. Miquel-Cases A, Schouten PC, Steuten LM, Retèl VP, Linn SC, van Harten WH. (Very) Early technology assessment and translation of predictive biomarkers in breast cancer. *Cancer Treat Rev* 2017; **52**: 117-27.
26. Lips EH, Kumar T, Megalios A, et al. Genomic analysis defines clonal relationships of ductal carcinoma in situ and recurrent invasive breast cancer. *Nat Genet* 2022; **54**(6): 850-60.
27. Poste G. Bring on the biomarkers. *Nature* 2011; **469**(7329): 156-7.
28. Cardoso F, van't Veer LJ, Bogaerts J, et al. 70-gene signature as an aid to treatment decisions in early-stage breast cancer. *N Engl J Med* 2016; **375**(8): 717-29.
29. Sparano JA, Gray RJ, Makower DF, et al. Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer. *N Engl J Med* 2018; **379**(2): 111-21.
30. Curigliano G, Burstein HJ, Winer EP, et al. De-escalating and escalating treatments for early-stage breast cancer: the St. Gallen International Expert Consensus Conference on the Primary Therapy of Early Breast Cancer 2017. *Ann Oncol* 2017; **28**(8): 1700-12.
31. Krop I, Ismaila N, Andre F, et al. Use of Biomarkers to Guide Decisions on Adjuvant Systemic Therapy for Women With Early-Stage Invasive Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline Focused Update. *J Clin Oncol* 2017; **35**(24): 2838-47.
32. van Steenhoven JEC, Kuijer A, Schreuder K, et al. The Changing Role of Gene-Expression Profiling in the Era of De-escalating Adjuvant Chemotherapy in Early-Stage Breast Cancer. *Ann Surg Oncol* 2019; **26**(11): 3495-501.
33. Sparano JA, Gray RJ, Ravdin PM, et al. Clinical and genomic risk to guide the use of adjuvant therapy for breast cancer. *N Engl J Med* 2019; **380**(25): 2395-405.
34. Retèl VP, Bueno-de-Mesquita JM, Hummel MJ, et al. Constructive Technology Assessment (CTA) as a tool in coverage with evidence development: the case of the 70-gene prognosis signature for breast cancer diagnostics. *Int J Technol Assess Health Care* 2009; **25**(1): 73-83.

35. Retèl VP, Joore MA, Knauer M, Linn SC, Hauptmann M, Harten WH. Cost-effectiveness of the 70-gene signature versus St. Gallen guidelines and Adjuvant Online for early breast cancer. *Eur J Cancer* 2010; **46**(8): 1382-91.
36. Drukker CA, Bueno-de-Mesquita JM, Retèl VP, et al. A prospective evaluation of a breast cancer prognosis signature in the observational RASTER study. *Int J Cancer* 2013; **133**(4): 929-36.
37. Retèl VP, Joore MA, Drukker CA, et al. Prospective cost-effectiveness analysis of genomic profiling in breast cancer. *Eur J Cancer* 2013; **49**(18): 3773-9.
38. Ijzerman MJ, Steuten LM. Early assessment of medical technologies to inform product development and market access: a review of methods and applications. *Appl Health Econ Health Policy* 2011; **9**(5): 331-47.
39. Vallejo-Torres L, Steuten LM, Buxton MJ, Girling AJ, Lilford RJ, Young T. Integrating health economics modeling in the product development cycle of medical devices: a Bayesian approach. *Int J Technol Assess Health Care* 2008; **24**(4): 459-64.
40. Sculpher M, Drummond M, Buxton M. The Iterative Use of Economic Evaluation as Part of the Process of Health Technology Assessment. *J Health Serv Res Policy* 1997; **2**(1): 26-30.
41. Ijzerman MJ, Koffijberg H, Fenwick E, Krahn M. Emerging Use of Early Health Technology Assessment in Medical Product Development: A Scoping Review of the Literature. *PharmacoEconomics* 2017; **35**(7): 727-40.
42. Burstein HJ, Curigliano G, Thürlimann B, et al. Customizing local and systemic therapies for women with early breast cancer: the St. Gallen International Consensus Guidelines for treatment of early breast cancer 2021. *Ann Oncol* 2021; **32**(10): 1216-35.
43. Andre F, Ismaila N, Allison KH, et al. Biomarkers for Adjuvant Endocrine and Chemotherapy in Early-Stage Breast Cancer: ASCO Guideline Update. *J Clin Oncol* 2022; **40**(16): 1816-37.
44. Cardoso F, Kyriakides S, Ohno S, et al. Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2019; **30**(8): 1194-220.
45. Gradishar WJ, Moran MS, Abraham J, et al. NCCN guidelines® insights: Breast cancer, version 4.2021: Featured updates to the NCCN guidelines. *JNCCN* 2021; **19**(5): 484-93.
46. Piccart M, van 't Veer LJ, Poncet C, et al. 70-gene signature as an aid for treatment decisions in early breast cancer: updated results of the phase 3 randomised MINDACT trial with an exploratory analysis by age. *Lancet Oncol* 2021; **22**(4): 476-88.
47. Burstein HJ, Curigliano G, Loibl S, et al. Estimating the benefits of therapy for early-stage breast cancer: the St. Gallen International Consensus Guidelines for the primary therapy of early breast cancer 2019. *Ann Oncol* 2019; **30**(10): 1541-57.
48. Burstein HJ, Lacchetti C, Anderson H, et al. Adjuvant endocrine therapy for women with hormone receptor–positive breast cancer: American Society of Clinical Oncology clinical practice guideline update on ovarian suppression. *J Clin Oncol* 2016; **34**(14): 1689-701.
49. Chlebowski RT, Kim J, Haque R. Adherence to endocrine therapy in breast cancer adjuvant and prevention settings. *Cancer Prev Res (Phila)* 2014; **7**(4): 378-87.
50. Pistilli B, Paci A, Ferreira AR, et al. Serum Detection of Nonadherence to Adjuvant Tamoxifen and Breast Cancer Recurrence Risk. *J Clin Oncol* 2020; **38**(24): 2762-72.
51. Condorelli R, Vaz-Luis I. Managing side effects in adjuvant endocrine therapy for breast cancer. *Expert Rev Anticancer Ther* 2018; **18**(11): 1101-12.
52. Piccart MJ, Hilbers FS, Bliss JM, et al. Road Map to Safe and Well-Designed De-escalation Trials of Systemic Adjuvant Therapy for Solid Tumors. *J Clin Oncol* 2020; **38**(34): 4120-9.
53. Lopes Cardozo JMN, Schmidt MK, van 't Veer LJ, et al. Combining method of detection and 70-gene signature for enhanced prognostication of breast cancer. *Breast Cancer Res Treat* 2021; **189**(2): 399-410.

54. Leibig C, Brehmer M, Bunk S, Byng D, Pinker K, Umutlu L. Combining the strengths of radiologists and AI for breast cancer screening: a retrospective analysis. *Lancet Digit Health* 2022; **4**(7): e507-e19.
55. Lehman CD, Wellman RD, Buist DS, Kerlikowske K, Tosteson AN, Miglioretti DL. Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection. *JAMA Intern Med* 2015; **175**(11): 1828-37.
55. Pickles K, Hersch J, Nickel B, Vaidya JS, McCaffery K, Barratt A. Effects of awareness of breast cancer overdiagnosis among women with screen-detected or incidentally found breast cancer: a qualitative interview study. *BMJ Open* 2022; **12**(6): e061211.
56. Esserman LJ, Shieh Y, Rutgers EJ, et al. Impact of mammographic screening on the detection of good and poor prognosis breast cancers. *Breast Cancer Res Treat* 2011; **130**(3): 725-34.
57. Drukker CA, Schmidt MK, Rutgers EJT, et al. Mammographic screening detects low-risk tumor biology breast cancers. *Breast Cancer Res Treat* 2014; **144**(1): 103-11.
58. Esserman LJ, Anton-Culver H, Borowsky A, et al. The WISDOM Study: breaking the deadlock in the breast cancer screening debate. *NPJ Breast Cancer* 2017; **3**(1): 34.
59. Shieh Y, Eklund M, Madlensky L, et al. Breast Cancer Screening in the Precision Medicine Era: Risk-Based Screening in a Population-Based Trial. *J Natl Cancer Inst* 2017; **109**(5).
60. Yala A, Mikhael PG, Strand F, et al. Multi-Institutional Validation of a Mammography-Based Breast Cancer Risk Model. *J Clin Oncol* 2022; **40**(16): 1732-40.
61. Roux A, Cholerton R, Sicsic J, et al. Study protocol comparing the ethical, psychological and socio-economic impact of personalised breast cancer screening to that of standard screening in the "My Personal Breast Screening" (MyPeBS) randomised clinical trial. *BMC Cancer* 2022; **22**(1): 507.
62. French DP, Astley S, Brentnall AR, et al. What are the benefits and harms of risk stratified screening as part of the NHS breast screening Programme? Study protocol for a multi-site non-randomised comparison of BC-predict versus usual screening (NCT04359420). *BMC Cancer* 2020; **20**(1): 570.
63. Brooks JD, Nabi HH, Andrulis IL, et al. Personalized Risk Assessment for Prevention and Early Detection of Breast Cancer: Integration and Implementation (PERSPECTIVE I&I). *J Pers Med* 2021; **11**(6).
64. Rainey L, van der Waal D, Broeders MJM. Dutch women's intended participation in a risk-based breast cancer screening and prevention programme: a survey study identifying preferences, facilitators and barriers. *BMC Cancer* 2020; **20**(1): 965.
65. Rainey L, van der Waal D, Jervaeus A, et al. European women's perceptions of the implementation and organisation of risk-based breast cancer screening and prevention: a qualitative study. *BMC Cancer* 2020; **20**(1): 247.
66. Borst P, Wessels L. Do predictive signatures really predict response to cancer chemotherapy? *Cell Cycle* 2010; **9**(24): 4836-40.



Annex

Summary

This PhD dissertation aims to understand the factors that may affect the use of interventions that lend themselves to de-escalating low-value treatment for early breast cancer and ductal carcinoma in situ (DCIS). The chapters cover three complimentary themes in the management of early-stage breast cancer. The first theme focuses on screen-detected primary DCIS, with select chapters characterizing disease etiology, treatment and surveillance outcomes, real-world health care utilization, and potential of biomarkers to select low-risk women for an active surveillance strategy. Research was performed within the PRECISION (PREvent ductal Carcinoma In Situ Invasive Overtreatment Now) Consortium. Issues such as overtreatment and the willingness to harness the potential of new prognostic technologies and biomarkers to guide treatment decisions are also covered by projects within this PhD. The second major theme focuses on treatment de-escalation for early-stage breast cancer, based on the first results of the EORTC 10041/BIG 3-04 MINDACT (Microarray in Node-Negative and 1 to 3 Positive Lymph Node Disease May Avoid Chemotherapy) phase 3 randomized control trial of the 70-gene signature. Finally, a complementary final theme and chapter highlights a promising new technology: artificial intelligence for to improve cancer detection at breast cancer screening to decrease the interval cancer rate.

Through collaborations with individuals from across the spectrum of care, this dissertation explores new approaches to optimizing treatment and follow-up care from a multitude of perspectives. It applies elements of health technology assessment: from mathematical modelling of disease processes and treatment outcomes, to gathering insights from patients and providers. Through producing this evidence, the final aim of this PhD is to inform decision makers about the many pathways towards optimizing care for women with early-stage breast cancer and DCIS.

The potential of an active surveillance strategy for low-risk DCIS

Chapter 2 introduces the disease etiology of DCIS and life years associated with different treatment strategies using a statistical modeling approach called multi-state modeling. The multi-state model forms the basis for future cost-effectiveness modeling, and is based on the principle that diseases can progress through several possible stages. Women begin their journey at diagnosis, with some experiencing disease progression which may take different possible manifestations. The model covers disease-related processes until death from the disease itself or another cause. We refer to these disease stages as *health states*, and an individual experiencing a disease can be in any given state at any given time. Their progression to different states or time spent in a state depends on a combination of many factors,

which are unique to each individual. This depends on e.g., one's age, the treatment one receives, down to the molecular-level characteristics of their disease. In cost-effectiveness modeling, we use information about time spent in, and transitions between health states. Each health state is associated with specific costs, and can be modified by the quality-of-life associated with each.

The real-world results from the multi-state modeling reveal that women with low-risk DCIS features demonstrate minimal differences by treatment strategy in the probability of surviving ipsilateral invasive breast cancer-free at 5 and 10 years. This study provides evidence beyond previously published studies using trial-based and observational data which provided limited direct comparison of no treatment and standard interventional treatment strategies. With this article, we provide an opportunity beyond prospective clinical trials to understand DCIS and the potential impact of an active surveillance strategy.

Chapter 3 covers a multi-institutional study based on US-based data from the National Cancer Database DCIS Special Study that investigates imaging surveillance in women following breast conservation treatment for DCIS. We analyzed data from over 12,000 patients with the aim to identify factors that influence regular imaging surveillance, particularly sociodemographic and clinical factors. Almost 50% of women with DCIS did not adhere to imaging surveillance guidelines over 5 years following breast conservation treatment, and non-adherence to early surveillance was associated with a delay in the detection of invasive recurrence. While more surveillance was associated with higher diagnosis of ipsilateral invasive breast cancer, it was also associated with racial and ethnic factors, private insurance, and receipt of adjuvant therapy. Women with more resources or who were motivated to have adjuvant therapy were more likely to follow surveillance imaging guidelines. This study highlights race and ethnicity disparities in care which may be due to limited access and/or longstanding racial/ethnic inequities in the United States. Establishing a regular pattern of surveillance soon after diagnosis may promote timely detection of ipsilateral recurrence in the long term.

Chapter 4 provides an opportunity to understand how acceptable de-escalation strategies are for recently diagnosed women with low-risk DCIS in the Netherlands. Preferences for treatment strategies for low-risk DCIS, including a new active surveillance strategy, were elicited with a discrete choice experiment among recently-diagnosed women and oncologists involved in the care of women with DCIS. 172 women participating in a prospective active surveillance trial for DCIS, and 30 radiation and surgical oncologists involved in the care of women with DCIS completed

the experiment. Patients exhibited strong preferences for active surveillance and seemed prepared to accept much higher levels of 10-year risk of developing ipsilateral invasive breast cancer than oncologists. Both patients and oncologists showed a strong aversion toward more extensive locoregional treatments (i.e., breast conserving surgery followed by radiotherapy, and mastectomy), while both groups demonstrated a strong preference toward shorter follow-up intervals. We report on surprising discrepancies in preferences, especially related to weighing the risk of recurrence. If an active surveillance strategy is deemed safe and effective based on the findings of the ongoing prospective active surveillance studies or in the future, incorporating patients' preferences in treatment decision making will serve to improve treatment compliance and satisfaction.

Chapter 5 draws together all the findings presented in the preceding chapters 2-4. We characterize the costs and quality-adjusted health outcomes associated with (non)-interventional strategies for women with low-risk DCIS. A semi-Markov model was constructed based on the multi-state modelling approach employed in chapter 2. The cost-effectiveness analysis explores two opportunities for selecting low-risk women with primary DCIS who could opt for an active surveillance strategy. The base-case model uses standard pathological information on DCIS grade (low-to-intermediate) and estrogen-receptor-positive status, similar to the eligibility criteria in the LORD trial described in Chapter 4. 50% of all women with screen-detected primary DCIS have low-risk features based on these criteria, and would be eligible for the active surveillance strategy. In the scenario analysis, models used information on COX-2 protein expression and breast adipocyte size instead to select low-risk women to forgo surgery. In this scenario, a smaller proportion of women would be deemed eligible to forgo surgery, but would have a 10-year risk of ipsilateral invasive breast cancer similar to the general population. The results were presented both as incremental life years gained, and quality-adjusted life years (QALYs) gained to provide an important contrast when applying utilities. Whereas incremental life years are reflected in the differences in overall survival between groups, QALYs adjust this metric by incorporating the impact of patient preference and resulting quality of life. In the base-case model, introducing an active surveillance strategy would result in life years lost across the cohort. The application of utilities however positively shifts the health effects towards an average QALY gain. Strategies involving active surveillance for low-risk women were cost-saving. A headroom analysis based on a simulation of the impact of a hypothetical perfect biomarker using COX-2 and adipose area information was also performed. For such a biomarker-based strategy to remain cost-effective at a WTP threshold of €20,000, the upper ceiling price for this could be set at €6,227.

The potential of the 70-gene signature to guide treatment strategies for low-risk early breast cancer

Chapter 6 similarly describes how minimal survival differences are balanced against quality-of-life gains per patient and considerable cost savings when using the 70-gene signature compared to clinical assessment in guiding treatment decisions for adjuvant chemotherapy. We report the results of a cost-effectiveness and budget impact analysis of treatment strategies guided by the 70-gene signature versus treatment decisions based on clinical risk assessment alone for a target group of patients with estrogen receptor-positive, HER2-negative early breast cancer. The analysis is based on patient-level outcome data from the MINDACT trial, information on breast cancer-specific quality of life, as well as costs for six countries: Belgium, France, Germany, the Netherlands, the United Kingdom, and the United States. A hybrid decision tree-Markov model simulated treatment strategies in accordance with the 70-gene signature with clinical assessment versus clinical assessment alone, over a 10-year time horizon. Primary outcomes were quality-adjusted life years (QALYs), country-specific costs and incremental cost-effectiveness ratios (ICERs) for each country. For all six countries, the 70-gene signature was found cost-effective. For five out of six countries, it was also found to be cost-saving.

Chapter 7 looks more closely at a select group of women in the MINDACT trial to better understand the added value of endocrine therapy among those with estrogen receptor-positive, HER2-negative, lymph node-negative tumours ≤ 2 cm. Most clinical guidelines recommend endocrine therapy for these women, despite their very low risk of distant metastasis. Within this study, we identified a subgroup of N=509 women who received no adjuvant systemic therapy following surgery, and matched them 1:1 to women with similar clinical characteristics who had received adjuvant endocrine therapy. The 8-year distant metastasis free interval rate in the untreated group was 94.8%, 2.5% lower than the matched group who received endocrine therapy. However, the cumulative incidence of locoregional recurrences and contralateral breast cancers at 8 years was 3% higher in patients who received no adjuvant treatment compared to patients who received endocrine therapy, contributing to an 8.3% difference in the breast cancer-free interval at 8 years when considering all breast cancer events together. The chapter discuses the associated overall poor quality of life and heightened risk of treatment-related adverse events associated with endocrine therapy, illustrating important considerations surrounding treatment de-escalation for low-risk women.

The potential of artificial intelligence to improve cancer detection on mammography

Chapter 8 covers a study on future perspectives incorporating artificial intelligence (AI)-based technology for optimization of breast cancer screening. This was conducted on data from the German national breast cancer screening program, in collaboration with Vara (MX Healthcare GmbH) and the North Mammography Reference Center in Oldenburg, Germany North Mammography Reference Center in Oldenburg, Germany, during the latter part of the PhD study time frame. The study was deemed fitting to include in the thesis by the University of Twente supervising team in view of the topics concerned. This study sought to understand how many retrospectively visible cancers that were overlooked by radiologists could be detected and found by AI. A cohort of N=2,396 interval cancers were used for this study, 842 of them were considered retrospectively visible. When the algorithm was set at an operating level to approximate the specificity of the screening program, it could detect up to 40% of missed cancers. Evaluating any AI at this specificity is necessary to demonstrate how AI can be brought into a screening program without inducing a detrimental impact on false positive rates. This is important to women participating in screening, as the emotional burden of unnecessary recalls can result in fewer women consistently attending screening. Furthermore, the interval cancer rate is an important quality indicator for breast cancer screening. Compared to screen-detected cancers, interval cancers are more likely to have unfavorable prognosis, requiring more intensive treatment. AI-augmented breast cancer screening can provide value to the healthcare system, as breast cancers detected earlier have better long-term outcomes and can be less invasive to treat, as illustrated throughout chapters 2 to 7.

Discussion and conclusion

The final chapter of the PhD provides an overview of further research, policy and practice implications resulting from each of the chapters. It concludes with three major opportunities that have emerged as future directions for optimizing surveillance, diagnosis, and treatment of early breast cancer and DCIS. Opportunity Number 1: Apply risk assessment at the point of diagnosis, and possibly at the start of the breast cancer screening journey. De-escalate treatment for low-risk women to maintain their quality of life, and to free-up resources to find meaningful treatment for high-risk women, and advanced and metastatic cancer. Opportunity Number 2: Let fewer cancers slip through screening and surveillance undetected. Use new technologies during screening that catch more aggressive cancers earlier.

Optimize post-diagnosis surveillance to improve adherence and ensure that any recurrences are caught as early as possible.

This PhD acknowledges the challenges in finding and translating predictive biomarkers into clinical practice to inform decisions for women with breast cancer and DCIS and using AI technology without inducing further overdiagnosis among screen-detected cancer. Researchers must continue verifying the cost-effectiveness of approaches based on their use before and after prospective validation. Early HTA has earned itself an essential role in facilitating the adoption of technologies that improve health outcomes at the population level and bring cost savings to the healthcare system.

Samenvatting

Dit proefschrift heeft tot doel inzicht te krijgen in de factoren die van invloed kunnen zijn op het gebruik van interventies die zich lenen voor het de-escaleren van laagwaardige, of onzinnige behandeling voor borstkanker in een vroeg stadium en ductaal carcinoma in situ (DCIS). Dit proefschrift behandelt drie complementaire thema's. Het eerste thema richt zich op screen-gedetecteerde primaire DCIS, met geselecteerde hoofdstukken die de etiologie van de ziekte, de resultaten van behandeling en monitoring (nacontrole), het gebruik van gezondheidszorg in de praktijk, en het potentieel van biomarkers om vrouwen met een laag risico te selecteren voor een actieve surveillance strategie. Het onderzoek werd uitgevoerd binnen het PRECISION (PREvent ductal Carcinoma In Situ Invasive Overtreatment Now) Consortium. Kwesties zoals overbehandeling en de bereidheid om het potentieel van nieuwe prognostische technologieën en biomarkers te benutten om beslissingen over behandeling te sturen, komen ook aan bod in projecten binnen dit proefschrift. Het tweede grote thema richt zich op de-escalatie van behandeling voor borstkanker in een vroeg stadium, op basis van de esultaten van de EORTC 10041/BIG 3-04 MINDACT (Microarray in Node-Negative and 1 to 3 Positive Lymph Node Disease May Avoid Chemotherapy) gerandomiseerde fase 3 studie van de 70-gene signature. Ten derde wordt in een aanvullend hoofdstuk de aandacht gevestigd op een veelbelovende nieuwe technologie: kunstmatige intelligentie voor het verbeteren van de opsporing van kanker bij borstkankerscreening om het percentage intervalkanker te verlagen.

Door samen te werken met mensen uit het hele spectrum van de zorg, verkent dit proefschrift nieuwe benaderingen voor het optimaliseren van screening, behandeling en nazorg. Het past elementen toe van Health Technology Assessment: van wiskundige modellering van ziekteprocessen en behandeleffecten, tot het verzamelen van voorkeuren van patiënten en zorgverleners. Het uiteindelijke doel van dit proefschrift is om besluitvormers te informeren over de vele manieren waarop de zorg voor vrouwen met borstkanker in een vroeg stadium en DCIS kan worden geoptimaliseerd.

Het potentieel van een actieve monitoring strategie voor laag-risico DCIS

Hoofdstuk 2 introduceert de ziekte-etiologie van DCIS en de 'levensjaren' geassocieerd met verschillende behandelingsstrategieën met behulp van een statistische modelbenadering die multi-state modeling wordt genoemd. Het multi-state model vormt de basis voor toekomstige kosteneffectiviteitsmodellen en is gebaseerd op het principe dat ziekten zich in verschillende stadia kunnen

ontwikkelen. Patiënten beginnen hun traject bij de diagnose, waarbij sommigen een ziekteprogressie doormaken die verschillende mogelijke verschijningsvormen kan aannemen. Het model bestrijkt ziektegerelateerde processen tot het overlijden aan de ziekte zelf of aan een andere oorzaak. We noemen deze ziektestadia gezondheidstoestanden, en een persoon die een ziekte doormaakt kan zich op elk moment in een bepaalde toestand bevinden. De progressie naar verschillende toestanden of de tijd die iemand in een toestand doorbrengt, hangt af van een combinatie van vele factoren, die voor elk individu uniek zijn. Dit hangt bijvoorbeeld af van iemands leeftijd, de behandeling die hij of zij krijgt, tot aan de kenmerken op moleculair niveau van zijn of haar ziekte. In kosteneffectiviteitsmodellen gebruiken we informatie over de tijd die iemand doorbrengt in een gezondheidstoestand en de overgang tussen gezondheidstoestanden. Elke gezondheidstoestand wordt gaat gepaard met specifieke kosten, en een specifieke levenskwaliteit.

De resultaten van de multi-state modellering op basis van echte data van Amerikaanse vrouwen laten zien dat vrouwen met laag risico DCIS-kenmerken minimale verschillen vertonen in overleving van ipsilaterale invasieve borstkanker-vrij op 5 en 10 jaar per behandelingsstrategie (monitoring versus actieve behandeling). Deze studie levert bewijs dat verder gaat dan eerder gepubliceerde studies die gebruik maken van op studie gebaseerde- en observationele gegevens die een beperkte directe vergelijking van geen behandeling versus standaard interventionele behandelingsstrategieën leverden. Met dit artikel bieden wij de mogelijkheid om, naast prospectieve klinische studies, inzicht te krijgen in DCIS en de potentiële impact van een actieve monitoring strategie.

Hoofdstuk 3 behandelt een multi-institutionele studie gebaseerd op Amerikaanse gegevens van de "National Cancer Database DCIS Special Study" die het opvolgen van monitoring onderzoekt bij vrouwen na borst besparende behandeling voor DCIS. Wij analyseerden gegevens van meer dan 12.000 patiënten met als doel om factoren te identificeren die regelmatige controle van beeldvorming beïnvloeden, met name sociodemografische en klinische factoren. Bijna 50% van de vrouwen met DCIS hield zich niet aan de richtlijnen voor beeldvormende controle gedurende 5 jaar na een borstbesparende behandeling, en het niet volgen van vroege controle was geassocieerd met een vertraging in de opsporing van een terugkeer van ziekte. Hoewel meer surveillance geassocieerd was met een hogere diagnose van ipsilaterale invasieve borstkanker, was dit ook geassocieerd met raciale en etnische factoren, particuliere verzekering, en ontvangst van adjuvante therapie. Vrouwen in een hogere sociaaleconomische klasse of die gemotiveerd waren om adjuvante therapie te ondergaan, hadden meer kans om de richtlijnen voor beeldvorming

te volgen. Deze studie belicht de verschillen in zorg tussen ras en etniciteit, die te wijten kunnen zijn aan beperkte toegang en/of reeds lang bestaande raciale/ etnische ongelijkheden in de Verenigde Staten. Het instellen van een regelmatig controlepatroon kort na de diagnose kan de tijdige opsporing van ipsilateraal recidief op de lange termijn bevorderen.

Hoofdstuk 4 biedt een mogelijkheid om te begrijpen hoe acceptabel de-escalatie strategieën zijn voor recent gediagnosticeerde vrouwen met laag-risico DCIS in Nederland. Voorkeuren voor behandelingsstrategieën voor laag-risico DCIS, inclusief een nieuwe actieve surveillance strategie, werden bepaald met een discrete keuze experiment onder recent gediagnosticeerde vrouwen en oncologen betrokken bij de zorg voor vrouwen met DCIS. 172 vrouwen die deelnamen aan een prospectieve actieve surveillance studie voor DCIS, en 30 radiotherapeutische en chirurgische oncologen die betrokken zijn bij de zorg voor vrouwen met DCIS, vulden het experiment in. Patiënten vertoonden een sterke voorkeur voor actieve monitoring en leken bereid om veel hogere niveaus van 10-jaars risico op het ontwikkelen van ipsilaterale invasieve borstkanker te accepteren dan oncologen. Zowel patiënten als oncologen toonden een sterke afkeer van uitgebreidere locoregionale behandelingen (d.w.z. borstsparende chirurgie gevolgd door radiotherapie, en mastectomie), terwijl beide groepen een sterke voorkeur toonden voor meer controle afspraken. Wij hebben deze voorkeuren in een vroeg stadium gerapporteerd (de klinische studie is nog niet afgerond) . Als een actieve surveillance strategie veilig en effectief wordt geacht op basis van de bevindingen van de lopende studie, zal het meenemen van de voorkeuren van patiënten in de besluitvorming over de behandeling dienen om de therapietrouw en -tevredenheid te verbeteren.

In **Hoofdstuk 5** worden alle bevindingen uit de voorgaande hoofdstukken 2-4 samengevoegd. We karakteriseren de kosten en de voor kwaliteit gecorrigeerde gezondheidsuitkomsten van (non)-interventionele strategieën voor vrouwen met laag-risico DCIS. Een semi-Markov model werd geconstrueerd op basis van de multi-state modelling benadering die in hoofdstuk 2 werd gebruikt. De kosteneffectiviteitsanalyse verkent twee mogelijkheden voor het selecteren van vrouwen met laag risico met primair DCIS die zouden kunnen kiezen voor een actieve surveillance strategie. Het basisscenario model maakt gebruik van standaard pathologische informatie over DCIS graad (laag tot gemiddeld) en oestrogeen-receptor-positieve status, vergelijkbaar met de geschiktheidscriteria in de LORD trial beschreven in hoofdstuk 4. 50% van alle vrouwen met screen-ontdekte primaire DCIS hebben op basis van deze criteria kenmerken met een laag risico, en zouden in aanmerking komen voor de actieve surveillance strategie. In de scenarioanalyse gebruikten de modellen informatie over COX-2 protein expression

en de grootte van borstadipocyten om vrouwen met een laag risico te selecteren voor chirurgie. In dit scenario zou een kleiner deel van de vrouwen in aanmerking komen voor actieve monitoring, maar zou het tienjaarsrisico van ipsilaterale invasieve borstkanker vergelijkbaar zijn met dat van de algemene bevolking. De resultaten werden zowel in gewonnen 'levensjaren' als in gewonnen voor 'kwaliteit gecorrigeerde levensjaren' (QALYs) gepresenteerd. Terwijl het marginale verschil in 'levensjaren' tot uiting komt tussen de groepen, wordt deze metriek in de QALYs aangepast door rekening te houden met de invloed van de voorkeur van de patiënt en de daaruit voortvloeiende levenskwaliteit. In het basisscenario zou de invoering van een actieve bewakingsstrategie leiden tot een verlies van levensjaren alleen. Als deze vervolgens gecorrigeerd worden door de kwaliteit van leven, verschuiven de gezondheidseffecten echter in positieve zin naar een gemiddelde QALY-winst. Naast de QALY winst werd ook geconcludeerd dat de monitoring strategie voor vrouwen met een laag risico DCIS kostenbesparend was. Er werd ook een scenario analyse uitgevoerd van het effect van een hypothetische perfecte biomarker op basis van COX-2 en informatie over het vetweefsel. Opdat een dergelijke strategie op basis van biomarkers kosteneffectief zou blijven bij een WTP-drempel van 20.000 euro, kon de maximumprijs hiervoor worden vastgesteld op 6.227 euro.

Het potentieel van de 70-genen signature als leidraad voor behandelingsstrategieën voor laag-risico borstkanker in een vroeg stadium

Hoofdstuk 6 beschrijft op vergelijkbare wijze hoe minimale overlevingsverschillen worden afgewogen tegen winst in kwaliteit van leven per patiënt en aanzienlijke kostenbesparingen bij gebruik van de 70-genen signature in vergelijking met klinische beoordeling bij het begeleiden van behandelbeslissingen voor adjuvante chemotherapie. Wij rapporteren de resultaten van een kosten-effectiviteits- en budgetimpactanalyse van behandelingsstrategieën geleid door de 70-genen signature in vergelijking met behandelingsbeslissingen alleen gebaseerd op klinische risicobeoordeling voor een doelgroep van patiënten met oestrogeen receptor-positieve, HER2-negatieve vroege borstkanker. De analyse is gebaseerd op uitkomstgegevens op patiëntniveau van de MINDACT trial, informatie over borstkankerspecifieke kwaliteit van leven, alsmede kosten voor zes landen: België, Frankrijk, Duitsland, Nederland, het Verenigd Koninkrijk, en de Verenigde Staten. Een hybride decision tree-Markov model simuleerde behandelingsstrategieën in overeenstemming met de 70-genen signature met klinische beoordeling versus klinische beoordeling alleen, over een tijdshorizon van 10 jaar. Primaire uitkomsten waren voor 'kwaliteit gecorrigeerde levensjaren' (QALYs), landspecifieke kosten en incrementele kosteneffectiviteitsratio's (ICERs) voor elk land. Voor alle zes landen

werd de 70-genen signature kosteneffectief bevonden. Voor vijf van de zes landen bleek het ook kostenbesparend te zijn.

Hoofdstuk 7 wordt nader ingegaan op een selecte groep vrouwen in de MINDACT trial om een beter inzicht te krijgen in de toegevoegde waarde van hormonale therapie bij vrouwen met oestrogeen receptor-positieve, HER2-negatieve, lymfeklier-negatieve tumoren ≤ 2 cm. De meeste klinische richtlijnen raden hormonale therapie aan voor deze vrouwen, ondanks hun zeer lage risico op afstandsmetastasen. In deze studie identificeerden wij een subgroep van N=509 vrouwen die geen adjuvante systemische therapie kregen na chirurgie, en koppelden hen 1:1 aan vrouwen met vergelijkbare klinische karakteristieken die adjuvante hormonale therapie hadden gekregen. Het percentage metastasevrije intervallen na 8 jaar in de onbehandelde groep was 94,8%, 2,5% lager dan in de gematchte groep die hormonale therapie had gekregen. De cumulatieve incidentie van locoregionale recidieven en contralaterale borstkankers na 8 jaar was echter 3% hoger bij patiënten die geen adjuvante behandeling kregen in vergelijking met patiënten die endocriene therapie kregen, wat bijdroeg aan een verschil van 8,3% in het borstkankervrije interval na 8 jaar wanneer alle borstkankergebeurtenissen samen worden genomen. Het hoofdstuk gaat in op de geassocieerde algehele slechte kwaliteit van leven en het verhoogde risico op behandelingsgerelateerde ongewenste bijwerkingen die gepaard gaan met hormonale therapie, en illustreert belangrijke overwegingen met betrekking tot de-escalatie van behandeling voor vrouwen met een laag risico.

Het potentieel van kunstmatige intelligentie om de opsporing van kanker bij mammografie te verbeteren

Hoofdstuk 8 behandelt een studie naar de toekomstperspectieven met behulp van kunstmatige intelligentie (AI) technologie voor de optimalisering van borstkankerscreening. Dit is uitgevoerd op gegevens van het Duitse nationale borstkankerscreeningsprogramma, in samenwerking met Vara (MX Healthcare GmbH) en het North Mammography Reference Center in Oldenburg, Duitsland, tijdens het laatste deel van de looptijd van het promotieonderzoek. De studie werd door het begeleidingsteam van de Universiteit Twente geschikt geacht om in het proefschrift op te nemen gezien de betrokken onderwerpen. Deze studie trachtte te begrijpen hoeveel retrospectief zichtbare kankers die door radiologen over het hoofd werden gezien, door AI konden worden opgespoord en gevonden. Een cohort van N=2.396 intervalkankers werd gebruikt voor deze studie, 842 daarvan werden beschouwd als retrospectief zichtbaar. Wanneer het algoritme werd ingesteld op een werkingsniveau dat de specificiteit van het screeningsprogramma benaderde, kon het tot 40% van de gemiste kankers opsporen. Het evalueren van AI op deze specificiteit is noodzakelijk om aan te tonen hoe AI in een screeningsprogramma

kan worden geimplementeerd zonder een nadelig effect op de fout-positieve percentages. Dit is belangrijk voor vrouwen die aan screening deelnemen, aangezien de emotionele belasting van onnodige terugroepingen ertoe kan leiden dat minder vrouwen consequent aan screening deelnemen. Bovendien is het intervalkankercijfer een belangrijke kwaliteitsindicator voor borstkankerscreening. In vergelijking met screen-opgespoorde kankers hebben intervalkankers vaker een ongunstige prognose, zodat een intensievere behandeling nodig is. Screening op borstkanker met behulp van AI kan waarde toevoegen aan het gezondheidszorgsysteem, aangezien borstkankers die eerder worden ontdekt betere resultaten op lange termijn hebben en minder invasief te behandelen zijn, zoals wordt geïllustreerd in de hoofdstukken 2 tot 7.

Discussie en conclusie

Het laatste hoofdstuk van het proefschrift geeft een overzicht van verdere implicaties voor onderzoek, beleid en praktijk die uit elk van de hoofdstukken voortvloeien. Het sluit af met twee belangrijke aanbevelingen die naar voren zijn gekomen als toekomstige richtingen voor het optimaliseren van screening, diagnose en behandeling van borstkanker en DCIS in een vroeg stadium. Aanbeveling nummer 1: Pas risicobeoordeling toe op het punt van de diagnose, en mogelijk aan het begin van het traject van borstkankerscreening. De behandeling voor vrouwen met een laag risico de-escaleren om hun levenskwaliteit te handhaven en middelen vrij te maken om een zinvolle behandeling te vinden voor vrouwen met een hoog risico en gevorderde en uitgezaaide kanker. Aanbeveling nummer 2: Minder gevallen van kanker onopgemerkt door screening en toezicht laten ontsnappen. Gebruik nieuwe technologieën tijdens screening die agressievere vormen van kanker eerder opsporen. Optimaliseer de controle na de diagnose om de therapietrouw te verbeteren en ervoor te zorgen dat eventuele recidieven zo vroeg mogelijk worden ontdekt.

Dit proefschrift erkent de uitdagingen bij het vinden en vertalen van voorspellende biomarkers naar de klinische praktijk om beslissingen te nemen voor vrouwen met borstkanker en DCIS en het gebruik van AI-technologie zonder verdere overdiagnose te induceren bij screen-gedetecteerde kanker. Onderzoekers moeten de kosteneffectiviteit van benaderingen blijven verifiëren op basis van hun gebruik voor en na prospectieve validatie. Vroegtijdige HTA heeft een essentiële rol verdiend in het verbeteren van de invoering van technologieën die de gezondheidsresultaten op populatieniveau verbeteren en kostenbesparingen voor het gezondheidszorgsysteem opleveren.

List of Publications

Publications part of this dissertation

Lopes Cardozo JMN, **Byng D**, Drukker CA, et al. Outcome without any adjuvant systemic treatment in stage I ER+/HER2- breast cancer patients included in the MINDACT trial. *Ann Oncol* 2022; **33**(3): 310-20.

Byng D, Retèl VP, Engelhardt EG, et al. Preferences of Treatment Strategies among Women with Low-Risk DCIS and Oncologists. *Cancers* 2021; **13**(16): 3962.

Byng D, Retèl VP, Schaapveld M, Wesseling J, van Harten WH. Treating (low-risk) DCIS patients: What can we learn from real-world cancer registry evidence? *Breast Cancer Res Treat* 2021; **187**(1): 187-96.

Retèl VP*, **Byng D***, Linn SC, et al. Cost-effectiveness analysis of the 70-gene signature compared with clinical assessment in breast cancer based on a randomised controlled trial. *Eur J Cancer* 2020; **137**: 193-203. **(*contributed equally)**

In revision

Byng D, Thomas ST, Rushing CN, et al. Surveillance imaging after primary diagnosis of ductal carcinoma in situ. In revision at Radiology.

Under review

Byng D, Schaapveld M, Lips EH, et al. An early economic evaluation of active surveillance for low-risk DCIS.

Other publications

Byng D, Strauch B, Gnas L, et al. AI-based prevention of interval cancers in a national mammography screening program. *Eur J Radiol* 2022; **152**: 110321.

Leibig C, Brehmer M, Bunk S, **Byng D**, Pinker K, Umutlu L. Combining the strengths of radiologists and AI for breast cancer screening: a retrospective analysis. *Lancet Digit Health* 2022; **4**(7): e507-e19.

Byng D, Lutter JI, Wacker ME, et al. Determinants of healthcare utilization and costs in COPD patients: first longitudinal results from the German COPD cohort COSYCONET. *Int J Chron Obstruct Pulmon Dis* 2019; **14**: 1423.

Kohler JC, Mitsakakis N, Saadat F, **Byng D**, Martinez MG. Does pharmaceutical pricing transparency matter? Examining Brazil's public procurement system. *Glob Health* 2015; **11**(1): 1-13.

Congress Presentations

Byng D, Bunk S, Schueler D, et al. Prospective post-marketing surveillance of AI for breast cancer screening in clinical practice. European Society of Radiology Congress 2022.

Cardozo JML, **Byng D**, Drukker CA, et al. Abstract PS11-01: Outcome without adjuvant systemic treatment in breast cancer patients included in the MINDACT trial. *Cancer Res* 2021; **81**(4_Supplement): PS11-01-PS11-01. San Antonio Breast Cancer Symposium 2020.

Byng D, Retel VP, Harten WvH, et al. Disparities in surveillance imaging after breast conserving surgery for primary DCIS. *J Clin Oncol* 2021; **39**(15_suppl): 6516-. American Society of Clinical Oncology (ASCO) Annual Meeting 2021.

Engelhardt E, **Byng D**, Klaver K, et al. Women diagnosed with Ductal Carcinoma In Situ (DCIS) and healthcare providers' views on active surveillance for DCIS. Results from focus groups and in-depth interviews. *Eur J Cancer* 2020; **138**: S36. European Breast Cancer Conference 2020.

Byng D, Retèl V, Schaapveld M, Wesseling J, van Harten W. Non-intervention vs. surgical interventions in (Low-Risk) Ductal Carcinoma In Situ: A DCIS multi-state model for decision analytics. *Eur J Cancer* 2020; **138**: S25. European Breast Cancer Conference 2020.

Acknowledgements

An accomplishment like a PhD would not have been possible without the support and encouragement of many different people throughout my life. Doors have been opened for me that have led me to have had incredibly fulfilling experiences as I establish my career as a researcher. Wim and Valesca, you created this PhD position for someone to continue the important work in breast cancer and HTA that you started at the NKI. You have both been trailblazers in bringing value-based healthcare to our institute and driving the field forward in the Netherlands. I have learned a lot from you and greatly benefited from your pragmatism, honesty, and openness. Thank you for your guidance, and for allowing me the creative freedom to develop our projects together. Most of all, thank you for believing in me, believing in my potential, and for the generous opportunity to work with you at the NKI.

I wish to show my appreciation to the professors making up my PhD graduation committee: prof. dr. C.J.M. Doggen, prof. dr. M.A. Joore, prof. dr. M.K. Schmidt, prof. dr. S. Siesling, and prof. dr. C.A. Uyl-de Groot. You each play an important role in developing the next generation of researchers, and I feel honoured that you have invested your time in participating in my PhD defence. Thank you for your meaningful feedback and for sharing your unique perspectives.

The PSOE department is filled with talented and motivated researchers, with some very prodigious leaders at the helm. Dr. M. Schaapveld, thank you for generously sharing with me your breast cancer and DCIS expertise as we navigated creative new ways to build a statistical model of DCIS. The assistance provided by prof. dr. M. Hauptmann and dr. K. Jóźwiak during our time working on MINDACT projects was also greatly appreciated. I gained a lot of inspiration for my research from participating in the EPI meetings led by prof. dr. F. van Leeuwen and the thoughtful and creative research staff, postdocs, and PhD researchers in our department.

I am endlessly inspired by the individuals who made up my OOA committee: prof. dr. E. Bleiker, prof. dr. E. Rutgers, prof. dr. L. van de Poll, and prof. dr. J. Wesseling. Each of you has made important strides in breast cancer research, and I am grateful to have benefitted from your expertise and wisdom. I always looked forward to our annual meetings and felt so energized by our discussions about the state of oncology research.

Part of this PhD was generously funded through the Grand Challenge PRECISION (PREvent ductal Carcinoma In Situ Invasive Overtreatment Now) Consortium, a recipient of the Cancer Research UK Grand Challenge Award 2017, jointly funded by

Cancer Research UK and Dutch Cancer Society (KWF). Research funding for MINDACT projects was also generously provided by Agendia BV. The accomplishments during this PhD were made possible by the interdisciplinary collaborations afforded through these research grants.

Ellen, Miranda, and Eveline, our weekly meetings and collaborations within the PRECISION consortium were a significant opportunity to understand the lived experiences of women with DCIS. Thank you for expanding my horizons and keeping our important research on track. The PRECISION consortium brought together so many talented scientists and medical professionals with the shared goal of further understanding DCIS and optimizing care for women with a DCIS diagnosis. I must express my deepest gratitude to Prof. dr. Shelley Hwang and prof. dr. Jelle Wesseling for opening doors for me to conduct research and develop my skills across borders. Prof. dr. F.H. van Duijnhoven and dr. E.H. Lips, thank you for sharing your unique insights and providing critical feedback – this was a perfect example of interdisciplinary work! Dr. C.G.M. Oudshoorn and dr. J.A. van Til – thank you Karin and Janine for providing your valuable expertise on preference-based research. Working with the PRECISION consortium and meeting scientists from around the world was the biggest highlight of my research career thus far. Thank you to the whole team for such an enriching experience.

Within MINDACT I was given the chance to work on interesting and uniquely challenging projects. Josephine, it was a pleasure to work so efficiently with you, I admire your professionalism and hope we'll have another chance to collaborate. Prof. dr. ir. H. Koffijberg, thank you Erik for committing a lot of time to help smooth out our cost-effectiveness modelling approaches and for being a patient mentor. Prof. dr. S. Linn, you were one of the first oncologists I interacted with at the NKI and I quickly developed a deep respect for you, your profession, and the important strides you've made as an individual working in this field. It was an honour to benefit from your experience and expertise during my early days at the NKI.

Marc, you have provided me with invaluable mentorship over the past two and a half years. Thank you for devoting so much time and effort – not only to our collaboration but also to helping shape me into a more conscientious and detail-oriented researcher. I hope that we will have more opportunities to work together in the future.

My horizons have continued to expand within the world of breast cancer research, thanks to the incredible opportunity to work at Vara. Special thanks to Rob, for

bringing me into the company and providing me guidance; Christian and Stefan for welcoming me into the machine learning community and helping me find my niche; and Anja for always looking out for me and making me feel at home.

Life and its many challenges happen alongside a PhD, and I've taken great joy in learning and growing alongside the wonderful and unique 'van Harten' group. It was truly something special to have this group lift each other and celebrate accomplishments, but also to be there for each other to commiserate during difficult moments. I will always cherish our memories together in Amsterdam. Many thanks (in alphabetical order) to Anke, Ann-Jean, Bruno, Emine, Hester, Joost, Julie, Inge, Laura, Lucas, Melanie, Nora, Simone, Willeke, and Wim G.

A PhD isn't just an intellectual endeavour, it comes with grit and determination. I am grateful to my parents for creating the environment for me to grow into someone who could achieve this and more. Dad, you instilled in me a love of learning, and Mom, you instilled in me a strong work ethic. Thank you for all your sacrifices and understanding and accepting my ambitions – even though it meant leaving Toronto all those years ago and starting a new life in Europe. My gratitude extends to my parents' partners, Mary and Jerome, for supporting my parents while they've had to deal with the sorrow of being separated from their daughter.

During my time in Amsterdam, I made friendships that were an important source of energy and inspiration while completing this PhD. Jessica, thank you for sharing a home with me during my first years in the city and for all the calming and insightful discussions. Adele, we made it through the worst of the pandemic together and lived through a truly historic period. I admire your resilience and persisting sense of humour that brought a lot of sunshine into our home during this very difficult time. Gina, you are a kindred spirit and I'm grateful that we could navigate the highs and lows of doctoral research together.

Finally, I could not have completed this journey without the support of my partner Tobi. Thank you for always pushing me forward in the most challenging moments. Now you can finally put 'Dr. Byng' on the mailbox!

I've benefited greatly from the path that researchers before me have carved out, and from the sacrifices made by family and friends to support the completion of this PhD. This accomplishment is dedicated to the many people who helped me along the way, and to the millions of women worldwide who will have a breast cancer diagnosis in their lifetime.

About the author

Danalyn Tulagan Byng is an internationally-trained cancer epidemiologist and health services researcher. Born in Toronto, Canada, she is a second-generation immigrant and a first-generation graduate student. Danalyn centres her work around improving access to value-based healthcare for all individuals.

Throughout her research career, she became increasingly aware of the disparities in healthcare delivery and health outcomes for underserved populations. It has driven her to focus her research on epidemiological questions about burden of disease and access to healthcare.

After graduating in 2013 from York University with an Honours Bachelor of Health Studies specializing in health policy, she began work as a Policy and Programs Assistant with the Ontario Ministry of Health's Assistive Devices Program. In 2014 she started a graduate program in health services research at the University of Toronto. During this time, she conducted research on pharmaceutical pricing transparency at the WHO Collaborating Centre for Governance, Transparency & Accountability in the Pharmaceutical Sector. In 2015, Danalyn moved to Munich, Germany to complete her M.Sc. in Epidemiology at the Ludwig-Maximilians University. She continued to develop her knowledge in health economics and health technology assessment (HTA) through scientific internships at the Helmholtz Center in Munich and Costello Medical Consulting Ltd. in Cambridge.

In 2017 Danalyn joined prof. dr. W.H. van Harten and dr. V.P. Retèl as a doctoral researcher in their HTA research group in the Department of Psychosocial Research and Epidemiology at the Netherlands Cancer Institute in Amsterdam. Through collaborations with individuals from across the spectrum of care, her Ph.D. research explored innovative approaches to optimizing diagnosis, treatment and follow-up care for early-stage breast cancer and DCIS from a multitude of perspectives. Today, Danalyn continues to focus on using innovative technologies to improve breast cancer care for underserved women across the world. She leads the clinical evidence generation strategy for Vara, an artificial intelligence start-up that has developed CE-marked AI software for breast cancer screening.